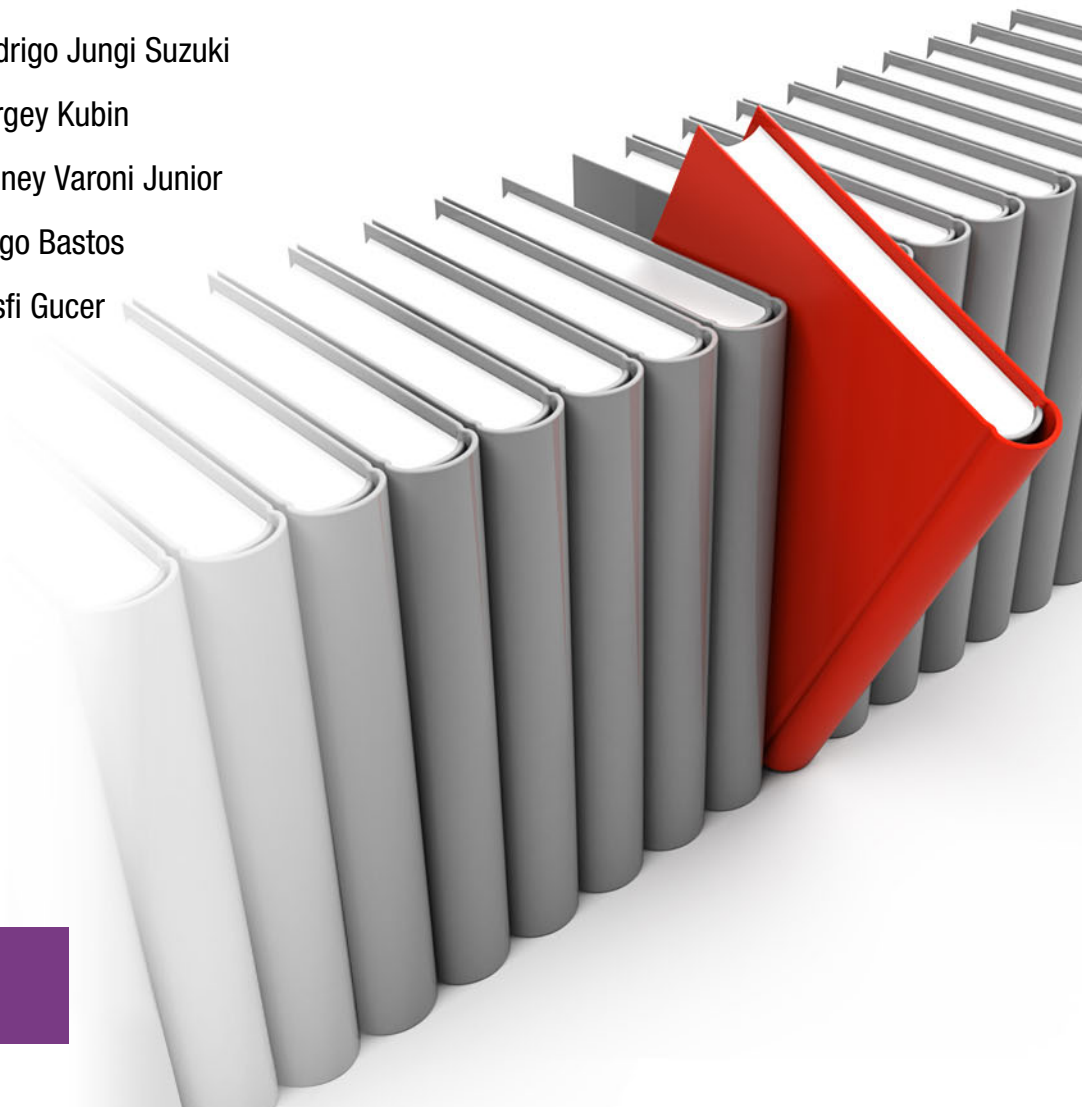


Implementing IBM FlashSystem with IBM Spectrum Virtualize V8.4

Corne Lottering
Denis Olshanskiy
Jackson Shea
Jordan Fincher
Hartmut Lonzer
Ibrahim Alade Rufai
Katja Kratt
Konrad Trojok
Leandro Torolho
Pawel Brodacki

Rodrigo Jungi Suzuki
Sergey Kubin
Sidney Varoni Junior
Tiago Bastos
Vasfi Gucer



Storage



IBM Redbooks

**Implementing the IBM FlashSystem with IBM Spectrum
Virtualize V8.4**

February 2021

Note: Before using this information and the product it supports, read the information in “Notices” on page xv.

First Edition (February 2021)

This edition applies to IBM Spectrum Virtualize Version 8.4.

© Copyright International Business Machines Corporation 2021. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Notices	xv
Trademarks	xvi
Preface	xvii
Authors	xvii
Now you can become a published author, too!	xxi
Comments welcome	xxii
Stay connected to IBM Redbooks	xxii
Chapter 1. Introduction and system overview	1
1.1 IBM Spectrum Virtualize	2
1.2 Latest changes and enhancements	3
1.3 IBM FlashSystem family	4
1.4 IBM FlashSystem 9200 overview	5
1.4.1 IBM FlashSystem 9200 hardware components	5
1.4.2 IBM FlashSystem 9200 Control Enclosure	7
1.4.3 IBM FlashSystem 9000 Expansion Enclosure Models AFF and A9F	10
1.5 IBM FlashSystem 9200R Rack Solution overview	12
1.5.1 Minimum IBM FlashSystem 9200R Rack Solution configuration in the rack	15
1.5.2 Maximum configuration of an IBM FlashSystem 9200R Rack Solution with Model A9F Expansion Enclosures	16
1.5.3 Maximum configuration of an IBM FlashSystem 9200R Rack Solution with Model AFF Expansion Enclosures	17
1.5.4 FC cabling and clustering	18
1.5.5 IBM FlashSystem 9200R Rack Solution FC configuration with IBM SAN32C-6 switches	20
1.5.6 IBM FlashSystem 9200R Rack Solution FC configuration with IBM SAN24B-6 switches	22
1.5.7 IBM FlashSystem 9200R Rack Solution SAS Expansion Enclosures cabling ...	23
1.6 IBM FlashSystem 7200 overview	25
1.6.1 IBM FlashSystem 7200 Control Enclosure	26
1.6.2 IBM FlashSystem 7200 Expansion Enclosures 12G, 24G, and 92G	27
1.6.3 IBM FlashSystem 7200 Utility Model U7C	30
1.7 IBM FlashSystem 5200 overview	31
1.8 IBM FlashSystem 5100 overview	33
1.9 IBM FlashSystem 5000 family overview	37
1.9.1 IBM FlashSystem 5015	37
1.9.2 IBM FlashSystem 5035	39
1.9.3 IBM FlashSystem 5010 hardware overview	40
1.9.4 IBM FlashSystem 5030 hardware overview	42
1.10 Features for storage efficiency and data reduction	44
1.10.1 Compression and deduplication	47
1.10.2 Features for enhanced data security	48
1.11 Features for application integration	49
1.12 Features for manageability	50
1.13 Copy services	54
1.13.1 HyperSwap	57
1.14 IBM FlashCore Module drives, NVMe SSDs, and SCM drives	57
1.15 Storage virtualization	61

1.16 Business continuity	64
1.16.1 Business continuity with HyperSwap	64
1.16.2 Business continuity with three-site replication.	65
1.17 Licensing	66
1.17.1 Licensing IBM FlashSystem 9200, IBM FlashSystem 9200R, and IBM FlashSystem 7200	66
1.17.2 Licensing IBM FlashSystem 5100.	67
1.17.3 Licensing IBM FlashSystem 5030 and IBM FlashSystem 5010	68
Chapter 2. Planning	71
2.1 General planning rules	72
2.2 Planning for availability	73
2.3 Physical installation planning	73
2.4 Planning for system management.	74
2.4.1 User password creation options	75
2.5 Connectivity planning	76
2.6 Fibre Channel SAN configuration planning	76
2.6.1 Physical topology	77
2.6.2 Zoning.	77
2.6.3 N_Port ID Virtualization.	78
2.6.4 Inter-node zone.	79
2.6.5 Back-end storage zones	79
2.6.6 Host zones	80
2.6.7 Zoning considerations for Metro Mirror and Global Mirror	81
2.6.8 Port designation recommendations.	81
2.7 IP SAN configuration planning	83
2.7.1 iSCSI and iSER protocols.	83
2.7.2 Priority flow control	84
2.7.3 RDMA clustering	85
2.7.4 iSCSI back-end storage attachment	85
2.7.5 IP network host attachment	86
2.7.6 Native IP replication	86
2.7.7 Firewall planning	87
2.8 Planning topology	87
2.8.1 High availability	87
2.8.2 Three-site replication	88
2.9 Back-end storage configuration	88
2.10 Internal storage configuration	90
2.11 Storage pool configuration	92
2.11.1 Child pools	94
2.11.2 The storage pool and cache relationship	95
2.12 Volume configuration	96
2.12.1 Planning for image mode volumes	96
2.12.2 Planning for fully allocated volumes	96
2.12.3 Planning for thin-provisioned volumes	96
2.12.4 Planning for compressed volumes	97
2.12.5 Planning for deduplicated volumes	97
2.13 Host attachment planning	98
2.13.1 Queue depth	98
2.13.2 Microsoft Offloaded Data Transfer	98
2.13.3 SAN boot support	99
2.13.4 Planning for large deployments	99
2.13.5 Planning for SCSI UNMAP	99

2.14	Planning copy services	100
2.14.1	FlashCopy guidelines	100
2.14.2	Planning for Metro Mirror and Global Mirror	100
2.15	Data migration	103
2.16	Performance monitoring with IBM Storage Insights	104
2.17	Configuration backup procedure	106
Chapter 3. Initial configuration		107
3.1	Prerequisites	108
3.2	System initialization	109
3.2.1	System initialization process	110
3.3	System setup	113
3.3.1	System setup wizard	113
3.4	Base configuration	123
3.4.1	Configuring Remote Direct Memory Access clustering	123
3.4.2	Adding an enclosure	127
3.4.3	Changing the system topology to HyperSwap	129
3.4.4	Configuring quorum disks or applications	131
3.4.5	Configuring the local Fibre Channel port masking	134
3.4.6	Automatic configuration for IBM SAN Volume Controller back-end storage	136
3.5	Configuring management access	139
3.5.1	Configuring secure communications	139
3.5.2	Configuring password policies	142
3.5.3	Configuring user authentication	146
Chapter 4. IBM Spectrum Virtualize GUI		155
4.1	Performing operations by using the GUI	156
4.1.1	Accessing the GUI	156
4.2	GUI introduction	160
4.2.1	Task menu	160
4.2.2	Suggested tasks	161
4.2.3	Notification icons and help	162
4.3	System Hardware - Overview window	166
4.3.1	Content-based organization	166
4.4	Monitoring menu	170
4.4.1	System Hardware overview	171
4.4.2	Easy Tier Reports	175
4.4.3	Events	177
4.4.4	Performance	178
4.4.5	Background Tasks	179
4.5	Pools	180
4.6	Volumes	180
4.7	Hosts	181
4.8	Copy Services	182
4.9	Access	182
4.9.1	Ownership groups	183
4.9.2	Users by groups	189
4.9.3	Audit log	194
4.10	Settings	196
4.10.1	Notifications	196
4.10.2	Network	200
4.10.3	Using the management GUI	203
4.10.4	Security	209

4.10.5	System menus	212
4.10.6	Support menu	225
4.10.7	GUI Preferences menu	226
4.11	Additional frequent tasks in the GUI	230
4.11.1	Renaming components	230
4.11.2	Working with enclosures	232
4.11.3	Restarting the GUI service	235
Chapter 5.	Storage pools	237
5.1	Working with storage pools	238
5.1.1	Creating storage pools	240
5.1.2	Managed disks in a storage pool	243
5.1.3	Actions on storage pools	244
5.1.4	Child pools	252
5.1.5	Encrypted storage pools	257
5.2	Working with internal drives and arrays	257
5.2.1	Working with drives	257
5.2.2	RAID and distributed redundant array of independent disks	265
5.2.3	Creating arrays	273
5.2.4	Actions on arrays	279
5.3	Working with external controllers and MDisks	286
5.3.1	External storage controllers	286
5.3.2	Actions for external storage controllers	289
5.3.3	Working with external MDisks	290
5.3.4	Actions for external MDisks	292
Chapter 6.	Volumes	299
6.1	Introduction to volumes	300
6.2	Volume characteristics	300
6.2.1	Volume type	301
6.2.2	Managed mode and image mode	302
6.2.3	VSize	305
6.2.4	Performance	306
6.2.5	Volume copies	306
6.2.6	I/O operations data flow	309
6.2.7	Storage efficiency	311
6.2.8	Encryption	316
6.2.9	Cache mode	316
6.2.10	I/O throttling	317
6.2.11	Volume protection	318
6.2.12	Secure data deletion	318
6.3	Virtual volumes	318
6.4	Volumes in multi-site topologies	319
6.5	Operations on volumes	321
6.5.1	Creating volumes	321
6.5.2	Creating custom volumes	330
6.5.3	HyperSwap volumes	334
6.5.4	I/O throttling	339
6.5.5	Volume protection	344
6.5.6	Modifying a volume	345
6.5.7	Deleting a volume	357
6.5.8	Mapping a volume to a host	358
6.5.9	Modify I/O Group or Non-disruptive Volume Move	363

6.5.10 Migrating a volume to another storage pool	368
6.6 Volume operations by using the CLI	376
6.6.1 Displaying volume information	376
6.6.2 Creating a volume	376
6.6.3 Creating a thin-provisioned volume	379
6.6.4 Creating a volume in image mode	379
6.6.5 Adding a volume copy	380
6.6.6 Splitting a mirrored volume	386
6.6.7 Modifying a volume	388
6.6.8 Deleting a volume	389
6.6.9 Volume protection	390
6.6.10 Expanding a volume	390
6.6.11 HyperSwap volume modification with CLI	391
6.6.12 Mapping a volume to a host	392
6.6.13 Listing volumes that are mapped to the host	394
6.6.14 Listing hosts that are mapped to the volume	394
6.6.15 Deleting a volume to host mapping	395
6.6.16 Migrating a volume	395
6.6.17 Migrating a fully managed volume to an image mode volume	396
6.6.18 Shrinking a volume	397
6.6.19 Listing volumes that use MDisks	398
6.6.20 Listing MDisks that are used by the volume	398
6.6.21 Listing volumes that are defined in the storage pool	398
6.6.22 Listing storage pools in which a volume has its extents	399
6.6.23 Tracing a volume from a host back to its physical disks	401
Chapter 7. Hosts	405
7.1 Host attachment overview	406
7.2 Host objects overview	407
7.3 NVMe over Fibre Channel	408
7.4 N_Port ID Virtualization support	410
7.4.1 NPIV prerequisites	413
7.4.2 Verifying the NPIV mode state for a new system installation	413
7.4.3 Enabling NPIV on an existing system	414
7.5 Hosts operations by using the GUI	418
7.5.1 Creating hosts	419
7.5.2 Host clusters	433
7.5.3 Actions on hosts	437
7.5.4 Actions on host clusters	448
7.5.5 Host management views	454
7.6 Performing hosts operations by using CLI	461
7.6.1 Creating a host by using the CLI	461
7.6.2 Host administration by using the CLI	463
7.6.3 Adding and deleting a host port by using the CLI	466
7.6.4 Host cluster operations	468
7.6.5 Adding a host or host cluster to an ownership group	470
7.7 Host attachment practical examples	471
7.7.1 Prerequisites	471
7.7.2 Fibre Channel host connectivity and capacity allocation	471
7.7.3 iSCSI host connectivity and capacity allocation	475
7.7.4 NVMe over Fabric host connectivity example	478
Chapter 8. Storage migration	485

8.1 Storage migration overview	486
8.1.1 Interoperability and compatibility	487
8.1.2 Prerequisites	488
8.2 Storage migration wizard	489
8.3 Enclosure Upgrade Migration	507
Chapter 9. Advanced features for storage efficiency	509
9.1 IBM Easy Tier	510
9.1.1 Easy Tier concepts	510
9.1.2 Implementing and tuning Easy Tier.	516
9.1.3 Monitoring Easy Tier activity	523
9.2 Thin-provisioned volumes	528
9.2.1 Concepts.	529
9.2.2 Implementation	529
9.3 UNMAP	530
9.3.1 The SCSI UNMAP command	530
9.3.2 Back-end SCSI UNMAP	531
9.3.3 Host SCSI UNMAP	531
9.3.4 Offloading I/O throttle	532
9.4 Data Reduction Pools	533
9.4.1 Introduction to DRP.	533
9.4.2 DRP benefits.	535
9.4.3 Planning for DRP	536
9.4.4 Implementing DRP with compression and deduplication	538
9.5 Saving estimations for compression and deduplication	545
9.5.1 Evaluating compression savings by using IBM Comprestimator	545
9.5.2 Evaluating compression and deduplication.	547
9.6 Overprovisioning and data reduction on external storage.	548
Chapter 10. Advanced Copy Services	553
10.1 IBM FlashCopy	554
10.1.1 Business requirements for FlashCopy	554
10.1.2 FlashCopy principles and terminology	556
10.1.3 FlashCopy mapping	556
10.1.4 Consistency groups	557
10.1.5 Crash consistent copy and hosts considerations	558
10.1.6 Grains and bitmap: I/O indirection.	559
10.1.7 Interacting with the cache	565
10.1.8 Background Copy Rate.	566
10.1.9 Incremental FlashCopy.	567
10.1.10 Starting FlashCopy mappings and consistency groups	568
10.1.11 Multiple target FlashCopy	570
10.1.12 Reverse FlashCopy	575
10.1.13 FlashCopy and image mode volumes.	577
10.1.14 FlashCopy mapping events	578
10.1.15 Thin-provisioned FlashCopy	580
10.1.16 Serialization of I/O by FlashCopy	581
10.1.17 Event handling	581
10.1.18 Asynchronous notifications	582
10.1.19 Interoperation with Metro Mirror and Global Mirror.	582
10.1.20 FlashCopy attributes and limitations.	583
10.2 Managing FlashCopy by using the GUI	584
10.2.1 FlashCopy presets	584

10.2.2	FlashCopy window	586
10.2.3	Creating a FlashCopy mapping	589
10.2.4	Single-click snapshot	599
10.2.5	Single-click clone	601
10.2.6	Single-click backup	603
10.2.7	Creating a FlashCopy consistency group	605
10.2.8	Creating FlashCopy mappings in a consistency group	606
10.2.9	Showing related volumes	609
10.2.10	Moving FlashCopy mappings across consistency groups	610
10.2.11	Removing FlashCopy mappings from consistency groups	611
10.2.12	Modifying a FlashCopy mapping	612
10.2.13	Renaming FlashCopy mappings	613
10.2.14	Deleting FlashCopy mappings	615
10.2.15	Deleting a FlashCopy consistency group	616
10.2.16	Starting FlashCopy mappings	617
10.2.17	Stopping FlashCopy mappings	618
10.2.18	Memory allocation for FlashCopy	619
10.3	Transparent Cloud Tiering	621
10.3.1	Considerations for using Transparent Cloud Tiering	622
10.3.2	Transparent Cloud Tiering as backup solution and data migration	622
10.3.3	Restoring data by using Transparent Cloud Tiering	623
10.3.4	Transparent Cloud Tiering restrictions	623
10.4	Implementing Transparent Cloud Tiering	624
10.4.1	Domain Name System configuration	624
10.4.2	Enabling Transparent Cloud Tiering	624
10.4.3	Creating cloud snapshots	627
10.4.4	Managing cloud snapshots	630
10.4.5	Restoring cloud snapshots	631
10.5	Volume mirroring and migration options	634
10.6	Remote Copy	636
10.6.1	IBM SAN Volume Controller and IBM FlashSystem system layers	637
10.6.2	Multiple IBM Spectrum Virtualize systems replication	638
10.6.3	Importance of write ordering	641
10.6.4	Remote Copy intercluster communication	642
10.6.5	Metro Mirror overview	644
10.6.6	Synchronous Remote Copy	645
10.6.7	Metro Mirror features	645
10.6.8	Metro Mirror attributes	646
10.6.9	Practical use of Metro Mirror	646
10.6.10	Global Mirror overview	647
10.6.11	Asynchronous Remote Copy	648
10.6.12	Global Mirror features	649
10.6.13	Using Global Mirror with Change Volumes	651
10.6.14	Distribution of work among nodes	653
10.6.15	Background copy performance	654
10.6.16	Thin-provisioned background copy	654
10.6.17	Methods of synchronization	654
10.6.18	Practical use of Global Mirror	655
10.6.19	IBM Spectrum Virtualize HyperSwap topology	655
10.6.20	Consistency Protection for Global Mirror and Metro Mirror	656
10.6.21	Valid combinations of FlashCopy, Metro Mirror, and Global Mirror	657
10.6.22	Remote Copy configuration limits	657
10.6.23	Remote Copy states and events	658

10.7 Remote Copy commands	665
10.7.1 Remote Copy process	665
10.7.2 Listing available system partners	666
10.7.3 Changing the system parameters	666
10.7.4 System partnership	667
10.7.5 Creating a Metro Mirror/Global Mirror consistency group	668
10.7.6 Creating a Metro Mirror/Global Mirror relationship	669
10.7.7 Changing a Metro Mirror/Global Mirror relationship	669
10.7.8 Changing a Metro Mirror/Global Mirror consistency group	669
10.7.9 Starting a Metro Mirror/Global Mirror relationship	669
10.7.10 Stopping a Metro Mirror/Global Mirror relationship	670
10.7.11 Starting a Metro Mirror/Global Mirror consistency group	670
10.7.12 Stopping a Metro Mirror/Global Mirror consistency group	670
10.7.13 Deleting a Metro Mirror/Global Mirror relationship	671
10.7.14 Deleting a Metro Mirror/Global Mirror consistency group	671
10.7.15 Reversing a Metro Mirror/Global Mirror relationship	671
10.7.16 Reversing a Metro Mirror/Global Mirror consistency group	672
10.8 Native IP replication	672
10.8.1 Native IP replication technology	672
10.8.2 IP partnership limitations	674
10.8.3 IP Partnership and data compression	676
10.8.4 VLAN support	676
10.8.5 IP partnership and terminology	677
10.8.6 States of IP partnership	678
10.8.7 Remote Copy groups	679
10.8.8 Supported configurations	680
10.9 Managing Remote Copy by using the GUI	693
10.9.1 Creating a Fibre Channel partnership	695
10.9.2 Creating Remote Copy relationships	697
10.9.3 Creating a consistency group	704
10.9.4 Renaming Remote Copy relationships	705
10.9.5 Renaming a Remote Copy consistency group	706
10.9.6 Moving stand-alone Remote Copy relationships to a consistency group	707
10.9.7 Removing Remote Copy relationships from a consistency group	708
10.9.8 Starting Remote Copy relationships	709
10.9.9 Starting a Remote Copy consistency group	710
10.9.10 Switching a relationship copy direction	710
10.9.11 Switching a consistency group direction	712
10.9.12 Stopping Remote Copy relationships	713
10.9.13 Stopping a consistency group	714
10.9.14 Deleting Remote Copy relationships	715
10.9.15 Deleting a consistency group	716
10.10 Remote Copy memory allocation	717
10.11 Troubleshooting Remote Copy	718
10.11.1 1920 error	718
10.11.2 1720 error	720
Chapter 11. Ownership groups	723
11.1 Ownership groups principles of operations	724
11.2 Implementing ownership groups on a new system	726
11.2.1 Creating an ownership group	726
11.2.2 Assigning users to an ownership group	726
11.2.3 Creating ownership group resources	728

11.2.4	Listing ownership group resources	729
11.2.5	Actions on ownership groups	730
11.3	Migrating existing objects to ownership groups	731
Chapter 12.	Encryption	735
12.1	General types of encryption across IBM Spectrum Virtualize	736
12.1.1	Externally virtualized storage	736
12.1.2	Serial-attached SCSI internal storage	736
12.1.3	Non-Volatile Memory Express internal storage	736
12.2	Planning for encryption	737
12.3	Defining encryption of data-at-rest	737
12.3.1	Encryption methods	738
12.3.2	Encrypted data	738
12.3.3	Encryption keys	741
12.3.4	Encryption licenses	742
12.4	Activating encryption	742
12.4.1	Obtaining an encryption license	743
12.4.2	Starting the activation process during the initial system setup	743
12.4.3	Starting the activation process on a running system	746
12.4.4	Activating the license automatically	747
12.4.5	Activating the license manually	750
12.5	Enabling encryption	752
12.5.1	Starting the Enable Encryption wizard	753
12.5.2	Enabling encryption by using USB flash drives	755
12.5.3	Enabling encryption by using key servers	759
12.5.4	Enabling encryption by using both providers	771
12.6	Configuring more providers	774
12.6.1	Adding key servers as a second provider	774
12.6.2	Adding USB flash drives as a second provider	776
12.7	Migrating between providers	777
12.7.1	Changing from a USB flash drive provider to an encryption key server	778
12.7.2	Changing from an encryption key server to a USB flash drive provider	778
12.7.3	Migrating between different key server types	779
12.8	Recovering from a provider loss	780
12.9	Using encryption	781
12.9.1	Encrypted pools	781
12.9.2	Encrypted child pools	783
12.9.3	Encrypted arrays	784
12.9.4	Encrypted MDisks	785
12.9.5	Encrypted volumes	787
12.9.6	Restrictions	788
12.10	Rekeying an encryption-enabled system	789
12.10.1	Rekeying by using a key server	789
12.10.2	Rekeying by using USB flash drives	790
12.11	Disabling encryption	792
Chapter 13.	Reliability, availability, and serviceability, monitoring and logging, and troubleshooting	793
13.1	Reliability, availability, and serviceability	794
13.1.1	Node canisters	795
13.1.2	Expansion canisters	800
13.1.3	Dense Drawer Enclosures LED	800
13.1.4	Enclosure SAS cabling	801
13.1.5	IBM FlashCore Module drives	803

13.1.6 Power	804
13.2 Shutting down the IBM FlashSystem	805
13.2.1 Shutting down and powering on a complete infrastructure	805
13.3 Removing or adding a node from or to the system	805
13.4 Configuration backup	808
13.4.1 Backing up by using the CLI	809
13.4.2 Saving the backup by using the GUI	810
13.5 Software update	812
13.5.1 Precautions before the update	812
13.5.2 IBM FlashSystem update test utility	813
13.5.3 Updating your IBM FlashSystem to Version 8.4.0	814
13.5.4 Updating the IBM FlashSystem drive code	822
13.5.5 Manually updating the system	826
13.6 Health checker feature	827
13.7 Troubleshooting and fix procedures	828
13.7.1 Managing the event log	830
13.7.2 Running a fix procedure	832
13.7.3 Event log details	833
13.8 Monitoring	834
13.8.1 Email notifications and the Call Home function	835
13.8.2 Remote Support Assistance	844
13.8.3 SNMP configuration	848
13.8.4 Syslog notifications	850
13.9 Audit log	852
13.10 Collecting support information by using the GUI, CLI, and USB	855
13.10.1 Collecting information by using the GUI	855
13.10.2 Collecting logs by using the CLI	858
13.10.3 Collecting logs by using a USB flash drive	860
13.10.4 Uploading files to the IBM Support Center	860
13.11 Service Assistant Tool	862
13.12 IBM Storage Insights monitoring	865
13.12.1 Capacity monitoring	866
13.12.2 Performance monitoring	868
13.12.3 Logging support tickets by using IBM Storage Insights	870
13.12.4 Managing existing support tickets by using IBM Storage Insights and uploading logs	877
Appendix A. Performance data and statistics gathering	879
IBM Storage System performance overview	880
Performance considerations	880
IBM Spectrum Virtualize performance perspectives	881
Performance monitoring	882
Collecting performance statistics	882
Real-time performance monitoring	884
Performance data collection and IBM Spectrum Control	892
Appendix B. Terminology	895
Commonly encountered terms	896
Appendix C. Command-line interface setup	925
CLI setup	926
Basic setup on a Windows host	926
Basic setup on a Mac, UNIX, or Linux host	935

Related publications	939
IBM Redbooks	939
Help from IBM	939
Abbreviations and acronyms	941

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at <http://www.ibm.com/legal/copytrade.shtml>

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

AIX®	IBM FlashCore®	PowerHA®
Db2®	IBM FlashSystem®	PureSystems®
DS8000®	IBM Garage™	Real-time Compression Appliance®
Easy Tier®	IBM Research®	Redbooks®
FICON®	IBM Security™	Redbooks (logo)  ®
FlashCopy®	IBM Spectrum®	Scalable POWERparallel Systems®
HyperSwap®	Informix®	Storwize®
IBM®	Insight®	XIV®
IBM Cloud®	MicroLatency®	

The following terms are trademarks of other companies:

Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

ITIL is a Registered Trade Mark of AXELOS Limited.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

OpenShift, Red Hat, are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

VMware, VMware vSphere, and the VMware logo are registered trademarks or trademarks of VMware, Inc. or its subsidiaries in the United States and/or other jurisdictions.

Other company, product, or service names may be trademarks or service marks of others.

Preface

Continuing its commitment to developing and delivering industry-leading storage technologies, IBM® introduces the IBM FlashSystem® solution that is powered by IBM Spectrum® Virtualize V8.4. This innovative storage offering delivers essential storage efficiency technologies and exceptional ease of use and performance, all integrated into a compact, modular design that is offered at a competitive, midrange price.

The solution incorporates some of the top IBM technologies that are typically found only in enterprise-class storage systems, which raise the standard for storage efficiency in midrange disk systems. This cutting-edge storage system extends the comprehensive storage portfolio from IBM and can help change the way organizations address the ongoing information explosion.

This IBM Redbooks® publication introduces the features and functions of an IBM Spectrum Virtualize V8.4 system through several examples. This book is aimed at pre-sales and post-sales technical support and marketing and storage administrators. It helps you understand the architecture, how to implement it, and how to take advantage of its industry-leading functions and features.

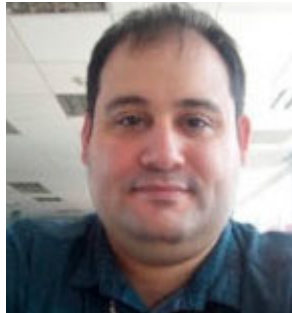
IBM Documentation: In this book, we provide links to IBM Documentation and a description of the relevant section that provides more information. Our starting point is the IBM FlashSystem 9200 family page, and the reader might have to select the product that applies to their environment.

Authors

This book was produced by a team of specialists from around the world.



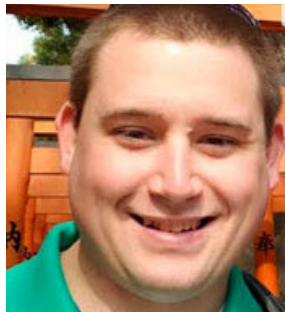
Corne Lottering is a Storage Client Technical Specialist in the US, focusing on technical sales in Texas and Louisiana within the Public Sector industry. He has been with IBM for more than 20 years, and has experience in a wide variety of storage technologies, including the IBM System Storage DS5000, IBM DS8000®, IBM Storwize®, IBM XIV®, IBM FlashSystem, IBM SAN switches, IBM Tape Systems, and software-defined storage software. Since joining IBM, he has fulfilled roles in support, implementation, and pre-sales support across various African and Middle Eastern countries. Corne is the author of several IBM Redbooks publications that are related to the midrange IBM System Storage DS Storage Manager range of products and IBM FlashSystem products.



Denis Olshanskiy is a Storage Specialist with a demonstrated history of working in the IT and services industry. His areas of expertise include storage area networks (SANs), data centers, storage solutions, Arduino, and Linux. He has a master's degree in Mechatronics, Robotics, and Automation Engineering from Budapest University of Technology and Economics.



Jackson Shea is a Level 2 certified IBM Information Technology Specialist/Architect performing design and implementation engagements through IBM Lab Services. He has been with IBM since April 2010. He was a Lead Storage Administrator with a large health insurance consortium in the Pacific Northwest, and has been working with IBM equipment since 2002. He has over 12 years of experience with IBM Spectrum Virtualize, formerly known as the IBM SAN Volume Controller (SVC), and related technologies. Jackson is based out of Portland, Oregon. He received his Bachelor of Science degree in Philosophy with minors in Communications and Chemistry from Lewis & Clark College. Jackson's professional focus is IBM Spectrum Virtualize, but he is conversant with SAN design, implementation, extension, and storage encryption.



Jordan Fincher is an SVC and IBM FlashSystem Level 3 Support Engineer. He has contributed to several IBM Redbooks publications and periodically speaks at IBM Technical University events.



Hartmut Lonzer is the IBM FlashSystem Territory Account Manager for Germany (D), Austria (A), and Switzerland (CH) (DACH). Before this position, he was OEM Alliance Manager for Lenovo at IBM Germany. He works at the IBM Germany headquarters in Ehningen. His main focus is on the IBM FlashSystem family and SVC. His experience with SVC and IBM FlashSystem products goes back to the beginning of these products. Hartmut has been with IBM in various technical and sales roles for 43 years.



Ibrahim Alade Rufai has expertise with designing, building, and implementing enterprise cloud and artificial intelligence (AI) projects, storage, and software-defined infrastructure systems for cognitive products. He helps clients across the Middle East and Africa design for cognitive business, build with collaborative innovation, and deliver through a cloud platform (private, public, hybrid, and multicloud).



Katja Kratt is a Product Field Engineer for IBM Spectrum Virtualize products and IBM V9000 in the EMEA Storage Competence Center (ESCC) in Kelsterbach, Germany. She has 33 years with IBM, mostly with technical support and technical education. She has co-authored several IBM Redbooks publications.



Konrad Trojok has been the technical team lead for the IBM Storage team at System Vertrieb Alexander GmbH for the last 9 years. His role includes being an active part in the daily IBM storage business, such as design, implementation, and taking care of storage solutions. He acts as a strategic advisor for storage solutions. He has worked on IBM Power Systems solutions for IBM Scalable POWERparallel Systems®, and Serial Storage Architecture storage before switching his technical focus to SAN and SAN storage.



Leandro Torolho is a Storage Client Technical Specialist for US Public Market (West). Before joining the technical sales team in 2015, he worked as a SAN and storage subject matter expert (SME) for several international clients. Leandro is an IBM Certified IT Specialist and holds a bachelor's degree in computer science, and a postgraduate degree in computer networks. He has 13 years of experience in storage services and support, and is a Certified Distinguished IT Specialist of The Open Group.



Pawel Brodacki is an Infrastructure Architect with 20 years of experience in IT who has worked for IBM Poland since 2003. His main focus for the last 5 years is on virtual infrastructure architecture from storage to servers to software-defined networks (SDNs). Before changing his profession to system architecture, he was an IBM Certified IT Specialist working on various infrastructure, virtualization, and disaster recovery (DR) projects. His experience includes SAN, storage, highly available (HA) systems, DR solutions, IBM System x and IBM Power Systems servers, and several types of operating systems (OSs) (Linux, IBM AIX®, and Microsoft Windows). Pawel has obtained certifications from IBM, Red Hat, and VMware. Pawel holds a master's degree in biophysics from the University of Warsaw College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences.



Rodrigo Jungi Suzuki is a SAN Storage specialist at IBM Brazil Global Delivery Center in Hortolandia. Currently, Rodrigo is a SME account focal point, and works with projects, implementations, and support for international clients. He has 20 years of IT Industry experience with the last five years in the SAN Storage area. He has a background in UNIX and IBM Informix® databases. He holds a bachelor's degree in computer science from Universidade Paulista in Sao Paulo, Brazil, and is an IBM Certified IT Specialist. Rodrigo also is certified in NetApp NCPA, IBM Storwize V7000 Technical Solutions V2, and the Information Technology Infrastructure Library (ITIL).



Sergey Kubin is a SME for IBM Storage and SAN technical support. He holds an Electronics Engineer degree from Ural Federal University in Russia, and has more than 15 years of experience in IT. At IBM, he works for IBM Technology Support Services, where he provides support and guidance about IBM Spectrum Virtualize family systems for customers in Europe, the Middle East, and Russia. His expertise includes SAN, block-level, and file-level storage systems and technologies. He is IBM Certified Specialist for IBM FlashSystem Family Technical Solutions.



Sidney Varoni Junior is a Storage Technical Advisor for IBM Systems in IBM Brazil. He has over 14 years of experience working with complex IT environments, having worked with both mainframe and open systems platforms. He currently works with clients from Brazil and other countries in Latin America, advising them about how to best use their IBM Storage products. He holds a bachelor's degree in computer science from Faculdade Politecnica de Jundiai. His areas of expertise include high availability (HA), DR, and business continuity solutions, and performance analysis.



Tiago Bastos is a SAN and Storage Disk specialist at IBM Brazil. He has over 20 years in the IT arena, and is an IBM Certified Master IT Specialist. Certified for IBM Storwize, he works on storage as service implementation projects. His areas of expertise include planning, configuring, and troubleshooting DS8000, IBM FlashSystem, SVC, and IBM XIV; lifecycle management; and copy services.



Vasfi Gucer is an IBM Technical Content Services Project Leader with IBM Garage™ for Systems. He has more than 20 years of experience in the areas of systems management, networking hardware, and software. He writes extensively and teaches IBM classes worldwide about IBM products. His focus has been primarily on cloud computing, including cloud storage technologies for the last 6 years. Vasfi is also an IBM Certified Senior IT Specialist, Project Management Professional (PMP), ITIL V2 Manager, and ITIL V3 Expert.

Thanks to the following for their contributions that made this book possible:

Bill Scales, Evelyn Perez, Jamie Pryde, Jon Tate, Greg Shepherd, Liam P Moyna, Lucy Harris, Matthew Smith, Suri Polisetti
IBM Hursley, UK

John Bernatz, Joe Consorti, Karen Brown, Mary Connell, Matt Key, Meagan M Miller, Richard Heffel
IBM US

Markus Oscheka
IBM Germany

Anil Nayak and Virendra P Kucheriya
IBM India

Eli Koren and Rivka Pollack
IBM Israel

Wade Wallace
ITSO Austin, IBM Garage for Systems, US

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an IBM Redbooks residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- ▶ Use the online **Contact us** review Redbooks form found at:

ibm.com/redbooks

- ▶ Send your comments in an email to:

redbooks@us.ibm.com

- ▶ Mail your comments to:

IBM Corporation, IBM Redbooks
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- ▶ Find us on Facebook:

<http://www.facebook.com/IBMRedbooks>

- ▶ Follow us on Twitter:

<http://twitter.com/ibmredbooks>

- ▶ Look for us on LinkedIn:

<http://www.linkedin.com/groups?home=&gid=2130806>

- ▶ Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>

- ▶ Stay current on recent Redbooks publications with RSS Feeds:

<http://www.redbooks.ibm.com/rss.html>



Introduction and system overview

This chapter contains an overview of each of the products that make up the IBM FlashSystem family, describes some major features and functions, and defines the storage virtualization concept.

The following topics are included in this chapter:

- ▶ 1.1, “IBM Spectrum Virtualize” on page 2
- ▶ 1.2, “Latest changes and enhancements” on page 3
- ▶ 1.3, “IBM FlashSystem family” on page 4
- ▶ 1.4, “IBM FlashSystem 9200 overview” on page 5
- ▶ 1.5, “IBM FlashSystem 9200R Rack Solution overview” on page 12
- ▶ 1.6, “IBM FlashSystem 7200 overview” on page 25
- ▶ 1.8, “IBM FlashSystem 5100 overview” on page 33
- ▶ 1.9, “IBM FlashSystem 5000 family overview” on page 37
- ▶ 1.10, “Features for storage efficiency and data reduction” on page 44
- ▶ 1.11, “Features for application integration” on page 49
- ▶ 1.12, “Features for manageability” on page 50
- ▶ 1.13, “Copy services” on page 54
- ▶ 1.14, “IBM FlashCore Module drives, NVMe SSDs, and SCM drives” on page 57
- ▶ 1.15, “Storage virtualization” on page 61
- ▶ 1.16, “Business continuity” on page 64
- ▶ 1.17, “Licensing” on page 66

1.1 IBM Spectrum Virtualize

IBM Spectrum Virtualize is a key member of the IBM Spectrum Storage portfolio. It is a highly flexible storage solution that enables rapid deployment of block storage services for new and traditional workloads, on-premises, off-premises, or a combination of both.

For more information: See the [IBM Spectrum Storage portfolio website](#).

With the introduction of the IBM Spectrum Storage family, the *software* that runs on IBM SAN Volume Controller (SVC) and on IBM FlashSystem products is called IBM Spectrum Virtualize. The name of the underlying *hardware* platform remains intact.

IBM FlashSystem storage systems are built with award-winning IBM Spectrum Virtualize software that simplifies infrastructure and eliminates differences in management, function, and even hybrid multicloud support.

IBM Spectrum Virtualize is an offering that has been available for years for the SVC and IBM FlashSystem family of storage solutions. It provides an ideal way to manage and protect huge volumes of data from mobile and social applications, enable rapid and flexible cloud services deployments, and deliver the performance and scalability that is needed to gain insights from the latest analytics technologies.

The benefits of IBM Spectrum Virtualize

IBM Spectrum Virtualize delivers leading benefits that improve storage infrastructure in many ways, including:

- ▶ Cost reduction of storing data by increasing utilization and accelerating applications to speed business insights. To achieve this goal, the solution:
 - Uses data reduction technologies to increase the amount of data that you can store in the same space.
 - Enables rapid deployment of cloud storage for disaster recovery (DR) along with the ability to store copies of local data.
 - Moves data to the most appropriate type of storage based on policies that you define by using IBM Spectrum Control to optimize storage.
 - Improves storage performance so you can get more done with your data.
- ▶ Data protection from theft or inappropriate disclosure while enabling a high availability (HA) strategy that includes protection for data and application mobility and DR. To achieve this goal, the solution:
 - Uses software-based encryption to improve data security.
 - Provides fully duplexed copies of data and automatic switchover across data centers to improve data availability.
 - Eliminates storage downtime with nondisruptive movement of data from one type of storage to another type.

- ▶ Data simplicity by providing a data strategy that is independent of your choice of infrastructure, delivering tightly integrated functions and consistent management across heterogeneous storage. To achieve this goal, the solution:
 - Integrates with virtualization tools such as VMware vCenter to improve agility with automated provisioning of storage and easy deployment of new storage technologies.
 - Enables supported storage to be deployed with Kubernetes and Docker container environments, including Red Hat OpenShift.
 - Consolidates storage regardless of the hardware vendor for simplified management, consistent functions, and greater efficiency.
 - Supports common capabilities across storage types, providing flexibility in storage acquisition by allowing a mix of vendors in the storage infrastructure.

Note: The benefits that are listed are not a complete list of features and functions that are available with IBM Spectrum Virtualize software.

1.2 Latest changes and enhancements

IBM Spectrum Virtualize V8.4 provides more features and updates to the IBM Spectrum Virtualize family of products of which IBM FlashSystem is part. The major software changes in Version 8.4 are:

- ▶ Data Reduction Pool (DRP) improvements:
 - A Data Reduction Child Pool allows for more flexibility, such as multi-tenancy.
 - IBM FlashCopy® with redirect-on-write (RoW) support, which uses the DRP internal deduplication referencing capabilities to reduce overhead by creating references instead of copying the data. RoW is an alternative to the existing copy-on-write (CoW) capabilities.

Note: At the time of writing, this capability may be used only for volumes with supported deduplication without mirroring relationships and within the same pool and I/O group. The mode selection (RoW/CoW) is automatically based on these conditions.

- Comprestimator is always on, which allows the systems to sample each volume at regular intervals and display the compressibility of the data in the GUI and IBM Storage Insights at any time.
- Redundant array of independent disks (RAID) Reconstruct Read, which increases reliability and availability by reducing the chances of DRP going offline because of fixable array issues. By using RAID capabilities, DRP asks for a specific data block reconstruction when detecting a potential corruption.
- ▶ Distributed redundant array of independent disks 1 (DRAID 1) support extends DRAID advantages to smaller pools of drives, which improves performance over traditional RAID (TRAID) 1 implementations, allowing a better use of flash technology. These DRAIDs can support as few as two drives with no rebuild area, and 3 - 16 drives with a single rebuild area.

Note: At the time of writing, DRAID 1 is supported only on IBM FlashSystem 7200 and IBM FlashSystem 9200, and it is not available for IBM FlashCore® Module (FCM) drives (FCM-XL) of 38.4 TB capacity.

- ▶ With Version 8.4, IBM FlashSystem 5100, IBM FlashSystem 7200, and IBM FlashSystem 9200 systems can support up to 12 storage-class memory (SCM) devices per enclosure with no slot restriction. Previously, the limit for all SCM drives was four per enclosure at the right side.

Note: With Version 8.3, IBM FlashSystem 5100, IBM FlashSystem 7200, and IBM FlashSystem 9200 systems can support up to 12 Z-SSD SCM drives or up to four Optane SCM drives.

- ▶ The expansion of mirrored virtual disks (VDisks) (also known as *volumes*) allows the VDisks capacity to be expanded or reduced online without requiring an offline format and sync. This function improves the availability of the volume for use because the new capacity is available immediately.
- ▶ Three-site replication with IBM HyperSwap® support providing improved availability for data in three-site implementations. This function expands on the DR capabilities that are inherent in this topology.

Important: Three-site replication that uses Metro Mirror (MM) was previously supported on Version 8.3.1 only in limited installations through the RPQ process. With Version 8.4.0, this implementation is generally available.

- ▶ Host attachment support with Non-Volatile Memory Express over Fibre Channel (FC-NVMe) in HyperSwap systems.
- ▶ Domain name server (DNS) support for Lightweight Directory Access Protocol (LDAP) and Network Time Protocol (NTP) with full DNS length (256 characters).
- ▶ Updates to maximum configuration limits, which double FlashCopy mapping from 5,000 to 10,000 and increases the HyperSwap volumes limit from 1,250 to 2,000.

1.3 IBM FlashSystem family

The IBM FlashSystem family, running IBM Spectrum Virtualize software, has been simplified with innovations and enterprise-class features for deployments of all sizes, from entry to mid-range to high-end. A one-platform system allows for ease-of-use to manage seamlessly and securely data across your entire IT infrastructure.

IBM FlashSystem 5010, IBM FlashSystem 5030, and IBM FlashSystem 5100 deliver entry enterprise solutions. IBM FlashSystem 7200 provides a midrange enterprise solution. IBM FlashSystem 9200 and the rack-based IBM FlashSystem 9200R provide two high-end enterprise solutions.

Even though all the IBM FlashSystem family systems are running the same IBM Spectrum Virtualize software, the feature set that is available with each of the models is different.

Figure 1-1 on page 5 shows the feature set that is provided by the IBM FlashSystem systems. Each of the features is described in more detail in further sections of this book.

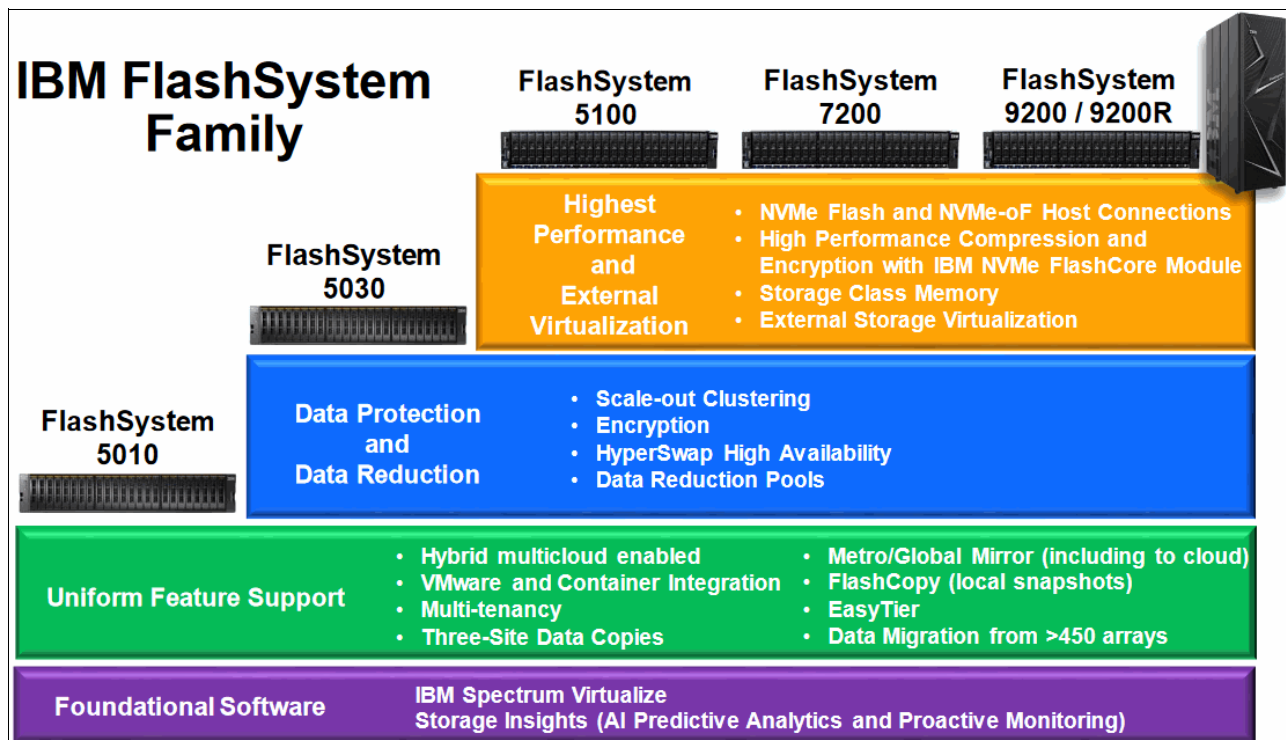


Figure 1-1 IBM FlashSystem

Note: For an analyst report about the IBM FlashSystem family, see [IBM FlashSystem Family: Ease of Use for All Environments](#).

1.4 IBM FlashSystem 9200 overview

This section describes the IBM FlashSystem 9200 architectural components, available models, and enclosure and software features.

1.4.1 IBM FlashSystem 9200 hardware components

IBM FlashSystem 9200 is an all-flash storage system that consists of a control enclosure that runs the IBM Spectrum Virtualize Software and manages your storage system, communicates with the hosts, and manages interfaces.

Figure 1-2 shows the IBM FlashSystem 9200 front and rear views.

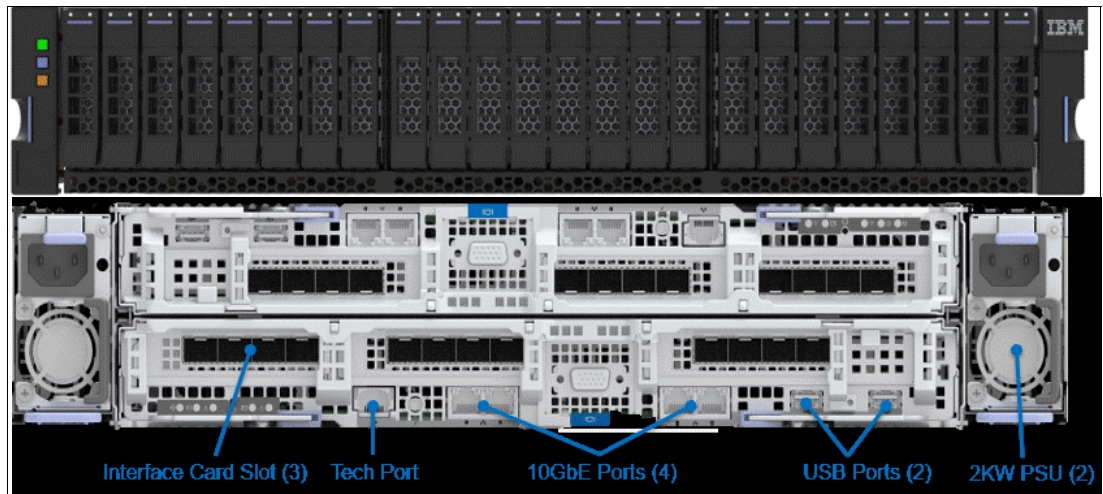


Figure 1-2 IBM FlashSystem 9200 front and rear views

Here are the core IBM FlashSystem 9200 components:

- ▶ IBM FlashSystem 9200 Control Enclosure:
 - Node canisters
 - Power supply units (PSUs)
 - Battery modules
 - Fan modules
 - Interface cards
 - Cascade Lake CPUs and memory slots
 - USB ports
 - Ethernet ports
- ▶ Non-Volatile Memory Express (NVMe)-capable flash drives
- ▶ IBM FlashSystem 9000 Expansion Enclosures (serial-attached Small Computer System Interface (SCSI) (SAS)-attached)

As shown in Figure 1-2, the IBM FlashSystem 9200 enclosure consists of redundant PSUs, node canisters, and fan modules to provide redundancy and HA.

Figure 1-3 shows the IBM FlashSystem 9200 internal architecture.

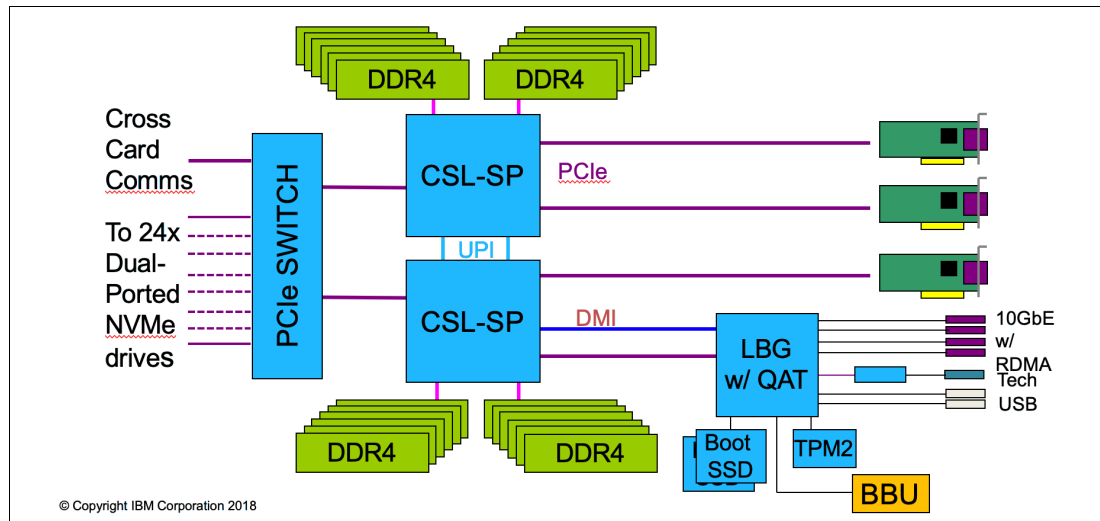


Figure 1-3 IBM FlashSystem 9200 internal architecture

Figure 1-4 shows a picture of the internal hardware components of a node canister. At the left of the picture is the front of the canister, where the fan modules and battery backup are, followed by two Cascade Lake CPUs and memory DIMM slots and Peripheral Component Interconnect Express (PCIe) risers for the adapters on the right.

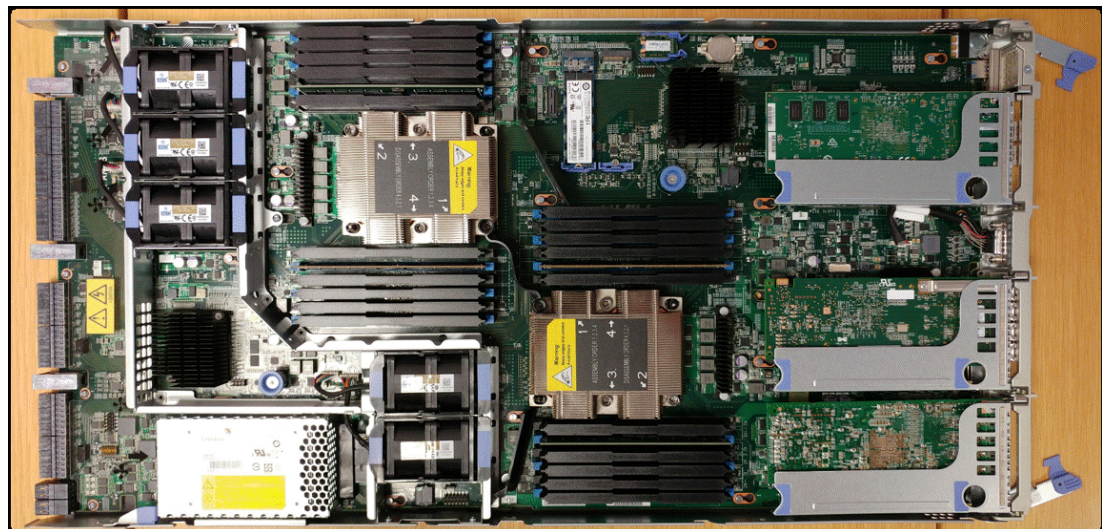


Figure 1-4 Internal hardware components

1.4.2 IBM FlashSystem 9200 Control Enclosure

The IBM FlashSystem 9200 system is a 2U model that can house up to 24 NVMe-capable flash drives of various capacities and be configured with up to 1.5 TB of cache.

An IBM FlashSystem 9200 clustered system can contain up to four IBM FlashSystem 9200 systems and up to 3,040 drives in expansion enclosures. IBM FlashSystem 9200 systems can be clustered with existing Storwize V7000 systems models 524, 624, or 724.

The IBM FlashSystem 9200 Control Enclosure node canisters are configured for active-active redundancy. The node canisters provide a web interface, Secure Shell (SSH) access, and Simple Network Management Protocol (SNMP) connectivity through external Ethernet interfaces. By using the web and SSH interfaces, administrators can monitor system performance and health metrics, configure storage, and collect support data, among other features.

Mixed cluster naming

You need clear rules about the behavior in mixed clusters to ensure that the system agrees on the name of the entire system. The new extended rule is that the new or highest system type overrules anything else in the cluster. For example, if you add an IBM FlashSystem 9200 system to an IBM FlashSystem 9100 system, the system reports itself as an IBM FlashSystem 9200 system.

Here is the explicit order of priority:

IBM FlashSystem 9200 system > IBM FlashSystem 9100 system > IBM FlashSystem 7200 system > Storwize 7000 system

Here are some examples:

- ▶ Add an IBM FlashSystem 7200 I/O group to an existing Storwize V7000 cluster, and now the cluster is an IBM FlashSystem 7200 cluster.
- ▶ If you then add an IBM FlashSystem 9200 system, the cluster is an IBM FlashSystem 9200 system.

IBM FlashSystem 9200 Control Enclosure Model AG8

IBM FlashSystem 9200 Control Enclosure Model AG8 has the following components:

- ▶ Two node canisters, each with four 16-core 2.3 GHz Cascade Lake CPUs with compression assist up to 100 gigabits per second (Gbps)
- ▶ Cache options from 256 GB (128 GB per canister) to 1.5 TB (768 GB per canister)
- ▶ Eight 10 Gb Ethernet (GbE) onboard ports standard for internet Small Computer Systems Interface (iSCSI) connectivity or IP replication
- ▶ Up to three PCIe adapters (see the options below)
- ▶ Twenty-four slots for 2.5-inch NVMe flash drives
- ▶ 2U 19-inch rack mount enclosure with AC power supplies
- ▶ Two boot drives
- ▶ The PCIe adapter options are:
 - Four-port 16 Gb Fibre Channel (FC) / NVMe over Fabrics (NVMe-oF) card
 - Four-port 32 Gb FC / NVMe-oF card
 - Two-port 25 GbE iSCSI / iSCSI Extensions for Remote Direct Memory Access (RDMA) (iSER)/ RDMA over Converged Ethernet (RoCE) card
 - Two-port 25 GbE iSCSI / iSER / internet Wide Area RDMA Protocol (iWARP) card
 - 12 Gb SAS ports for expansion enclosure attachment

IBM FlashSystem 9200 Utility Model UG8

IBM FlashSystem 9200 Utility Model UG8 provides a variable capacity storage offering. These models offer a fixed capacity with a base subscription of 35% of the total capacity.

IBM Storage Insights is responsible for monitoring the system and reporting the capacity that was used beyond the base 35%, which is then billed on the capacity-used basis. You can grow or shrink usage, and pay only for the configured capacity.

The IBM FlashSystem Utility Model is provided for customers who can benefit from a variable capacity system, where billing is based only on actual provisioned space. The hardware is leased through IBM Global Finance on a three-year lease, which entitles the customer to use approximately 30 - 40% of the total system capacity at no additional cost (depends on the individual customer contract). If storage needs increase beyond that initial capacity, usage is billed based on the average daily provisioned capacity per terabyte per month, on a quarterly basis.

Example: Total system capacity of 115 TB

A customer has an IBM FlashSystem 9200 Utility Model with 4.8 TB NVMe drives for a total system capacity of 115 TB. The base subscription for such a system is 40.25 TB. During the months where the average daily usage is below 40.25 TB, there is no additional billing.

The system monitors daily provisioned capacity and averages those daily usage rates over the month term. The result is the average daily usage for the month.

If a customer uses 45 TB, 42.5 TB, and 50 TB in three consecutive months, IBM Storage Insights calculates the overage as shown in Table 1-1, rounding to the nearest terabyte.

Table 1-1 Billing calculations that are based on customer usage

Average daily	Base	Overage	To be billed
45 TB	40.25 TB	4.75 TB	5 TB
42.5 TB	40.25 TB	2.25 TB	2 TB
50 TB	40.25 TB	9.75 TB	10 TB

The total capacity that is billed at the end of the quarter is 17 TB per month in this example.

Flash drive expansions may be ordered with the system in all supported configurations.

Table 1-2 shows the feature codes that are associated with the IBM FlashSystem 9200 Utility Model UG8 billing.

Table 1-2 IBM FlashSystem 9200 Utility Model UG8 billing feature codes

Feature code	Description
#AE00	Variable Usage 1 TB per month
#AE01	Variable Usage 10 TB per month
#AE02	Variable Usage 100 TB per month

These features are used to purchase the variable capacity that is used in the IBM FlashSystem 9200 Utility Models. The features (#AE00, #AE01, and #AE02) provide terabytes of capacity beyond the base subscription on the system. Usage is based on the average capacity that is used per month. The total of the prior three months' usage should be totaled and the corresponding number of #AE00, #AE01, and #AE02 features ordered quarterly.

1.4.3 IBM FlashSystem 9000 Expansion Enclosure Models AFF and A9F

IBM FlashSystem 9000 Expansion Enclosures Models AFF and A9F can be attached to an IBM FlashSystem 9200 Control Enclosure to increase the available capacity. It communicates with the Control Enclosure through a dual pair of 12 Gbps SAS connections. These Expansion Enclosures can house many flash (solid-state drive (SSD)) SAS type drives.

IBM FlashSystem 9000 Expansion Enclosure Model AFF

IBM FlashSystem 9000 Expansion Enclosure Model AFF holds up to twenty-four 2.5-inch SAS flash drives in a 2U 19-inch rack mount enclosure. An intermix of capacity drives is allowed in any drive slot, and up to twenty AFF enclosures can be attached to the control enclosure to a total of 480 drives maximum.

Figure 1-5 shows the front view of the IBM FlashSystem 9000 Expansion Enclosure Model AFF.



Figure 1-5 IBM FlashSystem 9000 Expansion Enclosure Model AFF

IBM FlashSystem 9000 Expansion Enclosure Model A9F

The IBM FlashSystem 9000 Expansion Enclosure Model A9F holds up to ninety-two 3.5-inch SAS flash drives in a 5U 19-inch rack mount enclosure. An intermix of capacity drives is allowed in any drive slot, and up to eight A9F enclosures can be attached to the control enclosure to a total of 736 drives maximum.

Figure 1-6 on page 11 shows the front view of the IBM FlashSystem 9000 Expansion Enclosure Model A9F.

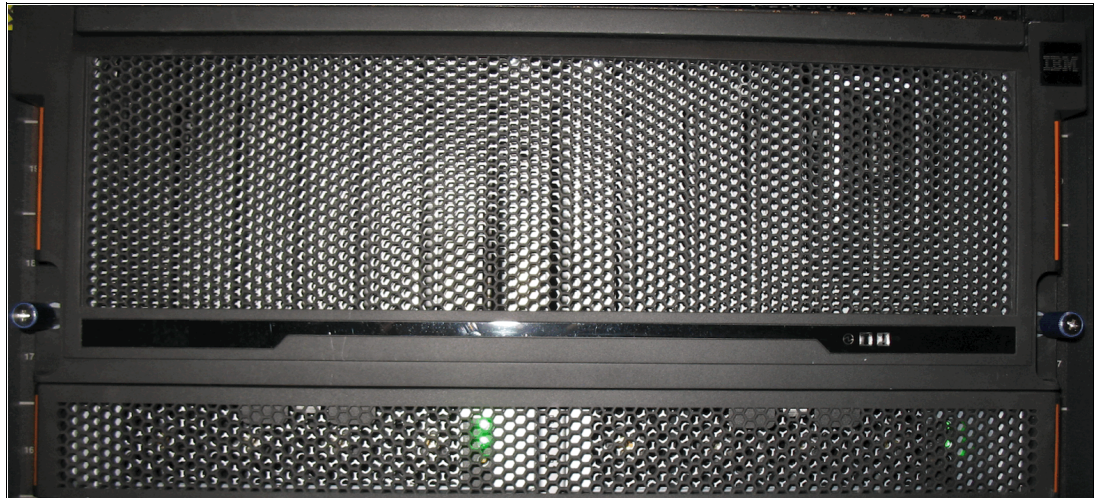


Figure 1-6 IBM FlashSystem 9000 Expansion Enclosure Model front view

SAS chain limitations

When attaching expansion enclosures to the control enclosure, you are not limited by the type of the enclosure. The only limitation for each of the two SAS chains is its *chain weight*. Each type of enclosure has its own chain weight:

- ▶ The IBM FlashSystem 9000 Expansion Enclosure Model AFF has a chain weight of 1.
- ▶ The IBM FlashSystem 9000 Expansion Enclosure Model A9F has a chain weight of 2.5.

The maximum chain weight is 10.

For example, you can combine seven IBM FlashSystem 9000 Expansion Enclosure Model AFF and one IBM FlashSystem 9000 Expansion Enclosure Model A9F expansions ($7 \times 1 + 1 \times 2.5 = 9.5$ chain weight) or two IBM FlashSystem 9000 Expansion Enclosure Model A9F enclosures and five IBM FlashSystem 9000 Expansion Enclosure Model AFF expansions ($2 \times 2.5 + 5 \times 1 = 10$ chain weight).

An example of chain weight 4.5 with two IBM FlashSystem 9000 Expansion Enclosure Model AFF enclosures and one IBM FlashSystem 9000 Expansion Enclosure Model A9F enclosure all correctly cabled is shown in Figure 1-7, which shows an IBM FlashSystem 9200 system connecting through SAS cables to the expansion enclosures while complying with the maximum chain weight.

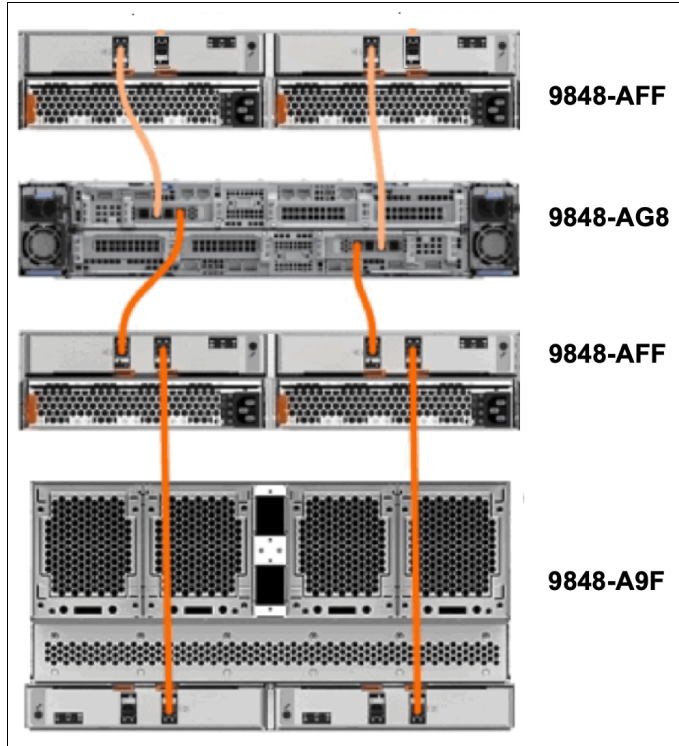


Figure 1-7 IBM FlashSystem 9200 system that is connected to expansion enclosures

1.5 IBM FlashSystem 9200R Rack Solution overview

IBM FlashSystem 9200R is a pre-cabled, pre-configured rack solution that contains multiple IBM FlashSystem 9200 Control Enclosures and uses IBM Spectrum Virtualize to linearly scale performance and capacity through clustering. For more information about this product, see *IBM FlashSystem 9200R Rack Solution Product Guide*, REDP-5593.

The IBM FlashSystem 9200R Rack Solution system has a dedicated FC network for clustering and optional expansion enclosures, which are delivered assembled in a rack. Available with two, three, or four clustered IBM FlashSystem 9200 systems and up to four expansion enclosures, it can be ordered as an IBM FlashSystem 9202R, IBM FlashSystem 9203R, or IBM FlashSystem 9204R system, with the last number denoting the number of AG8 controller enclosures in the rack.

The final configuration occurs on site following the delivery of the systems. More components can be added to the rack after delivery to meet the growing needs of the business.

Note: Other than the IBM FlashSystem 9200 control enclosure and its expansion enclosures, the additional components of this solution are not covered under Enterprise Class Support (ECS). Instead, they have their own warranty, maintenance terms, and conditions.

Rack rules

The IBM FlashSystem 9200R Rack Solution product represents a limited set of possible configurations. Each IBM FlashSystem 9200R Rack Solution order must contain these components:

- ▶ Two, three, or four 9848 Model AG8 Control Enclosures.
- ▶ Two IBM SAN24B-6 or two IBM SAN32C-6 FC switches.
- ▶ Optionally, 0 - 4 9848 Model AFF Expansion Enclosures, with no more than one expansion enclosure per Model AG8 Control Enclosure and no mixing with the 9848 Model A9F Expansion Enclosure.
- ▶ Optionally, 0 - 2 9848 Model A9F Expansion Enclosures, with no more than one expansion enclosure per Model AG8 Control Enclosure and no mixing with 9848 Model A9F Expansion Enclosure.
- ▶ One 7965-S42 rack with the appropriate power distribution units (PDUs) that are required to power components within the rack.
- ▶ All components in the rack must include feature codes #FSRS and #4651.
- ▶ For Model AG8, AFF, and A9F Control Enclosures, the first and largest capacity enclosure includes feature code #AL01, with subsequent enclosures that use #AL02, #AL03, and #AL04 in capacity order. The 9848 Model AG8 Control Enclosure with #AL01 must also have #AL0R included.

Following the initial order, each 9848 Model AG8 Control Enclosures can be upgraded through a miscellaneous equipment specification (MES).

More components can be ordered separately and added to the rack within the configuration limitations of the IBM FlashSystem 9200 system. Clients must ensure that the space, power, and cooling requirements are met. If assistance is needed with the installation of these additional components beyond the service that is provided by your IBM System Services Representative (IBM SSR), IBM Lab Services are available.

Table 1-3 shows the IBM FlashSystem 9200R Rack Solution combinations, the MTMs, and their associated feature codes.

Table 1-3 IBM FlashSystem 9200R Rack Solution combinations

Machine type and model (MTM)	Description	Quantity
7965-S42	IBM Enterprise Slim Rack	1
8960-F24	IBM SAN24B-6 FC switch (Brocade)	2 ^a
8977-T32	IBM SAN32C-6 FC switch (Cisco)	2 ^a
9848-AFF	IBM FlashSystem 9000 2U small form factor (SFF) Expansion Enclosure with 3-year Warranty and ECS	0 - 4 ^b

Machine type and model (MTM)	Description	Quantity
9848-AG8	IBM FlashSystem 9200 Control Enclosure with 3-year Warranty and ECS	2, 3, or 4
9848-A9F	IBM FlashSystem 9000 5U large form factor (LFF) high-density Expansion Enclosure with 3-year Warranty and ECS	0 - 2 ^b

- a. For the FC switch, choose either two of machine type (MT) 8977 or two of MT 8960.
b. For extra expansion enclosures, choose either model AFF, model A9F, or none. You cannot use both.

IBM FlashSystem 9200R Rack Solution configurations: Rack diagrams

This section shows the rack diagrams that show the minimum and maximum IBM FlashSystem 9200R Rack Solution configurations with both A9F and AFF Expansion Enclosures.

Key to figures

The key to the symbols that are used in the figures in this section are shown in Table 1-4.

Table 1-4 Key to the symbols that are used in the figures

Label	Description
CTL n	9848-AG8 Control Enclosure number n of 4. CTL1 and CTL2 are required. CTL3 and CTL4 are optional.
EXP n	9848 Expansion Enclosure number n . Optional.
AFF EXP n	9848-AFF 2U Expansion Enclosure number n of 4.
A9F EXP n	9848-A9F 5U Expansion Enclosure number n of 2.
FC SW n	FC switch n of 2. These switches are either both 8977-T32 or they are both 8960-F24.
PDU A, PDU B	PDUs. Both have the same rack feature code: #ECJJ, #ECJL, #ECJN, or 3ECJQ.

Figure 1-8 shows the legend that is used to denote the component placement and mandatory gaps for the figures that show the configurations.

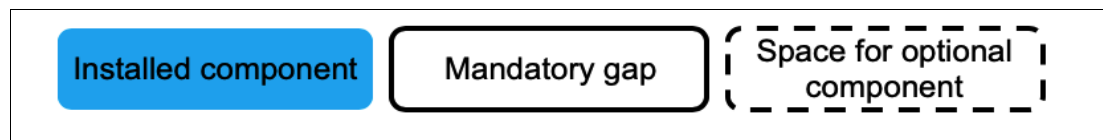


Figure 1-8 Legend to figures in this section

1.5.1 Minimum IBM FlashSystem 9200R Rack Solution configuration in the rack

Figure 1-9 shows the minimum IBM FlashSystem 9200R Rack Solution configuration in the rack.

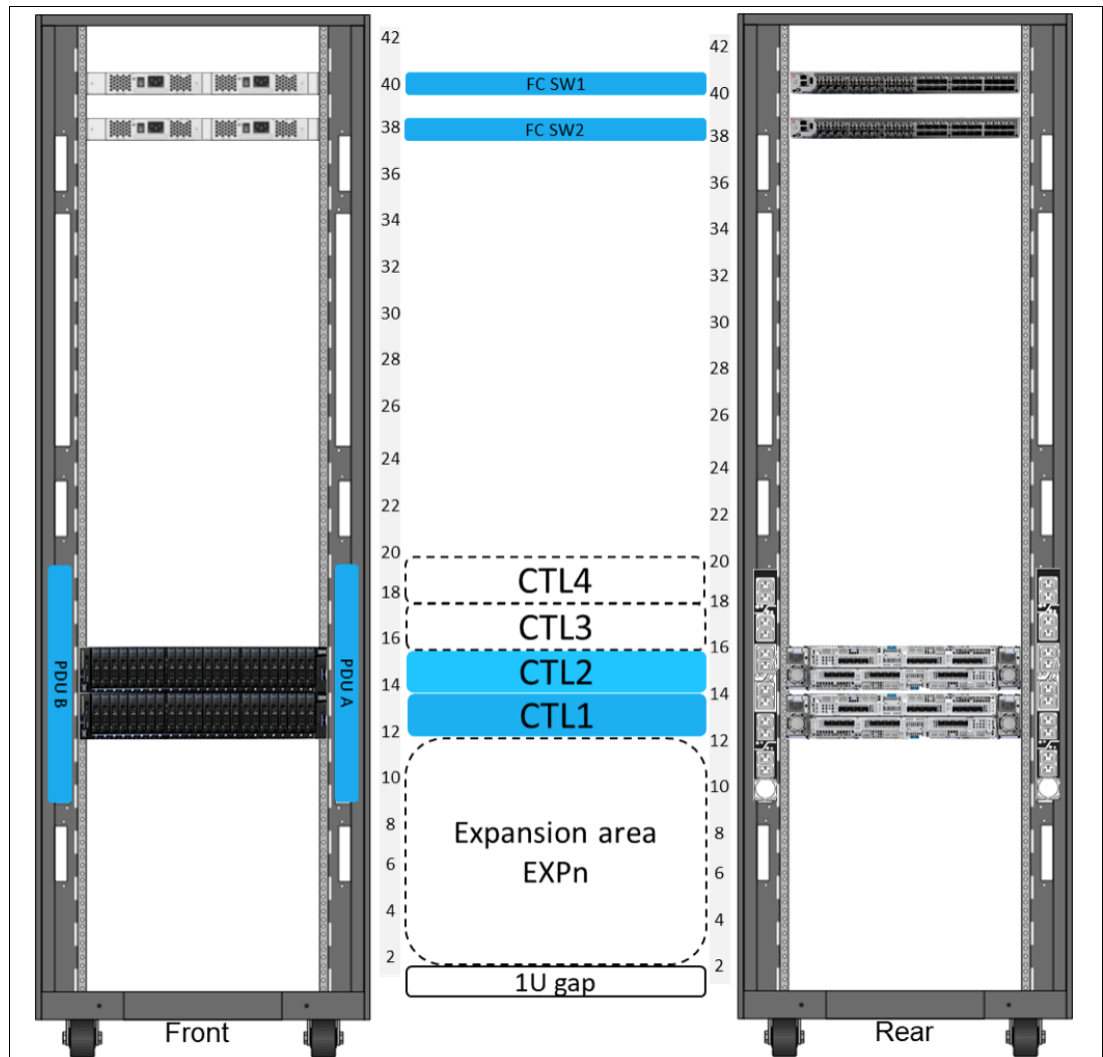


Figure 1-9 Minimum IBM FlashSystem 9200R Rack Solution configuration in the rack

Notes about the minimum configuration

- ▶ Control Enclosures (CTL) 1 and 2 are mandatory:
 - In adapter slots 1 and 2 are 32 G or 16 G FC adapters, 25 G Ethernet adapters, or a mix of both (for a mix, insert the FC adapter into slot 1).
 - Adapter slot 2 can be blank.
 - Adapter slot 3 is either an SAS adapter (it is required if CTLn is attached to EXPn), a choice of FC or 25 G Ethernet adapters, or blank.
- ▶ The product comes with cables that are appropriate for inter-system FC connectivity. You must order extra cables for host and Ethernet connectivity.
- ▶ The rack has either (0 - 0.2) A9F Expansion Enclosures or (0 - 4) AFF Expansion Enclosures.

- ▶ The PDUs and power cabling that are needed depends on what expansion enclosures are ordered:
 - For A9F configurations, a PDU with nine C19 outlets is required. This PDU also has three C13 outlets on the forward-facing side.
 - For other configurations, the PDU with 12 C13 outlets is selected by default.
- ▶ FC SW1 and FC SW2 are a pair of IBM SAN32C-6 or IBM SAN24B-6 FC switches.
- ▶ You may allocate different amounts of storage (drives) to each CTL and expansion components.
- ▶ A gap of 1U is maintained below the expansion area to allow for power cabling routing.

1.5.2 Maximum configuration of an IBM FlashSystem 9200R Rack Solution with Model A9F Expansion Enclosures

Figure 1-10 shows the maximum configuration of an IBM FlashSystem 9200R Rack Solution with Model A9F Expansion Enclosures.

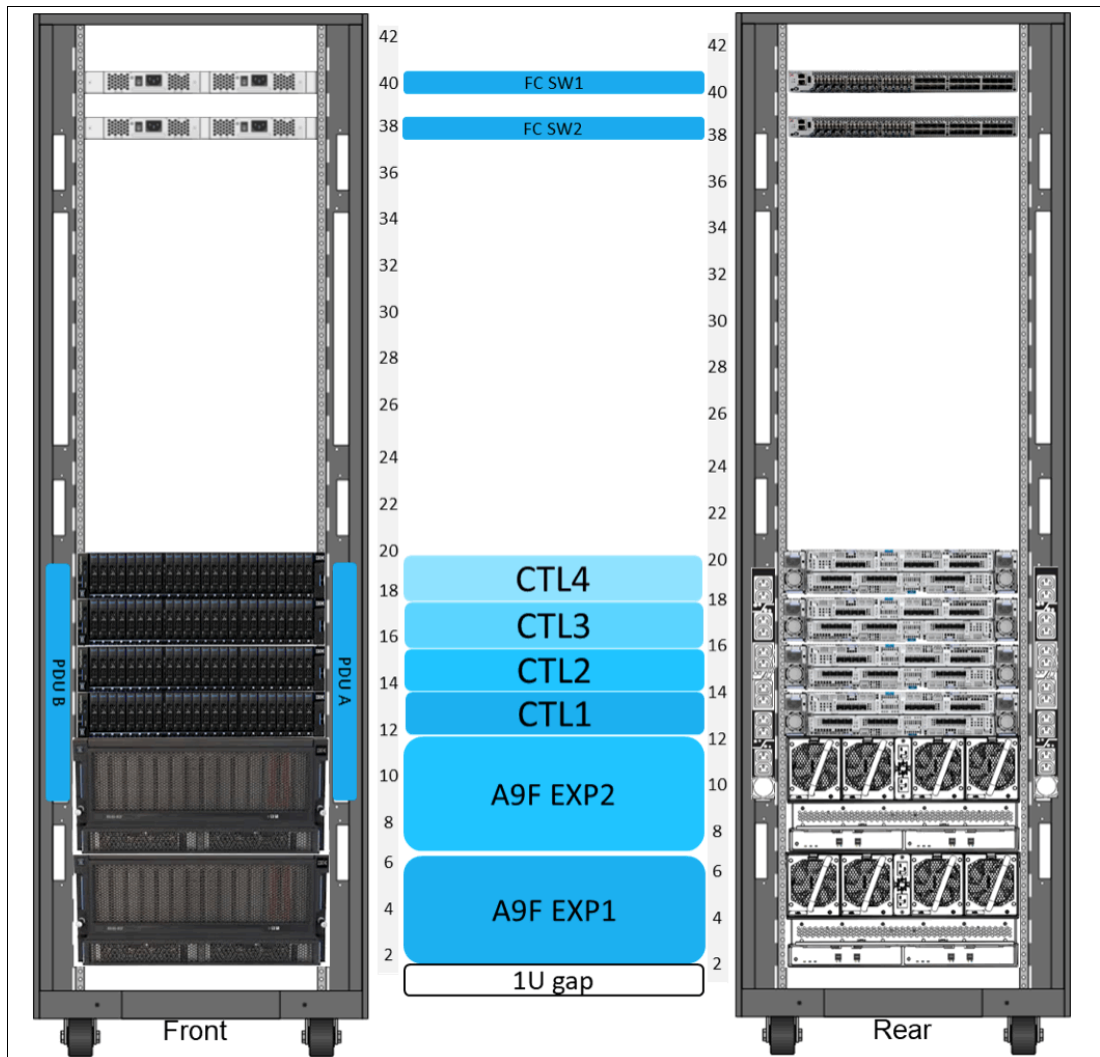


Figure 1-10 Maximum configuration of an IBM FlashSystem 9200R Rack Solution with Model A9F Expansion Enclosures

Notes about the maximum configuration with Model A9F Expansion Enclosures

- ▶ A PDU with nine C19 rear outlets and three C13 front outlets is required.
- ▶ The product comes with cables that are appropriate for inter-system FC connectivity. You must order extra cables for host and Ethernet connectivity.
- ▶ Any Model A9F Expansion Enclosures are installed in U2 - U6 and then U7-1.
- ▶ The CTLs and EXPs are stacked and cabled to the PDU power, with the highest capacity at the bottom. You go upwards, with EXPn attached to CTLn in a bottom-up order by using an SAS adapter on CTLn and cables.

1.5.3 Maximum configuration of an IBM FlashSystem 9200R Rack Solution with Model AFF Expansion Enclosures

Figure 1-11 shows the maximum configuration of an IBM FlashSystem 9200R Rack Solution with Model AFF Expansion Enclosures.

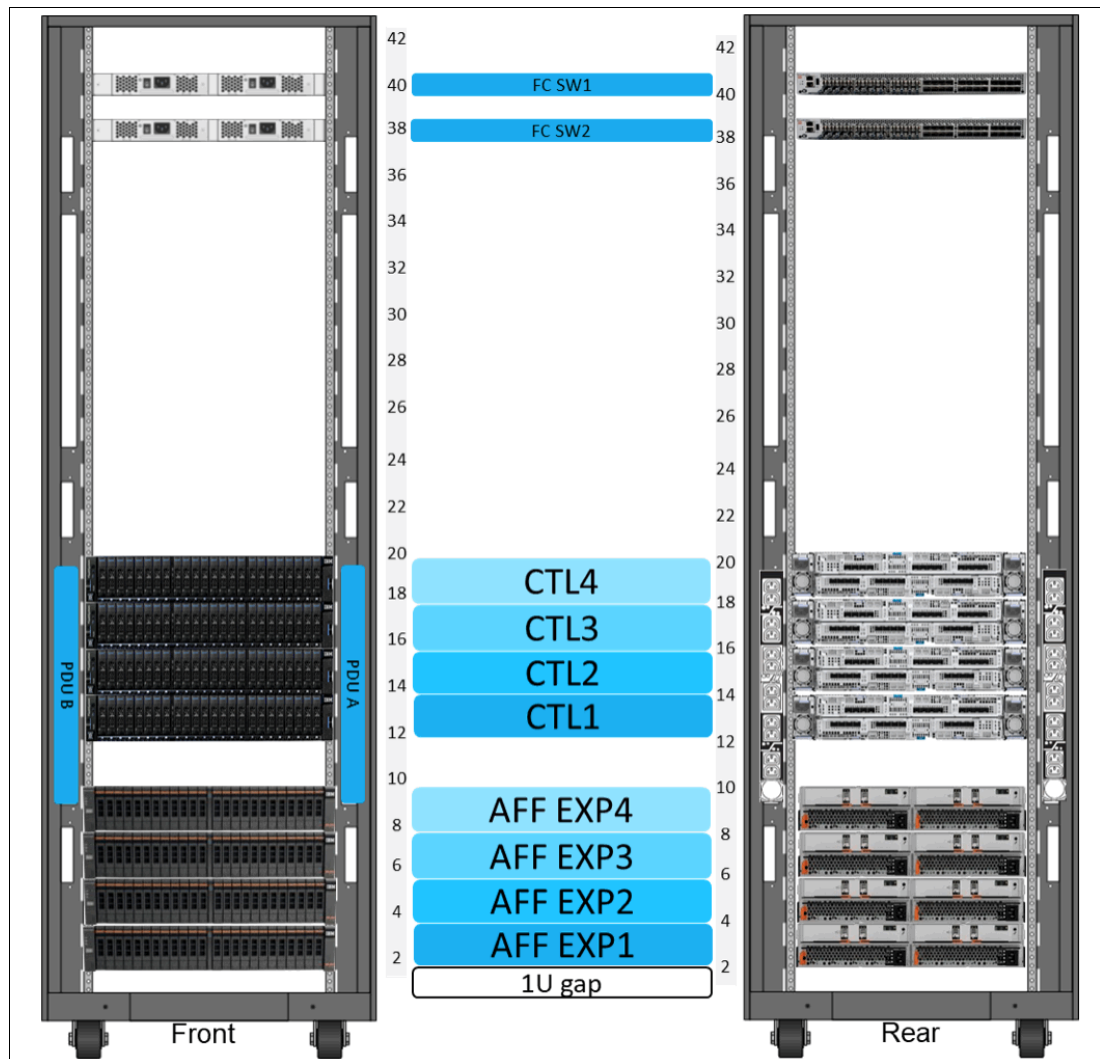


Figure 1-11 Maximum configuration of an IBM FlashSystem 9200R Rack Solution with Model AFF Expansion Enclosures

Notes about the maximum configuration with Model AFF Expansion Enclosures

- ▶ The 12x C13 PDU is selected on order.
- ▶ The product comes with cables that are appropriate for inter-system FC connectivity. You must order extra cables for host and Ethernet connectivity.
- ▶ AG8 1 and AG8 2 are mandatory. From there, AG8 3 and AG8 4 can be optionally and incrementally added.
- ▶ Adapter slot 1 of AG8 is dedicated to 32 Gb clustering usage.
- ▶ Adapter slot 2 of AG8 is used for your choice of a host adapter.
- ▶ Adapter slot 3 is an SAS adapter if it is required, or one of your choice if it is not.

1.5.4 FC cabling and clustering

The IBM FlashSystem 9200R Rack Solution has specialized internal cabling that is supplied by the manufacturing plant before the machine is shipped to the customer. The plugging of the inter-system internal FC cables is done on site at installation time.

Figure 1-12 shows the FC cabling at the rear of the IBM FlashSystem 9200R Control Enclosure.

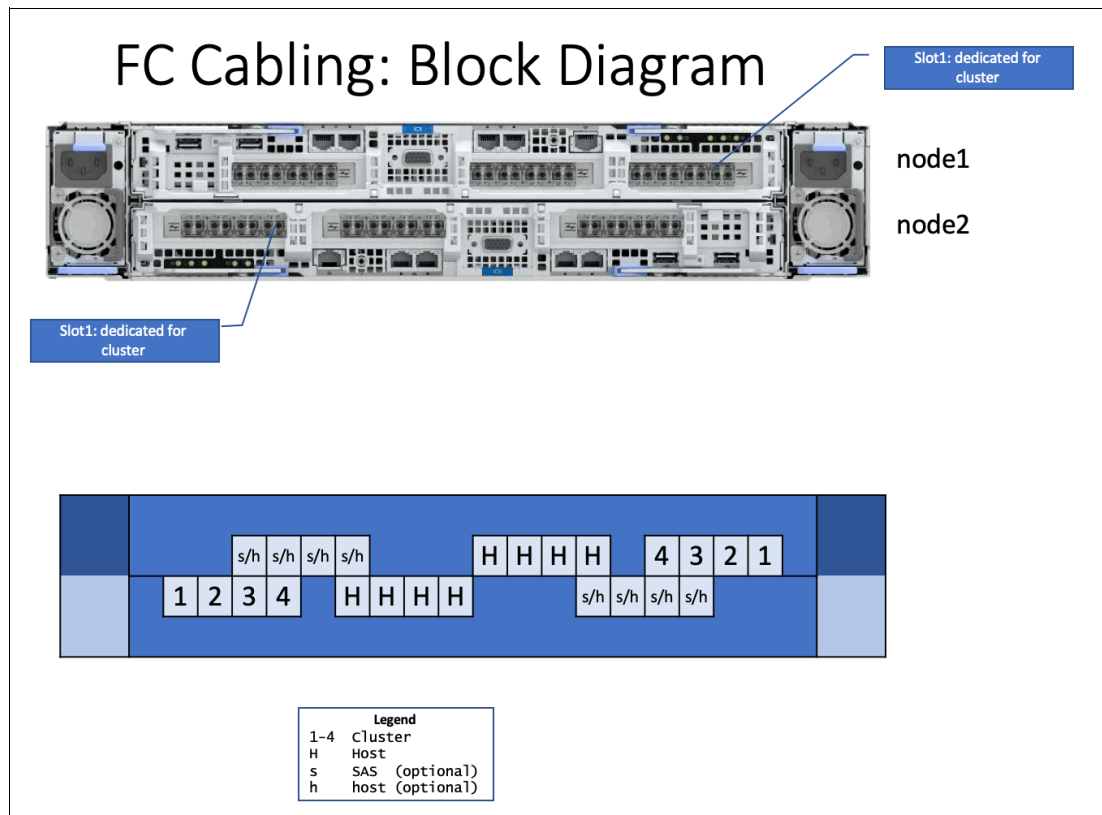


Figure 1-12 FC cabling at the rear of the IBM FlashSystem 9200R Control Enclosure

Notes on FC cabling and clustering

- ▶ In a control enclosure, the upper node canister is upside down, so port composition is reversed compared to the lower node canister.
- ▶ Adapter slot 1 is dedicated to clustering.
- ▶ Adapter slot 2 is a host connectivity adapter choice, for example, 32 Gb FC or 25 GbE.
- ▶ Adapter slot 3 is SAS for hybrid, or it is a host connectivity adapter choice.

Note: If there are multiple adapters, install the 32 G FC adapter first, then the 16 G FC adapter, and then the 25 G Ethernet adapter.

From the top image, this “block diagram” depicts the rear composition of the IBM FlashSystem 9200 system. It shows a simple composition to draw attention to the ports for cabling.

- The upper canister (for example, node1) is numbered right to left.
- The lower canister (for example, node2) is numbered left to right.
- Numbers 1, 2, 3, and 4 are used to denote inter-cluster cabling. The items for CE, IBM SSR, and LBS to refer to the cabling for the cluster switch.
- *H* depicts host-facing ports, which are a customer responsibility and a required selection (otherwise, the hosts cannot use the storage).
- *s/h* is for attaching optional SAS expansion enclosures or more SAS hosts. The ones with the lowercase *h* are an optional choice.
- Where a SAS adapter is not installed, use slot 3 for optional extra host-facing ports. The *h* means that they are optional.

Figure 1-13 shows up to four IBM FlashSystem 9200 Control Enclosures and the port notation for the inter-cluster connections.

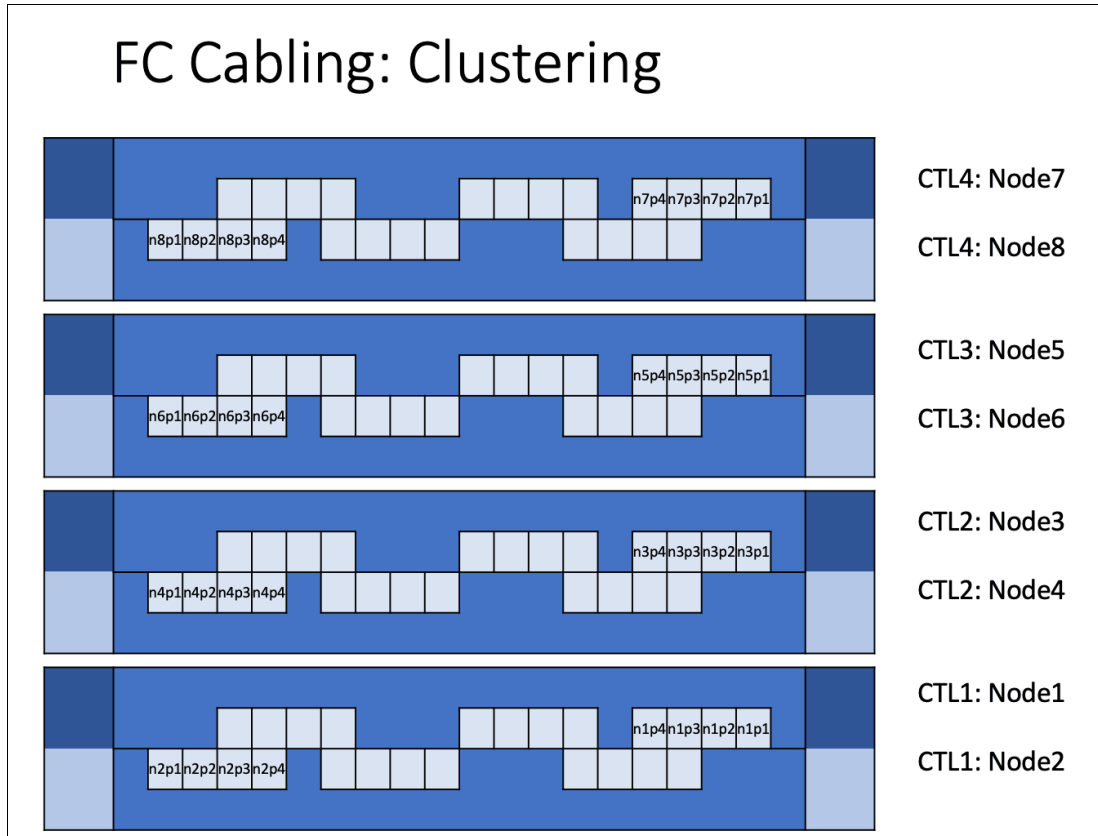


Figure 1-13 IBM FlashSystem 9200R Rack Solution inter-cluster connections

Figure 1-13 through Figure 1-17 on page 24 shows the numeric cabling for clustering:

- ▶ CTL1 - CTL4 represent the relative rack position of 1 - 4 (min - max) IBM FlashSystem 9200 Control Enclosures within the rack.
- ▶ To denote the cable ports:
 - N1P1 represents Node1 port 1, which is the farthest right port of the upper node canister.
 - N2P1 represents Node2 port 1, which is the lower node canister, farthest left port.

1.5.5 IBM FlashSystem 9200R Rack Solution FC configuration with IBM SAN32C-6 switches

Figure 1-14 on page 21 shows the IBM FlashSystem 9200R Rack Solution inter-system FC ports and the connections to the IBM SAN32C-6 switch ports.

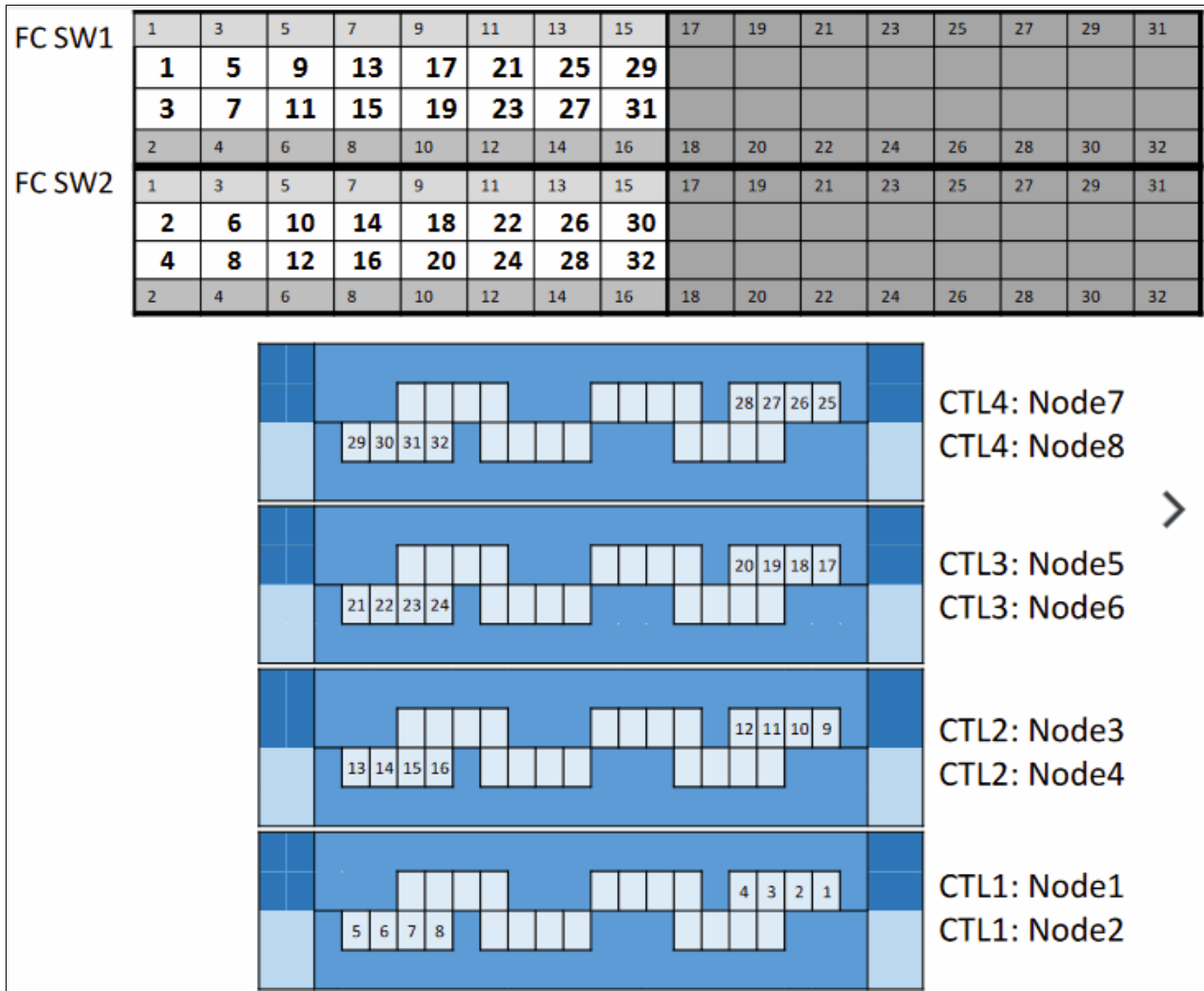


Figure 1-14 IBM FlashSystem 9200R Rack Solution FC with IBM SAN32C-6 switches

Notes about IBM FlashSystem 9200R Rack Solution FC with IBM SAN32C-6 switches

- ▶ Shows the SAN32C-6 cabling diagram, where the top part of Figure 1-14 represents the SAN32C-6 port configuration layout.
- ▶ “1” is a cable number, which starts from CTL1 node2 port1 and goes to the top SAN32C-6 switch (SW1) port 1.
- ▶ In a minimal order of an IBM FlashSystem 9200R Rack Solution product, you order two control enclosures (CTL1 and CTL2). Optionally, you may order (with the original order or as an MES later) the third and fourth control enclosures.

1.5.6 IBM FlashSystem 9200R Rack Solution FC configuration with IBM SAN24B-6 switches

Figure 1-15 shows the IBM FlashSystem 9200 Rack Solution inter-system FC ports and the connections to the IBM SAN24B-6 switch ports.

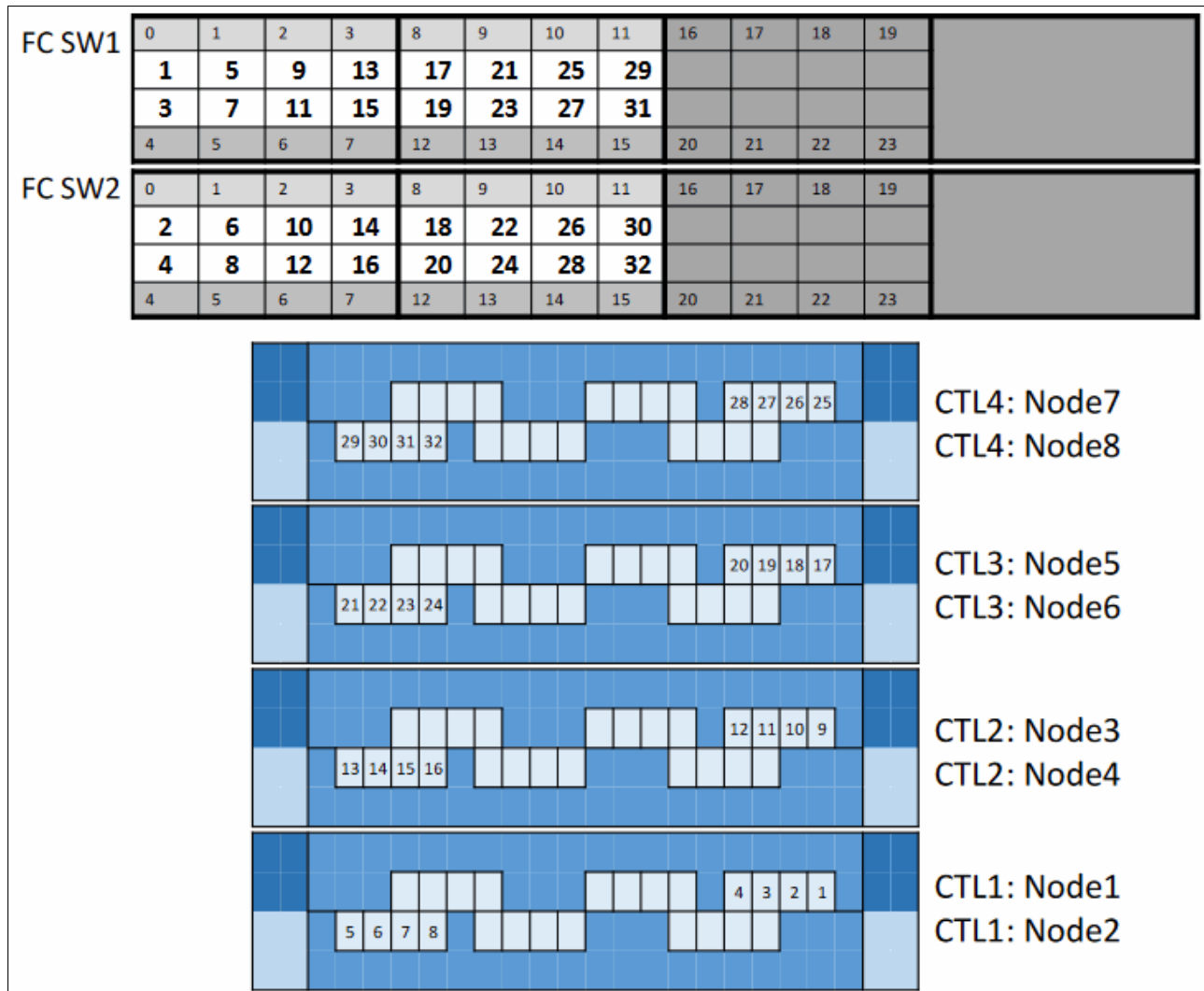


Figure 1-15 IBM FlashSystem 9200R Rack Solution with IBM SAN24B-6 switches

Notes about IBM FlashSystem 9200R Rack Solution FC configuration with IBM SAN24B-6 switches

- ▶ Shows the SAN24B-6 cabling diagram, where the top part of Figure 1-15 represents the SAN24B-6 port configuration layout.
- ▶ “1” is cable number, which starts from CTL1 node2 port1, and goes to the top SAN24B-6 switch (SW1) port 0.
- ▶ In a minimal order of an IBM FlashSystem 9200R Rack Solution product, you order two control enclosures (CTL1 and CTL2). Optionally, you may order (with the original order or as an MES later) the third and fourth control enclosures.

1.5.7 IBM FlashSystem 9200R Rack Solution SAS Expansion Enclosures cabling

Figure 1-16 shows the SAS port connections for both the Model AFF and Model A9F Expansion Enclosures.

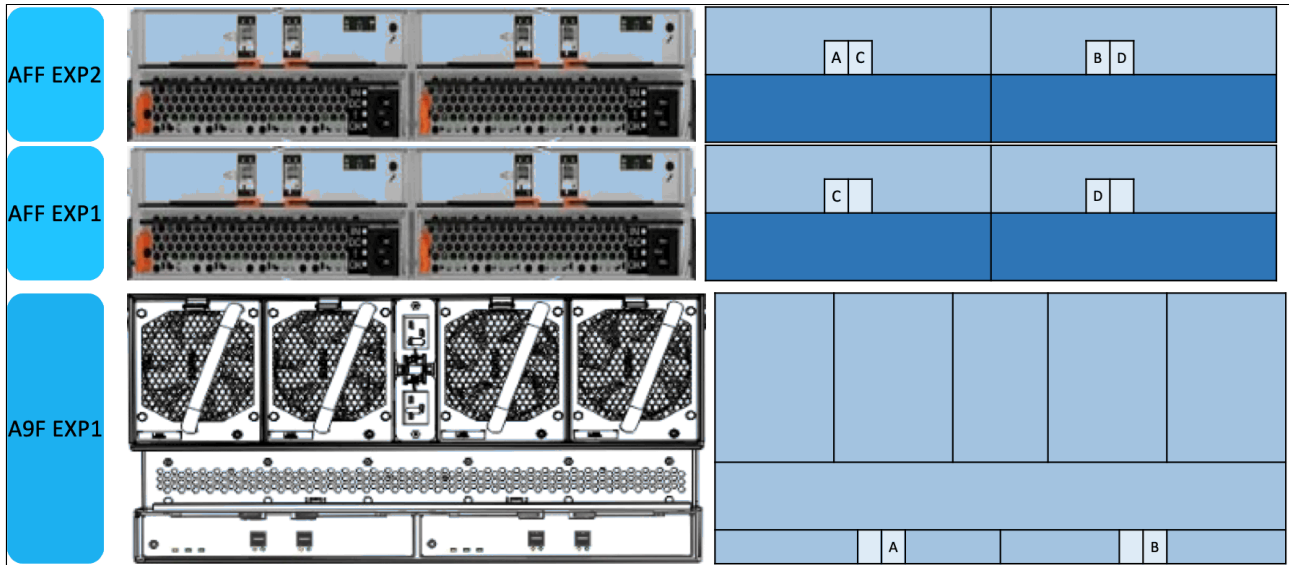


Figure 1-16 IBM FlashSystem 9200R Rack Solution Model AFF and Model A9F SAS Expansion Enclosure ports

Ports A - H refer to the connections that are made to the IBM FlashSystem 9200 Control Enclosures.

Figure 1-17 shows the cabling matrix from the IBM FlashSystem 9200 Control Enclosures and the A9F and AFF Expansion Enclosures. These SAS cabling connections are performed by the IBM SSR at installation time.

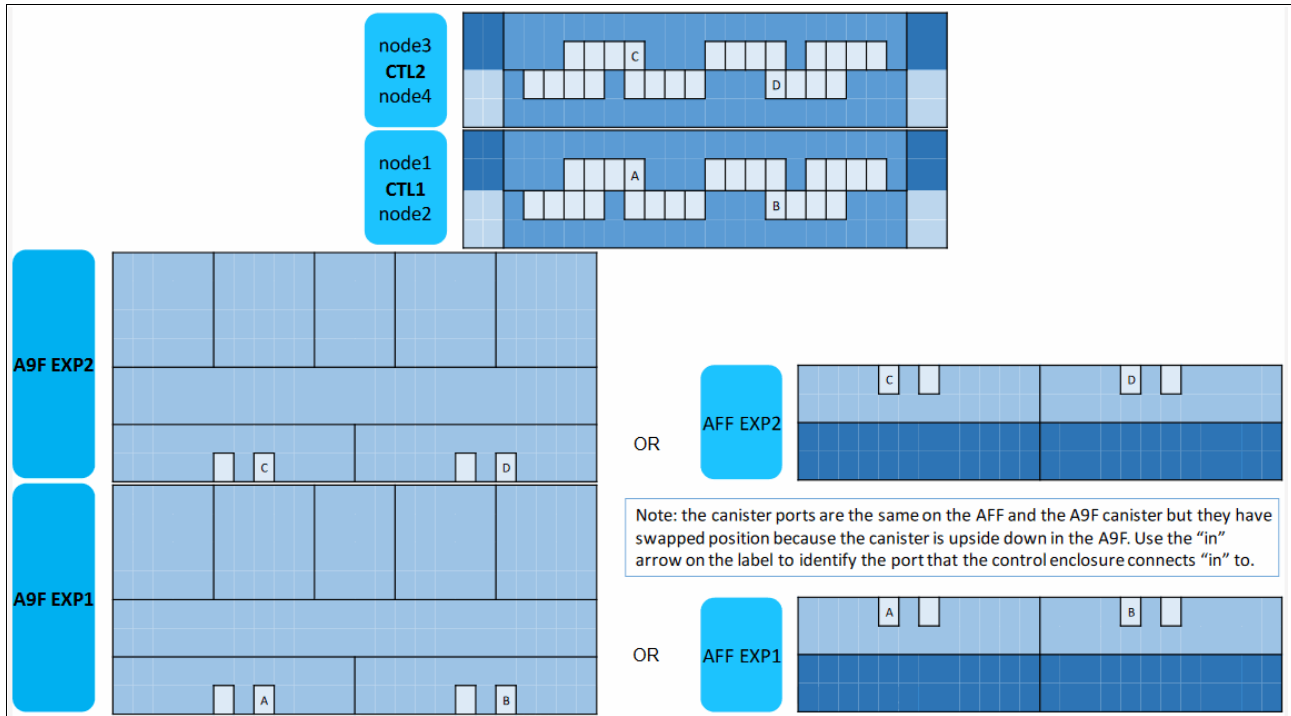


Figure 1-17 IBM FlashSystem 9200 SAS cabling matrix

- ▶ For the Model A9F expansion cabling:
 - Cable A, depicted as “A”, connects CTL1 Node1 port1 to EXP1’s canister 1, port 1.
 - Cable B, depicted as “B”, connects CTL1 Node2 port1 to EXP1’s canister 2, port 1.
- ▶ For the Model AFF expansion cabling:
 - Cable C, depicted as “C”, connects CTL2 Node3 port1 to EXP2’s canister 1, port 1.
 - Cable D, depicted as “D”, connects CTL2 Node4 port1 to EXP2’s canister 2, port 1.

If required, use the same pattern to connect CTL3 to EXP3 and CTL4 to EXP4.

Because you can choose whether the EXP1 is either the Model A9F or the Model AFF, the cable patterns are relatively the same, with the diagrams on the left showing the Model A9F and the diagrams on the right showing the Model AFF.

1.6 IBM FlashSystem 7200 overview

Each IBM FlashSystem 7200 system consists of a control enclosure and NVMe-attached flash drives. The control enclosure is the storage server that runs the IBM Spectrum Virtualize software that controls and provides features to store and manage data.

Figure 1-18 shows the front and rear views of the IBM FlashSystem 7200 system.

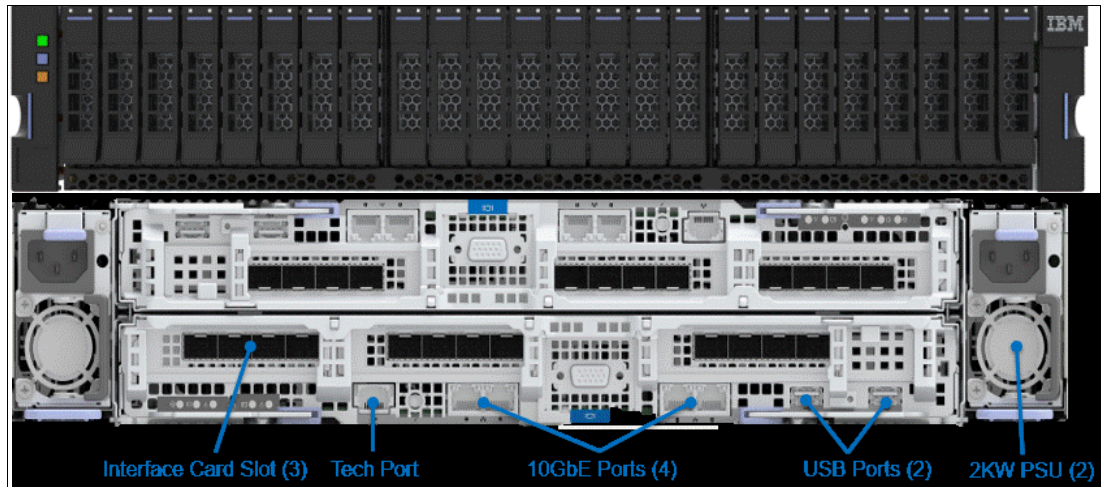


Figure 1-18 IBM FlashSystem 7200 front and rear views

Here are the core IBM FlashSystem 7200 components:

- ▶ IBM FlashSystem 7200 Control Enclosure:
 - PSUs
 - Node canisters
 - Battery modules
 - Fan modules
 - Interface cards
 - Cascade Lake CPUs and memory slots
- ▶ NVMe drives
- ▶ IBM FlashSystem 7000 Expansion Enclosures (SAS-attached)

Note: The IBM FlashSystem 7200 is also available with the optional purchase of the ECS, which gives enhanced customer service response times, the services of an IBM Technical Advisor, and IBM applied code that is purged through the Remote Code Load process.

As shown in Figure 1-18, the IBM FlashSystem 7200 enclosure consists of redundant PSUs, node canisters, and fan modules to provide redundancy and HA.

Figure 1-19 shows the IBM FlashSystem 7200 internal architecture.

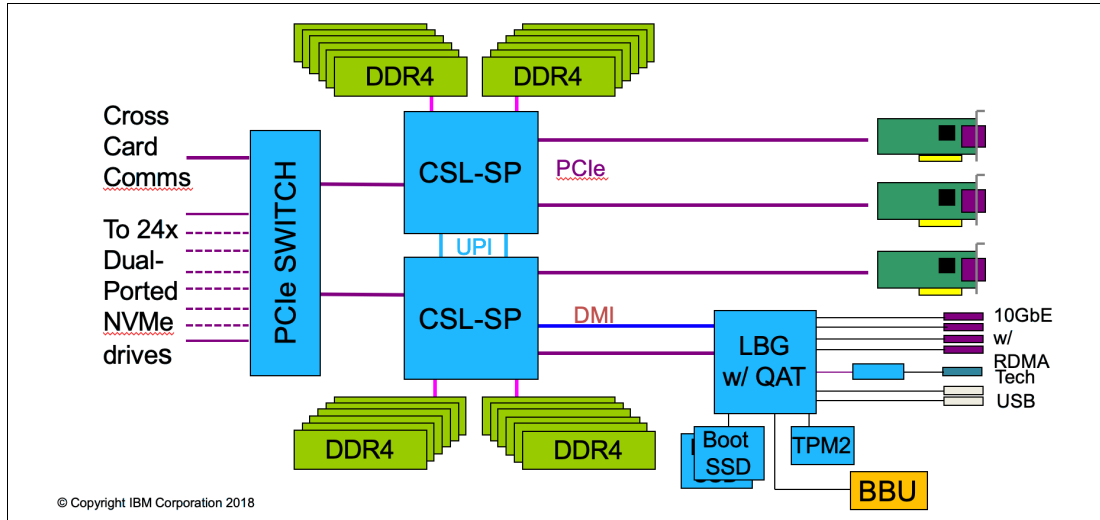


Figure 1-19 IBM FlashSystem 7200 internal architecture

Figure 1-20 shows a picture of the internal hardware components of a node canister. To the left of the picture is the front of the canister where fan modules and battery backup are, followed by two Cascade Lake CPUs and Dual Inline Memory Module (DIMM) slots and PCIe risers for adapters on the right.

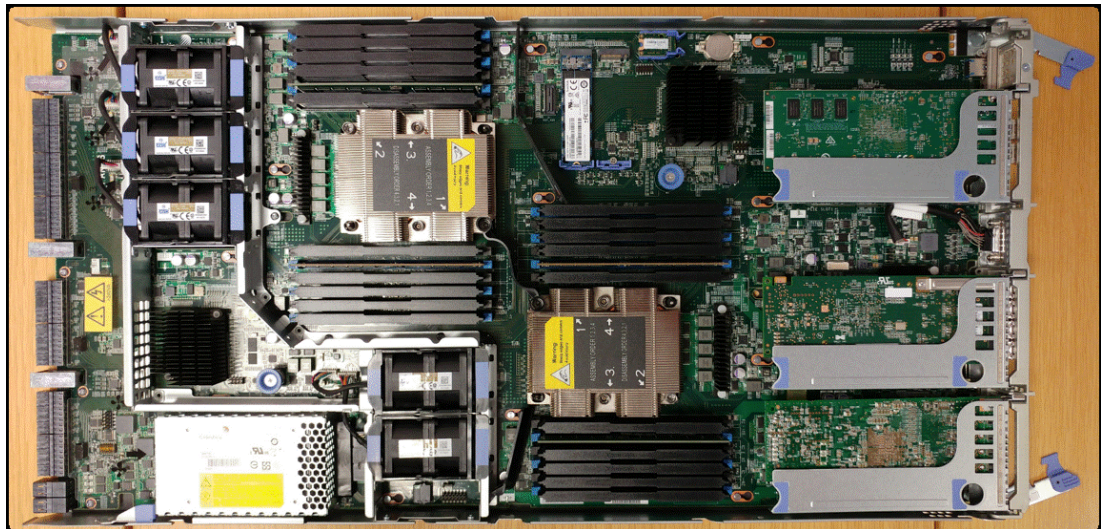


Figure 1-20 Internal hardware components

1.6.1 IBM FlashSystem 7200 Control Enclosure

IBM FlashSystem 7200 is a 2U model that can support up to 24 NVMe drives, either FCM drives with hardware compression and encryption or industry-standard NVMe drives of various capacities or even SCM drives. IBM FlashSystem 7200 can be configured with up to 1.5 TB of cache.

For more information about the drive types that are supported, see 1.14, “IBM FlashCore Module drives, NVMe SSDs, and SCM drives” on page 57.

An IBM FlashSystem 7200 clustered system can contain up to four IBM FlashSystem 7200 systems and up to 3,040 drives. IBM FlashSystem 7200 systems can be added into existing clustered systems that include Storwize V7000 systems.

Mixed cluster naming

You must understand the behavior in mixed clusters to ensure that the system agrees on the name of the entire system. This section is an extension of the work that was already done in Version 8.2.0 for licensing to allow IBM FlashSystem 9100 and 9200 products to cluster with the V7000 system. The new extended rule is that the newest / highest system type overrules anything else in the cluster. For example, if you add an IBM FlashSystem 7200 system to an IBM FlashSystem 9100 or Storwize V7000 system, the system reports itself as an IBM FlashSystem 9200 system.

Here is the explicit order of priority:

IBM FlashSystem 9200 > IBM FlashSystem 9100 > IBM FlashSystem 7200 > Storwize 7000

Here are some examples of the order of priority:

- ▶ In an existing Storwize V7000 cluster, you add an IBM FlashSystem 7200 I/O group. The cluster is now an IBM FlashSystem 7200 system.
- ▶ If you then add an IBM FlashSystem 9100 system, the cluster is an IBM FlashSystem 9100 system.
- ▶ If you then add an IBM FlashSystem 9200 system, the cluster is an IBM FlashSystem 9200 system.

IBM FlashSystem 7200 Model 824 system

The Model 824 system offers the following physical features:

- ▶ Two node canisters with four x8 cores 2.1 GHz Cascade Lake CPUs with compression assist up to 40 Gbps
- ▶ Cache options from 256 GB (128 GB per canister) to 1.5 TB (768 GB per canister)
- ▶ Eight 10 GbE on board ports standard for iSCSI connectivity or IP replication
- ▶ Up to three PCIe adapters (see options below)
- ▶ Twenty-four slots for 2.5-inch NVMe flash drives
- ▶ 2U 19-inch rack mount enclosure with AC power supplies
- ▶ One boot drive
- ▶ The PCIe adapter options are:
 - Four-port 16 Gb FC / NVMe-oF card
 - Four-port 32 Gb FC / NVMe-oF card
 - Two-port 25 GbE iSCSI / iSER / RoCE card
 - Two-port 25 GbE iSCSI / iSER / iWARP card
 - 12 Gb SAS ports for expansion enclosure attachment

1.6.2 IBM FlashSystem 7200 Expansion Enclosures 12G, 24G, and 92G

The following types of expansion enclosures are available:

- ▶ IBM FlashSystem 7200 LFF Expansion Enclosure Model 12G
- ▶ IBM FlashSystem 7200 SFF Expansion Enclosure Model 24G
- ▶ IBM FlashSystem 7200 LFF Expansion Enclosure Model 92G

The IBM FlashSystem 7200 LFF 12G Expansion Enclosure includes the following components:

- ▶ Two expansion canisters
- ▶ 12 Gb SAS ports for control enclosure and expansion enclosure attachment
- ▶ A total of 12 slots for 3.5-inch SAS drives
- ▶ 2U 19-inch rack-mounted enclosure with AC power supplies

Figure 1-21 shows a Model 12G front view.



Figure 1-21 IBM FlashSystem 7200 LFF Expansion Enclosure Model 12G

IBM FlashSystem 7200 SFF Expansion Enclosure Model 24G includes the following components:

- ▶ Two expansion canisters
- ▶ 12 Gb SAS ports for control enclosure and expansion enclosure attachment
- ▶ A total of 24 slots for 2.5-inch SAS drives
- ▶ 2U 19-inch rack mount enclosure with AC power supplies

The SFF Expansion Enclosure is a 2U enclosure that includes the following components:

- ▶ A total of twenty-four 2.5-inch drives (hard disk drives (HDDs) or SSDs).
- ▶ Two Storage Bridge Bay (SBB)-compliant Enclosure Services Manager (ESM) canisters.
- ▶ Two fan assemblies, which mount between the drive midplane and the node canisters. Each fan module is removable when the node canister is removed.
- ▶ Two power supplies.
- ▶ An RS232 port on the back panel (3.5 mm stereo jack), which is used for configuration during manufacturing.

Figure 1-22 shows the front of an SFF Expansion Enclosure.

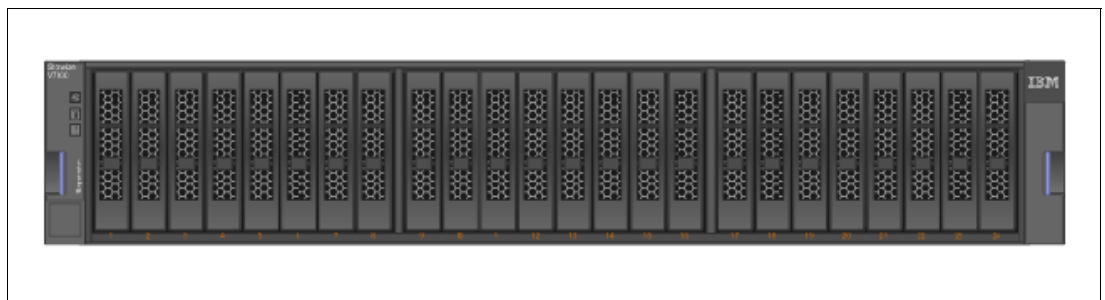


Figure 1-22 Front view of an IBM FlashSystem 7200 SFF Expansion Enclosure

Figure 1-23 on page 29 shows the rear view of the expansion enclosure.

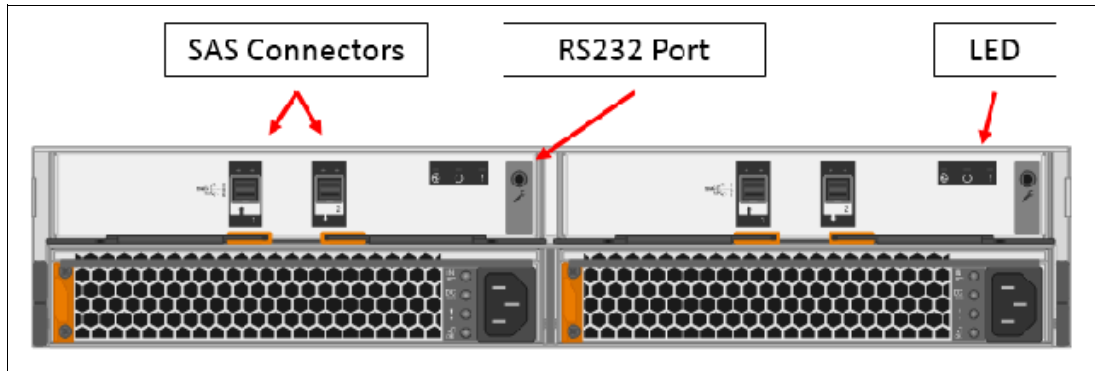


Figure 1-23 Rear of an IBM FlashSystem 7200 Expansion Enclosure

Dense Expansion Enclosure 92G

Dense expansion drawers, or dense drawers, are disk expansion enclosures that are 5U rack-mounted. Each chassis features two expansion canisters, two power supplies, two expander modules, and a total of four fan modules.

Each dense drawer can hold up to 92 drives that are positioned in four rows of 14 and another three rows of 12 mounted drives assemblies. Two Secondary Expander Modules (SEMs) are centrally located in the chassis. One Secondary Expander Module (SEM) addresses 54 drive ports, and the other addresses 38 drive ports.

The drive slots are numbered 1 - 14, starting from the left rear slot and working from left to right, back to front.

Each canister in the dense drawer chassis features two SAS ports numbered 1 and 2. The use of SAS port1 is mandatory because the expansion enclosure must be attached to an IBM FlashSystem 7200 node or another expansion enclosure. SAS connector 2 is optional because it is used to attach to more expansion enclosures.

Each IBM FlashSystem 7200 system can support up to four dense drawers per SAS chain.

Figure 1-24 shows a dense expansion drawer.



Figure 1-24 IBM Dense Expansion Drawer

SAS chain limitations

When attaching expansion enclosures to the control enclosure, you are not limited by the type of the enclosure. The only limitation for each of the two SAS chain is its *chain weight*. Each type of enclosure has its own chain weight:

- ▶ Enclosures 12G and 24G have a chain weight of 1.
- ▶ Enclosure 92G has a chain weight of 2.5.

The maximum chain weight is 10.

For example, you can combine seven 24G and one 92G expansions ($7 \times 1 + 1 \times 2.5 = 9.5$ chain weight), or two 92G enclosures, one 12G, and four 24G ($2 \times 2.5 + 1 \times 1 + 4 \times 1 = 10$ chain weight).

An example of chain weight 4.5 with one 24G, one 12G, and one 92G enclosures, all correctly cabled, is shown in Figure 1-25.

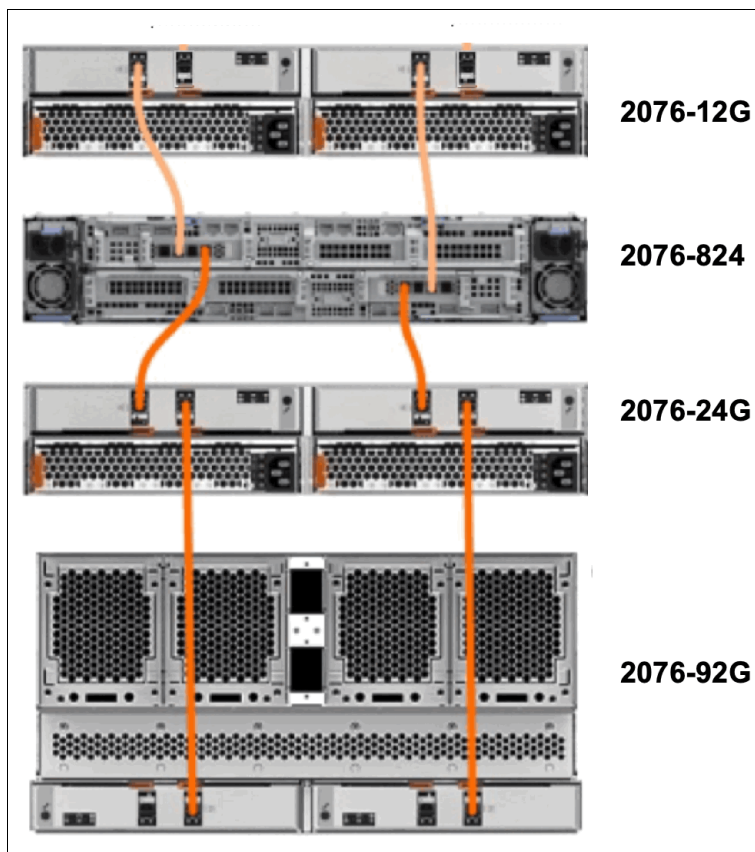


Figure 1-25 Connecting SAS cables while complying with the maximum chain weight

1.6.3 IBM FlashSystem 7200 Utility Model U7C

IBM FlashSystem 7200 Utility Model U7C provides a variable capacity storage offering. These models offer a fixed capacity with a base subscription of 35% of the total capacity.

IBM Storage Insights is responsible for monitoring the system and reporting the capacity that was used beyond the base 35%, which is then billed on the capacity-used basis. You can grow or shrink usage, and pay only for the configured capacity.

IBM FlashSystem Utility Model is provided for customers who can benefit from a variable capacity system, where billing is based only on actual provisioned space. The hardware is leased through IBM Global Finance on a three-year lease, which entitles the customer to use approximately 30 - 40% of the total system capacity at no additional cost (depends on the individual customer contract). If storage needs increase beyond that initial capacity, usage is billed based on the average daily provisioned capacity per terabyte per month, on a quarterly basis.

For an example of Utility Model billing, see “Example: Total system capacity of 115 TB” on page 9.

1.7 IBM FlashSystem 5200 overview

With IBM FlashSystem 5200, you can be ready for a technology transformation without sacrificing performance, quality, or security while simplifying your data management. This powerful and compact solution is focused on affordability with a wide range of enterprise-grade features of IBM Spectrum Virtualize that can easily evolve and extend as businesses grows. This system also has the flexibility and performance of flash and Non-Volatile Memory Express (NVMe) end-to-end, the innovation of IBM FlashCore technology, and Storage Class Memory (SCM) to help accelerate your business execution.

The innovative IBM FlashSystem family is based on a common storage software platform, IBM Spectrum Virtualize, that provides powerful all-flash and hybrid-flash solutions that offer feature-rich, cost-effective, and enterprise-grade storage solutions. Its industry-leading capabilities include a wide range of data services that can be extended to more than 500 heterogeneous storage systems: automated data movement, synchronous and asynchronous copy services either on-premises or to the public cloud, HA configurations, storage automated tiering, and data reduction technologies, including deduplication, among many others.

Available on IBM Cloud® and Amazon Web Services (AWS), IBM Spectrum Virtualize for Public Cloud works together with IBM FlashSystem 5200 to deliver consistent data management between on-premises storage and public cloud. You can move data and applications between on-premises and public cloud, implement new DevOps strategies, use public cloud for DR without the cost of a second data center, or improve cyberresiliency with “air gap” cloud snapshots.

IBM FlashSystem 5200 offers world-class customer support, product upgrades, and other programs:

- ▶ IBM Storage Expert Care service and support is simple. You can easily select the level of support and period that best fits your needs with predictable and upfront pricing that is a fixed percentage of the system cost.
- ▶ The IBM Data Reduction Guarantee helps reduce planning risks and lower storage costs with baseline levels of data compression effectiveness in IBM Spectrum Virtualize based offerings.
- ▶ The IBM Controller Upgrade Program enables customers of designated all-flash IBM storage systems to reduce costs while maintaining leading-edge controller technology for essentially the cost of ongoing system maintenance.

The IBM FlashSystem 5200 control enclosure supports up to twelve 2.5” NVMe-capable flash drives in a 1U high form factor.

There is one standard model of IBM FlashSystem 5200 (4662-6H2) and one utility model (4662-UH6).

Figure 1-26 shows the IBM FlashSystem 5200 control enclosure front view with 12 NVMe drives and a 3/4 ISO view as well.

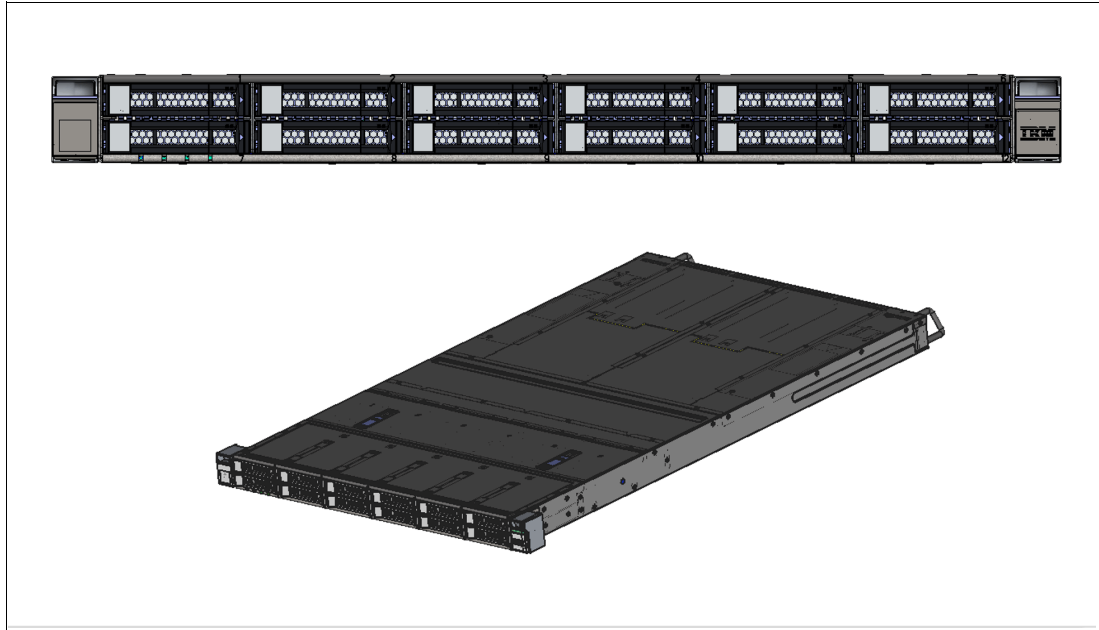


Figure 1-26 IBM FlashSystem 5200 control enclosure front and 3/4 ISO view

Table 1-5 gives a summary of the host connections, drive capacities, features, and standard options with IBM Spectrum Virtualize that are available on IBM FlashSystem 5200.

Table 1-5 IBM FlashSystem 5200 host, drive capacity, and functions summary

Feature / Function	Description
Host interface.	<ul style="list-style-type: none"> ▶ 10 Gbps Ethernet (iSCSI) ▶ 25 Gbps Ethernet (iSCSI, iSER - iWARP, and RoCE) ▶ 16 Gbps Fibre Channel (FC and FC-NVMe) ▶ 32 Gbps Fibre Channel (FC and FC-NVMe)
Control Enclosure Supported drives (12 maximum).	<ul style="list-style-type: none"> ▶ 2.5-inch NVMe self-compressing FCMs: <ul style="list-style-type: none"> – 4.8 TB, 9.6 TB, 19.2 TB, and 38.4 TB ▶ NVMe flash drives: <ul style="list-style-type: none"> – 800 GB, 1.92 TB, 3.84 TB, 7.68 TB, and 15.36 TB
SAS expansion enclosures. 760 per control enclosure. 1,520 per clustered system. Model 12G 2U 12 drives. Model 24G 2U 24 drives. Model 92G 5U 92 drives.	<ul style="list-style-type: none"> ▶ 2.5-inch flash drives supported: <ul style="list-style-type: none"> – 800 GB, 1.6 TB, 1.92 TB, 3.84 TB, 7.68 TB, 15.36 TB, and 30.72 TB ▶ 2.5-inch disk drives supported: <ul style="list-style-type: none"> – 600 GB, 900 GB, 1.2 TB, 1.8 TB, and 2.4 TB 10k SAS disk – 2 TB 7.2 K nearline SAS disk ▶ 3.5-inch disk drives supported: <ul style="list-style-type: none"> – 4 TB, 6 TB, 8 TB, 10 TB, 12 TB, 14 TB, 16 TB, and 18 TB 7.2 K nearline SAS disk
RAID levels.	Distributed RAID 5 and 6, TRAIID 1 and 10

Feature / Function	Description
Advanced features that are included with each system.	<ul style="list-style-type: none"> ▶ Virtualization of internal storage ▶ Data migration ▶ DRPs with thin provisioning ▶ UNMAP ▶ Compression and deduplication ▶ Metro Mirror (synchronous) and Global Mirror (asynchronous)
More available advanced features.	<ul style="list-style-type: none"> ▶ Remote mirroring ▶ IBM Easy Tier® compression ▶ External virtualization ▶ Encryption ▶ FlashCopy ▶ IBM Spectrum Control ▶ IBM Spectrum Protect Snapshot

For more information, see [V8.4.0.x Configuration Limits and Restrictions for IBM FlashSystem 5x00](#).

1.8 IBM FlashSystem 5100 overview

An IBM FlashSystem 5100 Control Enclosure consists of two node canisters that each run IBM Spectrum Virtualize Software. The IBM FlashSystem 5100 system provides affordable, highly functional, and high-performance storage solutions for enterprises of all sizes. The IBM FlashSystem 5100 Models 4H4 and UHB deliver improved latency and performance with the implementation of NVMe technology.

The IBM FlashSystem 5100 SFF Control Enclosure Models 4H4 and UHB feature the following components:

- ▶ Two node canisters, each with an 8-core processor and integrated hardware-assisted compression acceleration
- ▶ 64 GB cache (32 GB per canister) standard with the option of 192 GB - 576 GB (per system)
- ▶ Eight 10 GbE ports standard for iSCSI connectivity or IP replication
- ▶ 16 Gb or 32 Gb FC connectivity options with FC-NVMe support
- ▶ 25 GbE connectivity options with iSCSI or iSER and iSCSI Extensions for RDMA either through RoCe V2 or iWARP
- ▶ Support for up to twenty-four 2.5-inch NVMe flash drives
- ▶ 2U 19-inch rack-mounted enclosure

Figure 1-27 shows the front view of the IBM FlashSystem 5100 Control Enclosure.



Figure 1-27 Front view of an IBM FlashSystem 5100 Control Enclosure with 24 SSD drives

Figure 1-28 shows the rear view of an IBM FlashSystem 5100 Control Enclosure.



Figure 1-28 Rear view of an IBM FlashSystem 5100 Control Enclosure

IBM 2078 Model UHB is the IBM FlashSystem 5100 hardware component that is used in the Storage Utility Offering space. It is physically and functionally identical to the IBM FlashSystem 5100 Model 4H4, except for target configurations and variable capacity billing. The variable capacity billing uses IBM Storage Insights to monitor the system usage, enabling allocated storage usage above a base subscription rate to be billed per terabyte per month.

Allocated storage is identified as *storage that is allocated to a specific host (and unusable to other hosts), whether data is written or not*. For thin provisioning, the data that is written is considered *used*. For thick provisioning, total allocated volume space is considered *used*.

FCM drives integrate IBM MicroLatency® technology, advanced flash management, and reliability into a 2.5-inch SFF NVMe, with built-in, performance-neutral hardware compression and encryption.

The following 2.5-inch SFF NVMe SCM industry-standard drives are supported in IBM FlashSystem 5100 4H4 and UHB control enclosures:

- ▶ 375 GB NVMe SCM drive
- ▶ 750 GB NVMe SCM drive
- ▶ 800 GB NVMe SCM drive
- ▶ 1.6 TB NVMe SCM drive

The following 2.5-inch SFF NVMe FCM drives are supported in the IBM FlashSystem 5100 4H4 and UHB Control Enclosures:

- ▶ 4.8 TB NVMe FCM
- ▶ 9.6 TB NVMe FCM
- ▶ 19.2 TB NVMe FCM
- ▶ 38.4 TB NVMe FCM

The following 2.5-inch SFF NVMe industry-standard drives are supported in the IBM FlashSystem 5100 4H4 and UHB Control Enclosures:

- ▶ 800 GB 2.5-inch 3 Drive Write Per Day (DWPD) NVMe flash drive
- ▶ 1.92 TB 2.5-inch NVMe flash drive
- ▶ 3.84 TB 2.5-inch NVMe flash drive
- ▶ 7.68 TB 2.5-inch NVMe flash drive
- ▶ 15.36 TB 2.5-inch NVMe flash drive

For more information about the drive types, see 1.14, “IBM FlashCore Module drives, NVMe SSDs, and SCM drives” on page 57.

All drives are dual-port and hot-swappable. Drives can be intermixed where applicable. Expansion enclosures can be intermixed behind the SFF control enclosure.

IBM FlashSystem 5100 expansion enclosures

The IBM FlashSystem 5100 Model 4H4 attaches to Expansion Enclosure Models 12G (2U 12-drive), 24G (2U 24-drive), and 92G (5U 92-drive), which support SAS flash drives and SAS HDD drives.

Note: Attachment and intermixing of existing IBM Storwize V5100 / V5000 expansion enclosure models 12F, 24F, and 92F with IBM FlashSystem 5100 expansion enclosure models 12G, 24G, and 92G is supported by IBM FlashSystem 5100 Model 4H4 and with Storwize V5000 models 112, 124, 212, 224, 312, and 324 and Storwize V5100 Model 424.

Attachment and intermixing of existing IBM Storwize V5000 / V5100 expansion enclosure models AFF and A9F with IBM FlashSystem 5100 expansion enclosure models 24G and 92G is supported by Storwize V5000 Model AF3 and Storwize V5100 Model AF4.

The following 2.5-inch SFF flash drives are supported in the expansion enclosures:

400 GB, 800 GB, 1.6 TB, 1.92 TB, 3.2 TB, 3.84 TB, 7.68 TB, 15.36 TB, and 30.72 TB

The following 3.5-inch LFF flash drives are supported in the expansion enclosures:

1.6 TB, 1.92 TB, 3.2 TB, 3.84 TB, 7.68 TB, 15.36 TB, and 30.72 TB

- ▶ 3.5-inch SAS disk drives (Model 12G):
 - 900 GB, 1.2 TB, 1.8 TB, and 2.4 TB 10,000 rpm
 - 4 TB, 6 TB, 8 TB, 10 TB, 12 TB, 14 TB, and 16 TB 7,200 rpm
- ▶ 3.5-inch SAS drives (Model 92G):
 - 1.6 TB, 1.92 TB, 3.2 TB, 3.84 TB, 7.68 TB, 15.36 TB, and 30.72 TB flash drives
 - 1.2 TB, 1.8 TB, and 2.4 TB 10,000 rpm
 - 6 TB, 8 TB, 10 TB, 12 TB, 14 TB, and 16 TB 7,200 rpm
- ▶ 2.5-inch SAS disk drives (Model 24G):
 - 900 GB, 1.2 TB, 1.8 TB, and 2.4 TB 10,000 rpm
 - 2 TB 7,200 rpm
- ▶ 2.5-inch SAS flash drives (Model 24G):
 - 400 GB, 800 GB, 1.6 TB, 1.92 TB, 3.2 TB, 3.84 TB, 7.68 TB, 15.36 TB, and 30.72 TB

Host interface cards

There are two PCIe adapter slots. Table 1-6 shows the supported card combinations (nodes in the same I/O group must match).

Table 1-6 Supported card combinations

Supported number of cards	Ports	Protocol	Slot positions	Note
0 - 1	4	16 Gb FC	2	
0 - 1	4	32 Gb FC	2	
0 - 1	2	25 GbE (iWARP)	2	

Supported number of cards	Ports	Protocol	Slot positions	Note
0 - 1	2	25 GbE RoCE	2	
0 - 1	2	12 Gb SAS Expansion	1	<ul style="list-style-type: none"> ▶ Expansion only, no SAS host attached ▶ Four-port card, but only two are active

On board ports

Table 1-7 shows the onboard ports.

Table 1-7 Onboard ports

Onboard Ethernet Port	Speed	Function
1	10 GbE	Management IP, Service IP, and Host I/O (iSCSI only)
2	10 GbE	Secondary Management IP and Host I/O (iSCSI only)
3	10 GbE	Host I/O (iSCSI only) or IP Replication
4	10 GbE	Host I/O (iSCSI only) or IP Replication
T	10 GbE	Technician Port: DHCP / DNS for direct attach service management

Figure 1-29 shows all of the connectors of an IBM FlashSystem 5100 control bottom canister.

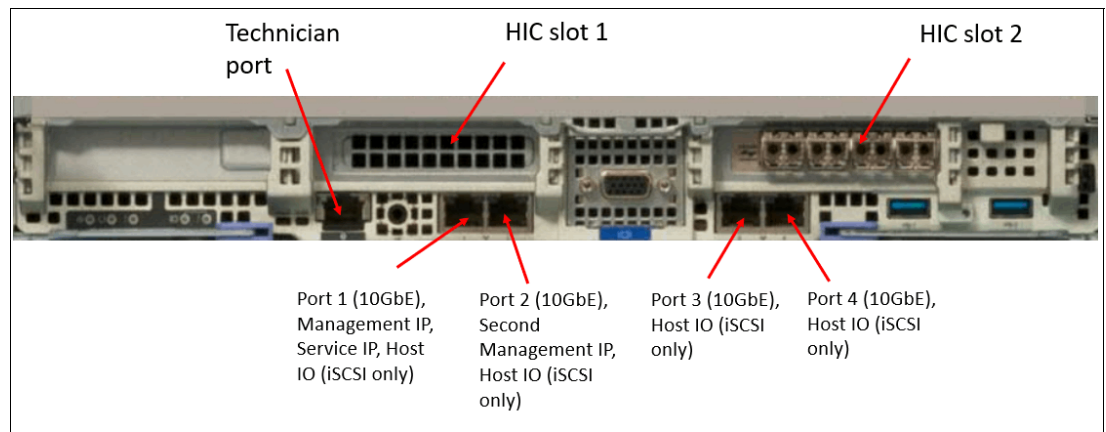


Figure 1-29 Connectors on an IBM FlashSystem 5100 control bottom canister

1.9 IBM FlashSystem 5000 family overview

IBM FlashSystem 5010 and IBM FlashSystem 5030 are all-flash and hybrid flash solutions that provide enterprise-grade functions without compromising affordability or performance, and have the rich features of IBM Spectrum Virtualize. IBM FlashSystem 5000 helps make modern technologies such as artificial intelligence (AI) accessible to enterprises of all sizes.

IBM FlashSystem 5000 is a member of the IBM FlashSystem family of storage solutions. IBM FlashSystem 5000 delivers increased performance and new levels of storage efficiency with superior ease of use. This entry storage solution enables organizations to overcome their storage challenges.

The solution includes technologies to complement and enhance virtual environments, which deliver a simpler, more scalable, and cost-efficient IT infrastructure. IBM FlashSystem 5000 features two node canisters in a compact, 2U 19-inch rack mount enclosure.

Important: At the time writing, IBM FlashSystem 5010 and IBM FlashSystem 5030 are End of Marketing (EOM) and were replaced by the IBM FlashSystem 5015 and IBM FlashSystem 5035. IBM FlashSystem 5015 and IBM FlashSystem 5035 offer superior CPU power and memory options, but the features and functions remain the same. We include the IBM FlashSystem 5015 and IBM FlashSystem 5035 charts only as a reference.

1.9.1 IBM FlashSystem 5015

IBM FlashSystem 5015 is an entry-level solution that is focused on affordability and ease of deployment and operation, with powerful scale-up features. It includes many IBM Spectrum Virtualize features and offers multiple flash and disk drive storage media and expansion options.

Figure 1-30 shows the IBM FlashSystem 5015 and IBM FlashSystem 5035 SFF control enclosure front view.



Figure 1-30 IBM FlashSystem 5015 and IBM FlashSystem 5035 SFF control enclosure front view

Figure 1-31 shows the IBM FlashSystem 5015 and 5035 LFF control enclosure front view.



Figure 1-31 IBM FlashSystem 5015 and 5035 LFF control enclosure front view

Table 1-8 shows the model comparison chart for the IBM FlashSystem 5000 family.

Table 1-8 Machine type and model comparison for the IBM FlashSystem 5000

MTM	Full name
2072-2N2	IBM FlashSystem 5015 LFF Control Enclosure
2072-2N4	IBM FlashSystem 5015 SFF Control Enclosure
2072-3N2	IBM FlashSystem 5035 LFF Control Enclosure
2072-3N4	IBM FlashSystem 5035 SFF Control Enclosure
2072-12G	IBM FlashSystem 5000 LFF Expansion Enclosure
2072-24G	IBM FlashSystem 5000 SFF Expansion Enclosure
2072-92G	IBM FlashSystem 5000 High-Density LFF Expansion Enclosure

Table 1-9 shows a summary of the host connections, drive capacities, features, and standard options with IBM Spectrum Virtualize that are available on IBM FlashSystem 5015.

Table 1-9 IBM FlashSystem 5015 host, drive capacity, and functions summary

Feature / Function	Description
Host interface.	<ul style="list-style-type: none"> ▶ 1 Gb iSCSI (on the system board). ▶ 16 Gbps Fibre Channel. ▶ 12 Gbps SAS. ▶ 25 Gbps iSCSI (iWARP or RoCE). ▶ 10 Gbps iSCSI.
Control Enclosure and SAS expansion enclosures supported drives.	<ul style="list-style-type: none"> ▶ For SFF enclosures, see Table 1-10. ▶ For LFF enclosures, see Table 1-11.
Cache per control enclosure / clustered system.	32 GB or 64 GB.
RAID levels.	DRAID 1, 5, and 6.
Maximum expansion enclosure capacity.	Up to 10 standard expansion enclosures per controller. Up to four high-density expansion enclosures per controller.

Feature / Function	Description
Advanced functions that are included with each system.	<ul style="list-style-type: none"> ▶ Virtualization of internal storage. ▶ DRPs with thin provisioning and UNMAP. ▶ One-way data migration.
More available advanced features.	<ul style="list-style-type: none"> ▶ Easy Tier. ▶ FlashCopy. ▶ Remote mirroring.

Table 1-10 shows the 2.5-inch supported drives for IBM FlashSystem 5000 family.

Table 1-10 2.5-inch supported drives for the IBM FlashSystem 5000 family

2.5-inch (SFF)	Capacity					
Tier 1 flash	800 GB	1.9 TB	3.84 TB	7.68 TB	15.36 TB	30.72 TB
High-performance enterprise disk drives (10K rpm)	900 GB	1.2 TB	1.8 TB	2.4 TB		
High capacity nearline disk drives (7.2 K rpm)	2 TB					

Table 1-11 shows the 3.5-inch supported drives for IBM FlashSystem 5000 family.

Table 1-11 3.5-inch supported drives for the IBM FlashSystem 5000 family

3.5-inch (LFF)	Speed	Capacity							
High-performance, enterprise class disk drives	10,000 RPM	900 GB	1.2 TB	1.8 TB	2.4 TB				
High capacity archival class nearline disk drives	7,200 RPM	4 TB	6 TB	8 TB	10 TB	12 TB	14 TB	16 TB	18 TB

1.9.2 IBM FlashSystem 5035

IBM FlashSystem 5035 provides powerful functions, including powerful encryption capabilities and DRPs with compression, deduplication, thin provisioning, and the ability to cluster for scale-up and scale-out.

Available with the IBM FlashSystem 5035 model, DRPs help transform the economics of data storage. When applied to new or existing storage, they can increase usable capacity while maintaining consistent application performance. DRPs can help eliminate or drastically reduce costs for storage acquisition, rack space, power, and cooling, and can extend the useful life of existing storage assets. Their capabilities include the following ones:

- ▶ Block deduplication that works across all the storage in a DRP to minimize the number of identical blocks.
- ▶ New compression technology that ensures consistent 2:1 or better reduction performance across a wide range of application workload patterns.
- ▶ SCSI UNMAP support that de-allocates physical storage when operating systems delete logical storage constructs such as files in a file system.

Table 1-12 summarizes the host connections, drive capacities, features, and standard options with IBM Spectrum Virtualize that are available on IBM FlashSystem 5035.

Table 1-12 IBM FlashSystem 5035 host, drive capacity, and functions summary

Feature / Function	Description
Host interface.	<ul style="list-style-type: none"> ▶ 10 Gb iSCSI (on the system board). ▶ 16 Gbps Fibre Channel. ▶ 12 Gbps SAS. ▶ 25 Gbps iSCSI (iWARP or RoCE). ▶ 10 Gbps iSCSI.
Control enclosure and SAS expansion enclosures supported drives.	<ul style="list-style-type: none"> ▶ For SFF enclosures, see Table 1-10. ▶ For LFF enclosures, see Table 1-11.
Cache per control enclosure / clustered system.	32 GB or 64 GB / 64 GB or 128 GB
RAID levels.	DRAID 1, 5 (CLI only), and 6
Maximum expansion enclosure capacity.	<ul style="list-style-type: none"> ▶ Up to 20 standard expansion enclosures per controller. ▶ Up to eight high-density expansion enclosures per controller.
Advanced functions that are included with each system.	<ul style="list-style-type: none"> ▶ Virtualization of internal storage. ▶ DRPs with thin provisioning. ▶ UNMAP, compression, and deduplication. ▶ One-way data migration. ▶ Dual-system clustering.
More available advanced features.	<ul style="list-style-type: none"> ▶ Easy Tier. ▶ FlashCopy. ▶ Remote mirroring. ▶ Encryption.

For more information, see [V8.4.0.x Configuration Limits and Restrictions for IBM FlashSystem 5015 and IBM FlashSystem 5035](#).

This next section provides hardware information about the IBM FlashSystem 5010 and 5030 models.

1.9.3 IBM FlashSystem 5010 hardware overview

IBM FlashSystem 5010 features two-core processors with up to 64 GB total cache, and the attachment of up to 10 2U expansion enclosures or up to four 5U expansion enclosures. This configuration delivers support for up to 392 drives.

Note: The IBM FlashSystem 5010 solution supports only one SAS expansion chain.

The IBM FlashSystem 5010 control enclosure features the following components:

- ▶ Two node canisters, each with a two-core processor
- ▶ 16 GB cache (8 GB per canister) with optional 32 GB cache (16 GB per canister) or 64 GB cache (32 GB per canister)
- ▶ 1 Gb iSCSI connectivity standard with optional 16 Gb FC, 12 Gb SAS, 10 Gb iSCSI (optical), or 25 Gb iSCSI (optical) connectivity
- ▶ 12 Gb SAS port for expansion enclosure attachment

- ▶ Twelve slots for 3.5-inch LFF SAS drives (Model 2H2) and 24 slots for 2.5-inch SFF SAS drives (Model 2H4)
- ▶ 2U, 19-inch rack mount enclosure with 100 - 240 V AC or -48 V DC power supplies

The LFF enclosure models support up to twelve 3.5-inch drives, and the SFF enclosure models support up to twenty-four 2.5-inch drives. High-performance disk drives, high-capacity nearline (NL) disk drives, and flash (SSDs) also are supported. Drives of the same form factor can be intermixed within an enclosure, which provides the flexibility to address performance and capacity needs in a single enclosure. You can also intermix LFF and SFF expansion enclosures behind any control enclosure.

Table 1-13 lists the supported 2.5-inch drives for IBM FlashSystem 5000.

Table 1-13 2.5-inch supported drives for IBM FlashSystem 5000

2.5-inch (SFF)	Capacity					
	Tier 1 flash	800 GB	1.9 TB	3.84 TB	7.68 TB	15.36 TB
High-performance enterprise disk drives (10 K RPM)	900 GB	1.2 TB	1.8 TB	2.4 TB		
High capacity NL disk drives (7.2 K RPM)	2 TB					

Table 1-14 shows the supported 3.5-inch (LFF) drives for IBM FlashSystem 5000.

Table 1-14 3.5-inch supported drives for IBM FlashSystem 5000

3.5-inch (LFF)	Speed	Capacity						
		High-performance enterprise class disk drives	10,000 RPM	900 GB	1.2 TB	1.8 TB	2.4 TB	
High capacity archival class NL disk drives	7,200 RPM	4 TB	6 TB	8 TB	10 TB	12 TB	14 TB	16 TB

Figure 1-32 shows the IBM FlashSystem 5010 SFF Control Enclosure with 24 drives.



Figure 1-32 Front view of IBM FlashSystem 5010

Figure 1-33 shows the rear view of an IBM FlashSystem 5010 control enclosure.



Figure 1-33 Rear view of an IBM FlashSystem 5010

Figure 1-34 shows the available connectors and light-emitting diodes (LEDs) on a single IBM FlashSystem 5010 canister.

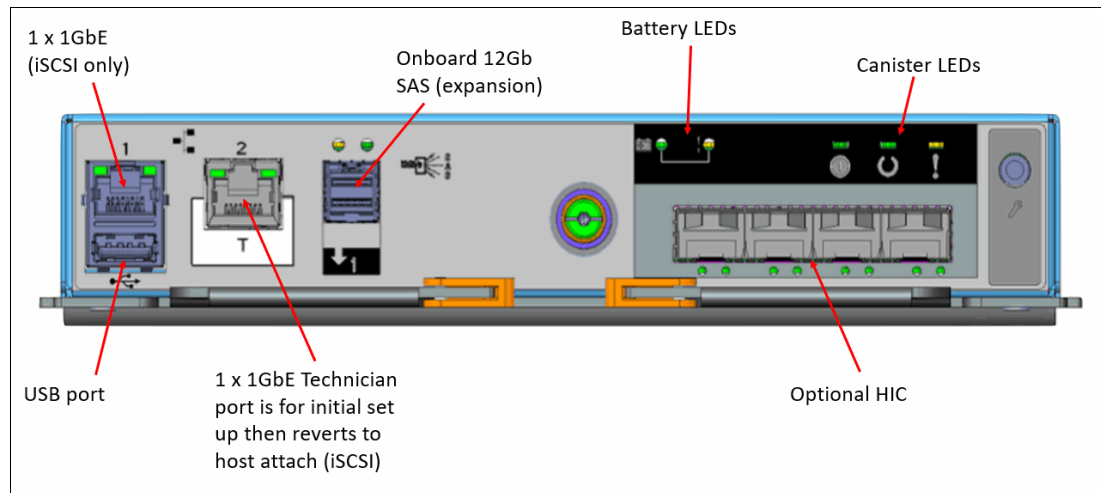


Figure 1-34 View of available connectors and LEDs on an IBM FlashSystem 5010 single canister

1.9.4 IBM FlashSystem 5030 hardware overview

The IBM FlashSystem 5030 control enclosure features the following components:

- ▶ Two node canisters, each with a six-core processor
- ▶ 32 GB cache (16 GB per canister) with optional 64 GB cache (32 GB per canister)
- ▶ 10 Gb iSCSI (copper) connectivity standard with optional 16 Gb FC, 12 Gb SAS, 10 Gb iSCSI (optical), or 25 Gb iSCSI (optical)
- ▶ 12 Gb SAS port for expansion enclosure attachment
- ▶ Twelve slots for 3.5-inch LFF SAS drives (Model 3H2) and 24 slots for 2.5-inch SFF SAS drives (Model 3H4)
- ▶ 2U, 19-inch rack mount enclosure with 100 - 240 V AC or -48 V DC power supplies

The IBM FlashSystem 5030 control enclosure models offer the highest level of performance, scalability, and functions and include the following features:

- ▶ Support for 760 drives per system with the attachment of eight IBM FlashSystem 5000 High-Density LFF Expansion Enclosures and 1,520 drives with a two-way clustered configuration
- ▶ DRPs with deduplication, compression,¹ and thin provisioning for improved storage efficiency

- Encryption of data-at-rest that is stored within the IBM FlashSystem 5030 system
- Figure 1-35 shows the IBM FlashSystem 5030 SFF Control Enclosure with 24 drives.



Figure 1-35 Front view of an IBM FlashSystem 5030

Figure 1-36 shows the rear view of an IBM FlashSystem 5030 Control Enclosure.



Figure 1-36 Rear view of an IBM FlashSystem 5030

Figure 1-37 shows the available connectors and LEDs on a single IBM FlashSystem 5030 canister.

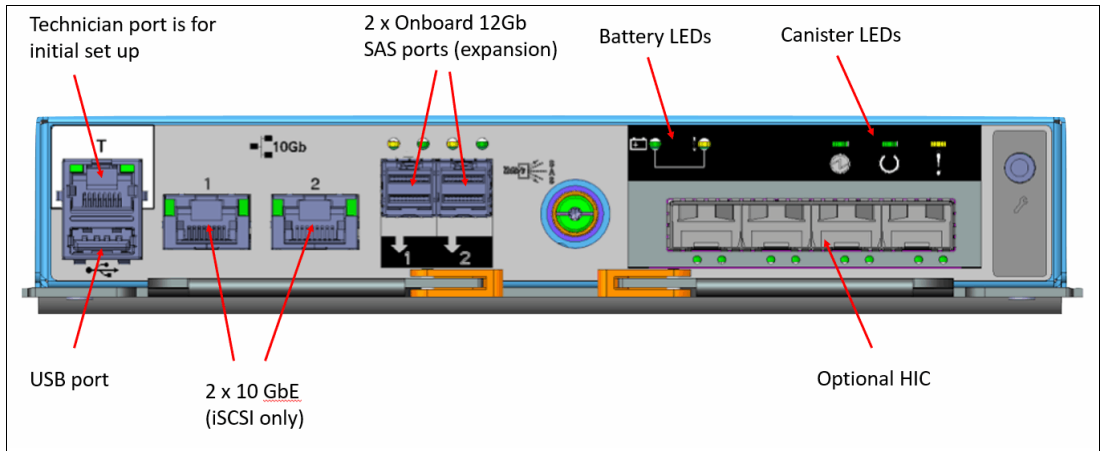


Figure 1-37 View of available connectors and LEDs on an IBM FlashSystem 5030 single canister

¹ Deduplication and compression require 64 GB of system cache.

1.10 Features for storage efficiency and data reduction

IBM Spectrum Virtualize software running in the IBM FlashSystem storage systems offers several functions for storage optimization and efficiency.

IBM Easy Tier

Many applications exhibit a significant skew in the distribution of I/O workload. A small fraction of the storage is responsible for a disproportionately large fraction of the total I/O workload of an environment.

In a tiered storage pool, IBM Easy Tier acts to identify this skew and automatically place data in the appropriate tier to take advantage of it. By moving the hottest data onto the fastest tier of storage, the workload on the remainder of the storage is reduced. By servicing most of the application workload from the fastest storage, Easy Tier acts to accelerate application performance.

Easy Tier is a performance optimization function that automatically migrates (move) extents that belong to a volume among different storage tiers based on their I/O load. The movement of the extents is online and unnoticed from a host perspective.

As a result of extent movement, the volume no longer has all its data in one tier, but rather in two or three tiers. Each tier provides optimal performance for the extent, as shown in Figure 1-38.

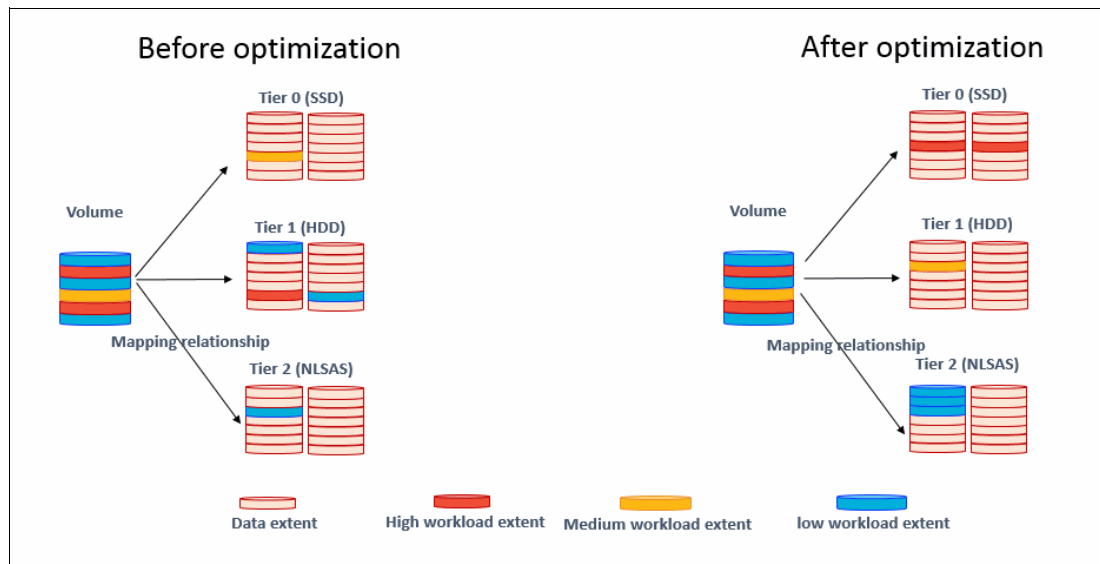


Figure 1-38 Easy Tier concept

Easy Tier monitors the I/O activity and latency of the extents on all Easy Tier enabled storage pools to create *heat maps*. Based on them, Easy Tier creates an extent migration plan and promotes (moves) high activity or hot extents to a higher disk tier within the same storage pool. It also demotes extents whose activity dropped off, or cooled, by moving them from a higher disk tier managed disk (MDisk) back to a lower tier MDisk.

Storage pools that contain only one tier of storage can also benefit from Easy Tier if they have multiple disk arrays (or MDisks). Easy Tier has a balancing mode: It moves extents from busy disk arrays to less busy arrays of the same tier, balancing I/O load.

All MDisks (disk arrays) belong to one of the tiers. They are classified as SCM, tier 0 flash, tier 1 flash, enterprise, or NL tier.

Data reduction and UNMAP

The UNMAP feature is a set of SCSI primitives that enables hosts to indicate to a SCSI target (storage system) that space that is allocated to a range of blocks on a target storage volume is no longer required. This command enables the storage controller to take measures and optimize the system so that the space can be reused for other purposes.

The most common use case, for example, is a host application, such as VMware, freeing storage in a file system. Then, the storage controller can perform functions to optimize the space, such as reorganizing the data on the volume so that space is better used.

When a host allocates storage, the data is placed in a volume. To free the allocated space back to the storage pools, the SCSI UNMAP feature is used. UNMAP enables host OSs to deprovision storage on the storage controller so that the resources can automatically be freed in the storage pools and used for other purposes.

A DRP increases infrastructure capacity usage by using new efficiency functions and reducing storage costs. By using the end-to-end SCSI UNMAP function, a DRP can automatically de-allocate and reclaim the capacity of thin-provisioned volumes that contain deleted data so that this reclaimed capacity can be reused by other volumes.

At its core, a DRP uses a Log Structured Array (LSA) to allocate capacity. An LSA enables a tree-like directory to be used to define the physical placement of data blocks independent of size and logical location. Each logical block device has a range of logical block addresses (LBAs), starting from 0 and ending with the block address that fills the capacity.

When written, you can use an LSA to allocate data sequentially and provide a directory that provides a lookup to match an LBA with a physical address within the array. Therefore, the volume that you create from the pool to present to a host application consists of a directory that stores the allocation of blocks within the capacity of the pool.

In DRPs, the maintenance of the metadata results in *I/O amplification*. I/O amplification occurs when a single host-generated read or write I/O results in more than one back-end storage I/O request because of advanced functions. A read request from the host results in two I/O requests: a directory lookup and a data read. A write request from the host results in three I/O requests: a directory lookup, a directory update, and a data write. This aspect must be considered when sizing and planning your data-reducing solution.

Standard pools, which make up a classic solution that is also supported by the IBM FlashSystem storage systems, do not use LSA. A standard pool works as a container that receives its capacity from MDisks (disk arrays), splits it into extents of the same fixed size, and allocates extents to volumes.

Standard pools do not cause I/O amplification and require less processing resource usage compared to DRPs. In exchange, DRPs provide more flexibility and storage efficiency.

Table 1-15 provides an overview of volume capacity saving types that are available with standard pools and DRPs.

Table 1-15 Volume types that are available in pools

Saving type	Standard pool	DRP
Fully allocated	Yes	Yes
Thin-provisioned	Yes	Yes

Saving type	Standard pool	DRP
Thin-provisioned compressed	No	Yes
Thin-provisioned deduplicated	No	Yes
Thin-provisioned compressed and deduplicated	No	Yes

Best practice: If you want to use deduplication, create thin-provisioned compressed and deduplicated volumes.

This book provides only an overview of DRP aspects. For more information, see *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430.

Fully allocated and thin-provisioned volumes

Volumes can be configured as thin-provisioned or fully allocated. Both can be configured in standard pools and DRPs.

In IBM FlashSystem family systems, each volume has *virtual capacity* and *real capacity* parameters. Virtual capacity is the volume storage capacity that is available to a host, and it is used by to create a file system. Real capacity is the storage capacity that is allocated to a volume from a pool. It shows the amount of space that is used on a physical storage volume.

Fully allocated volumes are created with the same amount of real capacity and virtual capacity. This type uses no storage efficiency features.

When a fully allocated volume is created on a DRP, it bypasses the LSA structure and works in the same manner as in a standard pool, so it has no processing impact and provides no data reduction options at the pool level.

When using fully allocated volumes on the IBM FlashSystem storage systems with FCM drives, whether a DRP or standard pool is used, capacity savings are achieved by compressing data with hardware compression that runs on the FCM drives. Hardware compression on FCM drives is always on and cannot be turned off. This configuration provides maximum performance in combination with outstanding storage efficiency.

A thin-provisioned volume presents a different capacity to mapped hosts than the capacity that the volume uses in the storage pool. Therefore, real and virtual capacities might not be equal. The virtual capacity of a thin-provisioned volume is typically significantly larger than its real capacity. As more information is written by the host to the volume, more of the real capacity is used. The system identifies read operations to unwritten parts of the virtual capacity, and returns zeros to the server without using any real capacity.

In a shared storage environment, thin provisioning is a method for optimizing the use of available storage. Thin provisioning relies on the allocation of blocks of data on demand, versus the traditional method of allocating all of the blocks up front. This method eliminates almost all white space, which helps avoid the poor usage rates that occur in the traditional storage allocation method where large pools of storage capacity are allocated to individual servers but remain unused (not written to).

If a thin-provisioned volume is created in a DRP, the system monitors it for reclaimable capacity from host unmap operations. This capacity can be reclaimed and redistributed into the pool. Space that is freed from the hosts is a process that is called *UNMAP*. A host can issue **SCSI UNMAP** commands when the user deletes files on a file system, which result in the freeing of all of the capacity that is allocated within that unmapping.

A thin-provisioned volume in a standard pool will not return unused capacity back to the pool with SCSI UNMAP.

A thin-provisioned volume can be converted non-disruptively to a fully allocated volume, or vice versa, by using the volume mirroring function. For example, you can add a thin-provisioned copy to a fully allocated primary volume, and then remove the fully allocated copy from the volume after they are synchronized.

Note: It is *not* recommended to use non-compressed thin-provisioned volumes on DRPs containing FCM drives. The system GUI prevents the creation of such types of configurations.

1.10.1 Compression and deduplication

When using DRPs on the IBM FlashSystem storage systems, host data can be compressed or compressed and deduplicated before it is written to the disk drives.

The IBM FlashSystem family DRP compression is based on the Lempel-Ziv lossless data compression algorithm that operates by using a real-time method. When a host sends a write request, the request is acknowledged by the write cache of the system, and then staged to the DRP.

As part of its staging, the write request passes through the compression engine and is stored in a compressed format. Therefore, writes are acknowledged immediately after they are received by the write cache with compression occurring as part of the staging to internal or external physical storage. This process occurs transparently to host systems, which makes them unaware of the compression.

The *IBM Comprestimator* tool is available to check whether your data is compressible. It estimates the space savings that are achieved when using compressed volumes. This utility provides a quick and easy view of showing the benefits of using compression. IBM Comprestimator can be run from the system GUI or command-line interface (CLI), and it checks data that is already stored on the system. In DRPs, IBM Comprestimator is always on starting at code level 8.4, so you can display the compressibility of the data in the GUI and IBM Storage Insights at any time. It is also available as a stand-alone, host-based utility that can analyze data on IBM or third-party storage devices. For more information, see [Comprestimator Utility Version 1.5.3.1](#).

Deduplication can be configured with thin-provisioned and compressed volumes in DRPs for added capacity savings. The deduplication process identifies unique chunks of data, or byte patterns, and stores a signature of the chunk for reference when writing new data chunks.

If the new chunk's signature matches an existing signature, the new chunk is replaced with a small reference that points to the stored chunk. The matches are detected when the data is written. The same byte pattern might occur many times, which greatly reduce the amount of data that must be stored.

To help with the profiling and analysis of existing workloads that must be migrated to an IBM FlashSystem system, IBM provides the Data Reduction Estimation Tool (DRET). DRET is a highly accurate command-line and host-based utility for estimating the data reduction savings on block storage devices. The tool scans target workloads on various storage arrays (from IBM or another storage vendor), merges all scan results, and provides a data reduction estimate.

Compression and deduplication are not mutually exclusive: One, both, or neither, features can be enabled. If the volume is deduplicated and compressed, data is deduplicated first, and then compressed. Therefore, deduplication references are created on the compressed data that is stored on the physical domain.

1.10.2 Features for enhanced data security

To protect data against the potential exposure of sensitive user data and user metadata that is stored on discarded, lost, or stolen storage devices, IBM FlashSystem storage systems support encryption of data-at-rest.

Encryption is performed by the IBM FlashSystem controllers for data that is stored within the entire system, the IBM FlashSystem Control Enclosure, all attached expansion enclosures, and for data that is stored as externally virtualized by the IBM FlashSystem storage systems.

Encryption is the process of encoding data so that only authorized parties can read it. Data encryption is protected by the Advanced Encryption Standard (AES) algorithm that uses a 256-bit symmetric encryption key in XTS mode, as defined in the IEEE 1619-2007 standard and NIST Special Publication 800-38E as XTS-AES-256.

There are two types of encryption on devices running IBM Spectrum Virtualize: *hardware encryption* and *software encryption*. Which method is used for encryption is chosen automatically by the system based on the placement of the data:

- ▶ **Hardware encryption:** Data is encrypted by using SAS hardware. It is used only for internal storage (drives).
- ▶ **Software encryption:** Data is encrypted by using the nodes' CPU (the encryption code uses the AES-NI CPU instruction set). It is used only for external storage that is virtualized by the IBM FlashSystem storage systems.

Both methods of encryption use the same encryption algorithm, key management infrastructure, and license.

Note: Only data-at-rest is encrypted. Host to storage communication and data that is sent over links that are used for remote mirroring are not encrypted.

The IBM FlashSystem also supports self-encrypting drives, where data encryption is completed in the drive itself.

Before encryption can be enabled, ensure that a license was purchased and activated.

The system supports two methods of configuring encryption:

- ▶ You can use a centralized external key server that simplifies creating and managing encryption keys on the system. This method of encryption key management is preferred for security and simplification of key management.

A *key server* is a centralized system that generates, stores, and sends encryption keys to the system. Some key server providers support replication of keys among multiple key servers. If multiple key servers are supported, you can specify up to four key servers that connect to the system over both a public network or a separate private network. The system supports IBM Security™ Key Lifecycle Manager or Gemalto SafeNet KeySecure key servers to handle key management.

- ▶ In addition, the system supports storing encryption keys on USB flash drives. USB flash drive-based encryption requires physical access to the systems and is effective in environments with a minimal number of systems. For organizations that require strict security policies regarding USB flash drives, the system supports disabling a canister's USB ports to prevent unauthorized transfer of system data to portable media devices. If you have such security requirements, use key servers to manage encryption keys.

1.11 Features for application integration

IBM FlashSystem storage systems include the following features, which enable tight integration with VMware:

- ▶ vCenter plug-in: Enables monitoring and self-service provisioning of the system from within VMware vCenter.
- ▶ vStorage application programming interfaces (APIs) for Array Integration (VAI) support: This function supports hardware-accelerated virtual machine (VM) copy / migration and hardware-accelerated VM initiation, and accelerates VMware Virtual Machine File System (VMFS).
- ▶ Microsoft Windows System Resource Manager (SRM) for VMware Site Recovery Manager: Supports automated storage and host failover, failover testing, and failback.
- ▶ VMware vSphere Virtual Volume (VVOL) integration for better usability: The migration of space-efficient volumes between storage containers maintains the space efficiency of volumes. Cloning a VM achieves a full independent set of VVOLS, and resiliency is improved for VMs if volumes start running out of space.

VMware vSphere Virtual Volumes

The system supports VVOLS, which allow VMware vCenter to automate the management of system objects like volumes and pools. It is an integration and management framework that virtualizes the IBM FlashSystem storage systems, enabling a more efficient operational model that is optimized for virtualized environments and centered on the application instead of the infrastructure.

VVOLS simplify operations through policy-driven automation that enables more agile storage consumption for VMs and dynamic adjustments in real time when they are needed. It simplifies the delivery of storage service levels to individual applications by providing finer control of hardware resources and native array-based data services that can be instantiated with VM granularity.

With VVOLS, VMware offers a paradigm in which an individual VM and its disks, rather than a logical unit number (LUN), becomes a unit of storage management for a storage system. It encapsulates VDisks and other VM files, and natively store the files on the storage system.

By using a special set of APIs called vSphere APIs for Storage Awareness (VASA), the storage system becomes aware of the VVOLs and their associations with the relevant VMs. Through VASA, vSphere and the underlying storage system establish a two-way out-of-band communication to perform data services and offload certain VM operations to the storage system. For example, some operations, such as snapshots and clones, can be offloaded.

For more information about VVOLs and the actions that are required to implement this feature on the host side, see the [VMware website](#).

IBM support for VASA is provided by IBM Spectrum Connect enabling communication between the VMware vSphere infrastructure and the IBM FlashSystem system. The IBM FlashSystem administrator can assign ownership of VVOLs to IBM Spectrum Connect by creating a user with the VASA Provider security role.

Although the system administrator can complete certain actions on volumes and pools that are owned by the VASA Provider security role, IBM Spectrum Connect retains management responsibility for VVOLs. For more information about IBM Spectrum Connect, see the [IBM Spectrum Connect documentation](#).

Note: At the time of writing, VVOLs are not supported on DRPs.

1.12 Features for manageability

IBM FlashSystem storage systems offer the following manageability and serviceability features:

- ▶ An intuitive GUI
- ▶ IBM Call Home
- ▶ IBM Storage Insights
- ▶ IBM Spectrum Virtualize RESTful API

The IBM FlashSystem family system GUI

IBM FlashSystem storage systems include an easy-to-use management GUI that runs on one of the node canisters in the control enclosure to help you monitor, manage, and configure your system. You can access the GUI by opening any supported web browser and entering the management IP addresses.

IBM FlashSystem use a GUI with the same look and feel across all platforms for a consistent management experience. The GUI has an improved overview dashboard that provides all information in an easy-to-understand format and enables visualization of effective capacity. With the GUI, you can quickly deploy storage and manage it efficiently.

Figure 1-39 on page 51 shows the IBM FlashSystem GUI dashboard view. This view is the default that is displayed after the user logs on to the system.

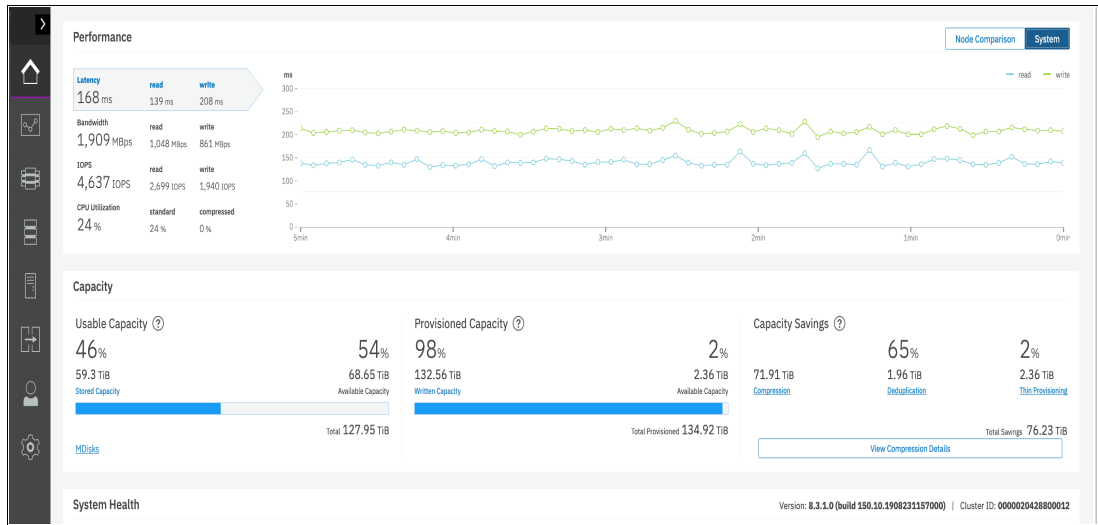


Figure 1-39 IBM FlashSystem GUI dashboard

The IBM FlashSystem storage systems also provide a CLI, which is useful for advanced configuration and scripting.

The systems support SNMP, email notifications that use Simple Mail Transfer Protocol (SMTP), and syslog redirection for complete enterprise management access.

IBM Call Home

Call Home connects the system to IBM Service Personnel who can monitor and respond to system events to ensure that your system remains running. The Call Home function opens a service alert if a serious error occurs in the system, automatically sending the details of the error and contact information to IBM Service Personnel.

If the system is entitled for support, a Problem Management Record (PMR) is automatically created and assigned to the appropriate IBM Support team. The information that is provided to IBM is an excerpt from the event log containing the details of the error, and client contact information from the system. IBM Service Personnel contact the client and arrange service on the system, which can greatly improve the speed of resolution by removing the need for the client to detect the error and raise a support call themselves.

The system supports two methods to transmit notifications to the support center:

- Call Home with cloud services

Call Home with cloud services sends notifications directly to a centralized file repository that contains troubleshooting information that is gathered from customers. Support personnel can access this repository and automatically be assigned issues as problem reports.

This method of transmitting notifications from the system to support removes the need for customers to create problem reports manually. Call Home with cloud services also eliminates email filters dropping notifications to and from support, which can delay resolution of problems on the system.

This method sends notifications only to the predefined support center.

- ▶ **Call Home with email notifications**

Call Home with email notification sends notifications through a local email server to support and local users or services that monitor activity on the system. With email notifications, you can send notifications to support and designate internal distribution of notifications, which alerts internal personnel about potential problems. Call Home with email notifications requires configuring at least one email server, and local users.

However, external notifications to the support center can be dropped if filters on the email server are active. To eliminate this problem, Call Home with email notifications is not recommended as the only method to transmit notifications to the support center. Call Home with email notifications can be configured together with cloud services.

IBM highly encourages all clients to take advantage of the Call Home feature so that you and IBM can collaborate for your success.

IBM Storage Insights

IBM Storage Insights is an IBM Cloud software as a service (SaaS) offering that can help you monitor and optimize the storage resources in the system and across your data center. IBM Storage Insights monitors your storage environment and provides information about the statuses of multiple systems in a single dashboard. You can view data from the perspectives of the servers, applications, and file systems. Two versions of IBM Storage Insights are available: IBM Storage Insights and IBM Storage Insights Pro.

When you order any IBM FlashSystem storage system, IBM Storage Insights is available at no additional cost. With this version, you can monitor the basic health, status, and performance of various storage resources.

IBM Storage Insights Pro is a subscription-based product that provides a more comprehensive view of the performance, capacity, and health of your storage resources. In addition to the features that are offered by IBM Storage Insights, IBM Storage Insights Pro provides tools for intelligent capacity planning, storage reclamation, storage tiering, and performance troubleshooting services. Together, these features can help you reduce storage costs and optimize your data center.

IBM Storage Insights is a part of the monitoring and helps to ensure continued availability of the IBM FlashSystem storage systems.

The tool provides a single dashboard that gives you a clear view of all your IBM block and file storage and some other storage vendors (the IBM Storage Insights Pro version is required to view other storage vendors' storage). You can make better decisions by seeing trends in performance and capacity. With storage health information, you can focus on areas that need attention. When IBM Support is needed, IBM Storage Insights simplifies uploading logs, speeds resolution with online configuration data, and provides an overview of open tickets, all in one place.

The following features are some of the ones that are available with IBM Storage Insights:

- ▶ **A unified view of IBM systems:**
 - Provides a single view to see all your system's characteristics.
 - See all of your IBM storage inventory.
 - Provides a live event feed so that you know in real time what is going on with your storage so that you can act fast.
- ▶ IBM Storage Insights collects telemetry data and Call Home data and provides real-time system reporting of capacity and performance.

- ▶ Overall storage monitoring by looking at the following information:
 - The overall health of the system.
 - Monitoring of the configuration to see whether it meets best practices.
 - System resource management determines which system is overtaxed, and provides proactive recommendations to fix it.
- ▶ IBM Storage Insights provides advanced customer service with an event filter that you can use to accomplish the following tasks:
 - You and IBM Support can view support tickets, open and close them, and track trends.
 - You can use the autolog collection capability to collect the logs and send them to IBM before IBM Support looks into the problem. This capability can save time in resolving the case.

Figure 1-40 shows a view of the IBM Storage Insights dashboard.

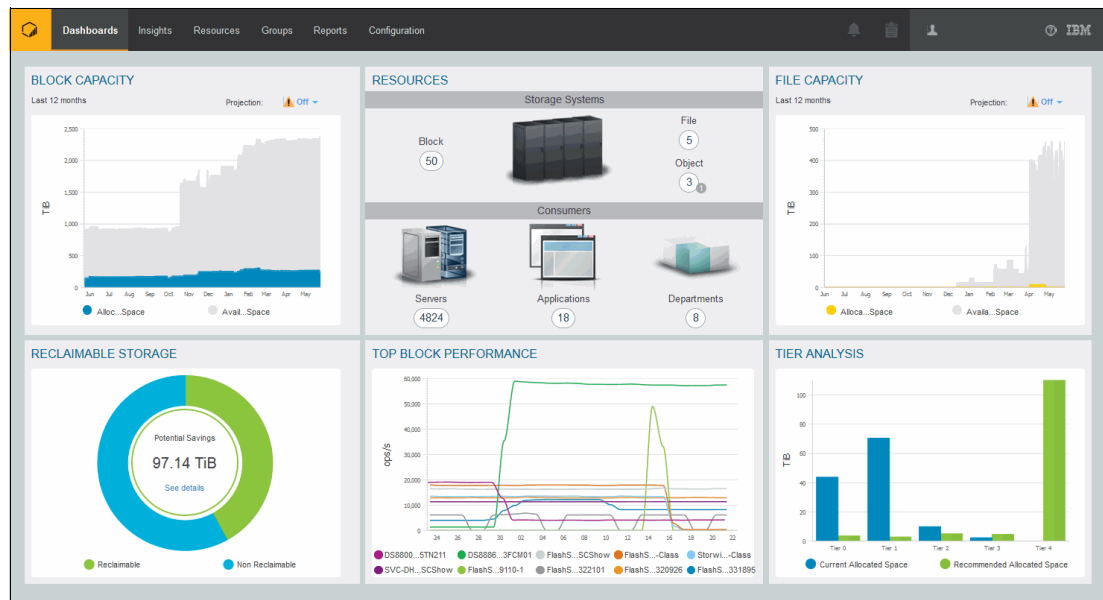


Figure 1-40 IBM Storage Insights dashboard

In order for IBM Storage Insights to operate, a lightweight data collector must be deployed in your data center to stream only system metadata to your IBM Cloud instance. The metadata flows in one direction: from your data center to IBM Cloud over HTTPS. The actual application data that is stored on the storage systems cannot be accessed by the data collector. In

IBM Cloud, your metadata is AES256-encrypted and protected by physical, organizational, access, and security controls.

IBM Storage Insights is ISO/IEC 27001 Information Security Management certified.

For more information about IBM Storage Insights, see the following websites:

- ▶ [IBM Storage Insights Fact Sheet](#)
- ▶ [Functional demonstration environment](#)
- ▶ [IBM Storage Insights security information](#)
- ▶ [IBM Storage Insights registration](#)

IBM Spectrum Virtualize RESTful API

The IBM Spectrum Virtualize Representational State Transfer (REST) model API consists of command targets that are used to retrieve system information and to create, modify, and delete system resources. These command targets allow command parameters to pass through unedited to the IBM Spectrum Virtualize CLI, which handles parsing parameter specifications for validity and error reporting and uses HTTPS to successfully communicate with the RESTful apiserver.

The RESTful apiserver does not consider transport security (such as Secure Sockets Layer (SSL)), but instead assumes that requests are initiated from a local, secured server. The HTTPS protocol provides privacy through data encryption. The RESTful API provides more security by requiring command authentication, which persists for 2 hours of activity or 30 minutes of inactivity, whichever occurs first.

Uniform Resource Locators (URLs) target different node objects on the system. The **HTTPS POST** method acts on command targets that are specified in the URL. To make changes or view information about different objects on the system, you must create and send a request to the system. You must provide certain elements for the RESTful apiserver to receive and transform the request into a command.

To interact with the system by using the RESTful API, make an HTTPS command request with a valid configuration node URL destination. Open TCP port 7443 and include the keyword **rest**, and then use the following URL format for all requests:

```
https://system_node_ip:7443/rest/command
```

Where:

- ▶ *system_node_ip* is the system IP address, which is the address that is taken by the configuration node of the system.
- ▶ The port number is always 7443 for the IBM Spectrum Virtualize RESTful API.
- ▶ **rest** is a keyword.
- ▶ *command* is the target command object (such as **auth** or **lsevenlog** with any parameters). The command specification follows this format:

```
command_name,method="POST",headers={'parameter_name': 'parameter_value',  
'parameter_name': 'parameter_value',...}
```

1.13 Copy services

IBM FlashSystem systems provide copy services functions that can be to improve availability and support DR.

Volume mirroring

By using volume mirroring, a volume can have two physical copies in one IBM FlashSystem system. Each volume copy can belong to a different pool and use a different set of capacity saving features.

When a host writes to a mirrored volume, the system writes the data to both copies. When a host reads a mirrored volume, the system picks one of the copies to read. If one of the mirrored volume copies is temporarily unavailable, the volume remains accessible to servers. The system remembers which areas of the volume are written, and resynchronizes these areas when both copies are available.

You can create a volume with one or two copies, and you can convert a non-mirrored volume into a mirrored volume by adding a copy. When a copy is added in this way, the system synchronizes the new copy so that the new copy is the same as the existing volume. Servers can access the volume during this synchronization process.

Volume mirroring can be used to migrate data to or from an IBM FlashSystem family system. For example, you can start with a non-mirrored image mode volume in the migration pool, and then add a copy to that volume in the destination pool on internal storage. After the volume is synchronized, you can delete the original copy that is in the source pool. During the synchronization process, the volume remains available.

Volume mirroring is also used to convert fully allocated volumes to use data reduction technologies, such as thin-provisioning, compression, or deduplication, or to migrate volumes between storage pools.

FlashCopy

The FlashCopy or snapshot function creates a point-in-time (PIT) copy of data that is stored on a source volume to a target volume. FlashCopy is sometimes described as an instance of a time-zero (T0) copy. Although the copy operation takes some time to complete, the resulting data on the target volume is presented so that the copy appears to have occurred immediately, and all data is available immediately. Advanced functions of FlashCopy allow operations to occur on multiple source and target volumes.

Management operations are coordinated to provide a common, single PIT for copying target volumes from their respective source volumes to create a consistent copy of data that spans multiple volumes.

The function also supports multiple target volumes to be copied from each source volume, which can be used to create images from different PITs for each source volume.

FlashCopy is used to create consistent backups of dynamic data and test applications, and to create copies for auditing purposes and for data mining. It can be used to capture the data at a particular time to create consistent backups of dynamic data. The resulting image of the data can be backed up, for example, to a tape device. When the copied data is on tape, the data on the FlashCopy target disks becomes redundant and can be discarded.

Another possible FlashCopy application is creating test environments. FlashCopy can be used to test an application with real business data before the existing production version of the application is updated or replaced. With FlashCopy, a fully functional and space-efficient clone of a volume containing real data can be created. It enables read and write access for the test environment while keeping the real production environment data both safe and untouched. After testing is complete, the clone volume can be discarded or retained for future use.

FlashCopy can perform a restore from any existing FlashCopy mapping. Therefore, you can restore (or copy) from the target to the source of your regular FlashCopy relationships. When restoring data from FlashCopy, this method can be qualified as reversing the direction of the FlashCopy mappings. This approach can be used for various applications, such as recovering a production database application after an errant batch process that caused extensive damage.

Remote mirroring

You can use remote mirroring (also referred as Remote Copy (RC)) function to set up a relationship between two volumes, where updates made to one volume are mirrored on the other volume. The volumes can be on two different systems (intersystem) or on the same system (intrasystem).

For an RC relationship, one volume is designated as the primary and the other volume is designated as the secondary. Host applications write data to the primary volume, and updates to the primary volume are copied to the secondary volume. Normally, host applications do not run I/O operations to the secondary volume.

The following types of remote mirroring are available:

- ▶ **Metro Mirror (MM)**

Provides a consistent copy of a source volume on a target volume. Data is written to the target volume synchronously after it is written to the source volume so that the copy is continuously updated.

With synchronous copies, host applications write to the primary volume but do not receive a confirmation that the write operation completed until the data is written to the secondary volume, which ensures that both volumes have identical data when the copy operation completes. After the initial copy operation completes, the MM function maintains a fully synchronized copy of the source data at the target site always. The MM function supports copy operations between volumes that are separated by distances up to 300 km.

For DR purposes, MM provides the simplest way to maintain an identical copy on both the primary and secondary volumes. However, as with all synchronous copies over remote distances, there can be a performance impact to host applications. This performance impact is related to the distance between primary and secondary volumes and depending on application requirements, its use might be limited based on the distance between sites.

- ▶ **Global Mirror (GM)**

Provides a consistent copy of a source volume on a target volume. The data is written to the target volume asynchronously and the copy is continuously updated. When a host writes to the primary volume, a confirmation of I/O completion is received before the write operation completes for the copy on the secondary volume. Due to this situation, the copy might not contain the most recent updates when a DR operation is completed.

If a failover operation is initiated, the application must recover and apply any updates that were not committed to the secondary volume. If I/O operations on the primary volume are paused for a small length of time, the secondary volume can become an exact match of the primary volume. This function is comparable to a continuous backup process in which the last few updates are always missing. When you use GM for DR, you must consider how you want to handle these missing updates.

The secondary volume is generally less than 1 second behind the primary volume, which minimizes the amount of data that must be recovered if a failover occurs. However, a high-bandwidth link must be provisioned between the two sites.

- ▶ **Global Mirror with Change Volumes (GMCV)**

Enables support for GM with a higher recovery point objective (RPO) by using change volumes. This function is for use in environments where the available bandwidth between the sites is smaller than the update rate of the replicated workload.

With GMCV, or GM with cycling, change volumes must be configured for both the primary and secondary volumes in each relationship. A copy is taken of the primary volume in the relationship to the change volume. The background copy process reads data from the stable and consistent change volume, copying the data to the secondary volume in the relationship.

CoW technology is used to maintain the consistent image of the primary volume for the background copy process to read. The changes that took place while the background copy process was active are also tracked. The change volume for the secondary volume can also be used to maintain a consistent image of the secondary volume while the background copy process is active.

GMCV provides fewer requirements to inter-site link bandwidth than other RC types, and it is mostly used when link parameters are not sufficient to maintain RC relationship without impacting host performance.

Intersystem replication is possible over an FC or IP link. The native IP replication feature enables replication between any family systems running IBM Spectrum Virtualize by using the built-in networking ports of the system nodes.

Note: All three types of RC are supported to work over an IP link, but the recommended type is GMCV.

1.13.1 HyperSwap

The IBM HyperSwap function is a HA feature that provides dual-site, active-active access to a volume and is available on systems that can support more than one I/O group.

With HyperSwap, a fully independent copy of the data is maintained at each site. When data is written by hosts at either site, both copies are synchronously updated before the write operation is completed. The HyperSwap function automatically optimizes itself to minimize data that is transmitted between two sites, and to minimize host read and write latency.

If the system or the storage at either site goes offline and an online and accessible up-to-date copy is left, the HyperSwap function can automatically fail over access to the online copy. The HyperSwap function also automatically resynchronizes the two copies when possible.

To construct HyperSwap volumes, active-active replication relationships are made between the copies at each site. These relationships automatically run and switch direction according to which copy or copies are online and up to date. The relationships provide access to whichever copy is up to date through a single volume, which has a unique ID. This volume is visible as a single object across both sites (I/O groups), and is mounted to a host system.

The HyperSwap function works with the standard multipathing drivers that are available on a wide variety of host types, with no additional host support that is required to access the highly available (HA) volume. Where multipathing drivers support *Asymmetric Logical Unit Access (ALUA)*, the storage system tells the multipathing driver which nodes are closest to it and should be used to minimize I/O latency. You tell the storage system which site a host is connected to, and it configures host pathing optimally.

1.14 IBM FlashCore Module drives, NVMe SSDs, and SCM drives

This section describes the three types of flash drives that can be installed in the control enclosures:

- ▶ FCM drives
- ▶ NVMe SSDs
- ▶ Storage-class memory (SCM) drives

Note: The SCM drives and XL FCM drives require IBM Spectrum Virtualize V8.3.1. or later to be installed on the IBM FlashSystem Control Enclosure.

The following IBM FlashSystem products can support all three versions of these drives as follows:

- ▶ IBM FlashSystem 9200 system
- ▶ IBM FlashSystem 9200R Rack Solution system
- ▶ IBM FlashSystem 7200 system
- ▶ IBM FlashSystem 5100 system

They are not supported in any of the expansion enclosures.

Figure 1-41 shows an FCM (NVMe) with a capacity of 19.2 TB that is built by using 64-layer Triple Level Cell (TLC) flash memory and an Everspin MRAM cache into a U.2 form factor.



Figure 1-41 IBM FlashCore Module (NVMe)

FCM drives are designed for high parallelism and optimized for 3D TLC and updated FPGAs. IBM also enhanced the FCM drives by adding read cache to reduce latency on highly compressed pages, and added four-plane programming to lower the overall power during writes. FCM drives offer hardware-assisted compression up to 3:1 and are FIPS 140-2 compliant.

FCM drives carry IBM Variable Stripe RAID (VSR) at the FCM level and use DRAID to protect data at the system level. VSR and DRAID together optimize RAID rebuilds by offloading rebuilds to DRAID, and they offer protection against FCM failures.

Table 1-16 shows the capacities of the FCM type drives.

Table 1-16 FCM type capacities

FCM module size	Physical size (TBu)
Small	4.8 TBu
Medium	9.6 TBu
Large	19.2 TBu
XLarge	38.4 TBu

Industry-standard SSD NVMe drives

All the IBM FlashSystem models that are described in this book provide an option to use industry-standard SSD NVMe drives, which are sourced from Samsung and Toshiba and available in the several capacity variations, as shown in Table 1-17.

Table 1-17 NVMe drive size options

Drive type	Physical size (TBu)
NVMe Flash Drive	800 GB
NVMe Flash Drive	1.92 TB
NVMe Flash Drive	3.84 TB
NVMe Flash Drive	7.68 TB
NVMe Flash Drive	15.36 TB

NVMe and adapter support

NVMe is a NUMA-optimized, high-performance, and highly scalable storage protocol that is designed to access non-volatile storage media by using a host PCIe bus. NVMe uses low-latency and available parallelism, and reduces I/O impact. NVMe supports multiple I/O queues up to 64 K queues, and each queue can support up to 64 K entries. Earlier generations of SAS and Serial Advanced Technology Attachment (SATA) support a single queue with only 254 and 32 entries and use many more CPU cycles to access data. NVMe handles more workload for the same infrastructure footprint.

NVMe-oF is a technology specification that is designed to enable NVMe message-based commands to transfer data between a host computer and a target SSD or system. Data is transferred over a network, such as Ethernet, FC, or InfiniBand.

Storage-class memory

SCM drives use persistent memory technologies that improve endurance and reduce the latency of flash storage device technologies. All SCM drives use the NVMe architecture. IBM Research® is actively engaged in researching these new technologies.

For more information about nanoscale devices, see [Storage Class Memory at Almaden](#).

For a comprehensive overview of the flash drive technology see the [SNIA Educational Library](#).

These technologies will fundamentally change the architecture of today's storage infrastructures. Figure 1-42 shows the different types of storage technologies versus the latency for Intel drives.

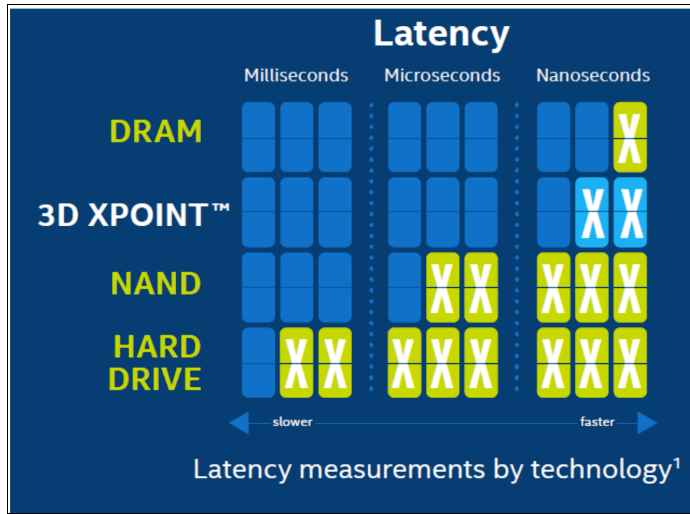


Figure 1-42 Storage technologies versus latency for Intel drives

IBM deploys two types of SCM class drives:

- ▶ 3D XPoint Intel Optane (375 GB and 750 GB)
- ▶ Z-SSD Samsung Z-NAND flash drive (800 GB and 1600 GB)

Table 1-18 shows the SCM drive size options.

Table 1-18 SCM drive options

Drive type	Physical size (TBu)
NVMe SCM Drive	375 GB
NVMe SCM Drive	750 GB
NVMe SCM Drive	800 GB
NVMe SCM Drive	1.6 TB

Easy Tier supports the SCM drives with a new tier that is called *tier_scm*.

Note: The SCM drive type supports only DRAID 6, DRAID 5, DRAID 1, and TRAIT 1 or 10.

1.15 Storage virtualization

Storage virtualization is a term that is used extensively throughout the storage industry. It can be applied to various technologies and underlying capabilities. In reality, most storage devices technically can claim to be virtualized in one form or another. Therefore, this chapter starts by defining the concept of storage virtualization as it is used in this book.

We describe storage virtualization in the following way:

- ▶ Storage virtualization is a technology that makes one set of resources resemble another set of resources, preferably with more wanted characteristics.
- ▶ Storage virtualization is a logical representation of resources that is not constrained by physical limitations and hides part of the complexity. It also adds or integrates new functions with services, and can be nested or applied to multiple layers of a system.

The virtualization model consists of the following layers:

- ▶ Application: The user of the storage domain.
- ▶ Storage domain:
 - File, record, and namespace virtualization, and file and record subsystem
 - Block virtualization
 - Block subsystem

Applications typically read and write data as vectors of bytes or records. However, storage presents data as vectors of blocks of a constant size (512 or in the newer devices, 4096 bytes per block).

The *file, record, and namespace virtualization* and *file and record subsystem* layers convert records or files that are required by applications to vectors of blocks, which are the language of the *block virtualization* layer. The block virtualization layer maps requests of the higher layers to physical storage blocks, which are provided by *storage devices* in the *block subsystem*.

Each of the layers in the storage domain abstracts away complexities of the lower layers and hides them behind an easy-to-use, standard interface that is presented to upper layers. The resultant decoupling of logical storage space representation and its characteristics that are visible to servers (storage consumers) from underlying complexities and intricacies of storage devices is a key concept of storage virtualization.

The focus of this publication is *block-level virtualization* at the *block virtualization layer*, which is implemented by IBM as IBM Spectrum Virtualize software that is running on an SVC and the IBM FlashSystem family. The SVC is implemented as a clustered appliance in the storage network layer. The IBM FlashSystem storage systems are deployed as modular systems that can virtualize their internally and externally attached storage.

IBM Spectrum Virtualize uses the SCSI protocol to communicate with its clients and presents storage space as SCSI logical units (LUs), which are identified by SCSI LUNs.

Note: Although LUs and LUNs are different entities, the term *LUN* in practice is often used to refer to a logical disk, that is, an LU.

Although most applications do not directly access storage but work with files or records, the operating system (OS) of a host must convert these abstractions to the language of storage, that is, vectors of storage blocks that are identified by LBAs within an LU.

Inside IBM Spectrum Virtualize, each of the externally visible LUs is internally represented by a volume, which is an amount of storage that is taken out of a storage pool. Storage pools are made of MDisks, that is, they are LUs that are presented to the storage system by external virtualized storage or arrays that consist of internal disks. LUs that are presented to IBM Spectrum Virtualize by external storage usually correspond to RAID arrays that are configured on that storage.

With storage virtualization, you can manage the mapping between logical blocks within an LU that is presented to a host, and blocks on physical drives. This mapping can be as simple or as complicated as required. A logical block can be mapped to one physical block, or for increased availability, multiple blocks that are physically stored on different physical storage systems, and in different geographical locations.

Importantly, the mapping can be dynamic: With Easy Tier, IBM Spectrum Virtualize can automatically change underlying storage to which groups of blocks (extent) are mapped to better match a host's performance requirements with the capabilities of the underlying storage systems.

IBM Spectrum Virtualize gives a storage administrator a wide range of options to modify volume characteristics, from volume resize to mirroring, creating a point-in-time (PiT) copy with FlashCopy, and migrating data across physical storage systems. Importantly, all the functions that are presented to the storage users are independent from the characteristics of the physical devices that are used to store data. This decoupling of the storage feature set from the underlying hardware and ability to present a single, uniform interface to storage users that masks underlying system complexity is a powerful argument for adopting storage virtualization with IBM Spectrum Virtualize.

IBM Spectrum Virtualize includes the following key features:

- ▶ Simplified storage management by providing a single management interface for multiple storage systems, and a consistent user interface for provisioning heterogeneous storage.
- ▶ Online volume migration. IBM Spectrum Virtualize enables moving the data from one set of physical drives to another set in a way that is not apparent to the storage users and without over-straining the storage infrastructure. The migration can be done within a specific storage system (from one set of disks to another set) or across storage systems. Either way, the host that uses the storage is not aware of the operation, and no downtime for applications is needed.
- ▶ Enterprise-level copy services functions. Performing copy services functions within IBM Spectrum Virtualize removes dependencies on the capabilities and interoperability of the virtualized storage subsystems. Therefore, it enables the source and target copies to be on any two virtualized storage subsystems.
- ▶ Improved storage space usage because of the pooling of resources across virtualized storage systems.
- ▶ Opportunity to improve system performance as a result of volume striping across multiple virtualized arrays or controllers, and the benefits of cache that is provided by IBM Spectrum Virtualize hardware.
- ▶ Improved data security by using data-at-rest encryption.
- ▶ Data replication, including replication to cloud storage by using advanced copy services for data migration and backup solutions.
- ▶ Data reduction techniques for space efficiency and cost reduction. Today, open systems typically use less than 50% of the provisioned storage capacity. IBM Spectrum Virtualize can enable savings, increase the effective capacity of storage systems up to five times, and decrease the floor space, power, and cooling that are required by the storage system.

IBM FlashSystem families are scalable solutions running on a HA platform that can use diverse back-end storage systems to provide all the benefits to various attached hosts.

External storage virtualization

You can use IBM FlashSystem to manage the capacity of other storage systems with external storage virtualization. When IBM FlashSystem virtualizes a storage system, its capacity is managed similarly to internal disk drives or flash modules. Capacity in external storage systems inherits all of the rich functions and ease of use of IBM FlashSystem.

You can use IBM FlashSystem to preserve your existing investments in storage, centralize management, and make storage migrations easier with storage virtualization and Easy Tier. Virtualization helps insulate applications from changes that are made to the physical storage infrastructure.

To verify whether your storage can be virtualized by IBM FlashSystem, see the [IBM System Storage Interoperation Center \(SSIC\)](#).

All the IBM FlashSystem family models can migrate data from external storage controllers, including migrating from any other IBM or third-party storage systems. IBM FlashSystem uses the functions that are provided by its external virtualization capability to perform the migration. This capability places external LUs under the control of an IBM FlashSystem system. Then, hosts continue to access them through the IBM FlashSystem system, which acts as a proxy.

The migration process typically consists of the following steps:

1. Input/output (I/O) to the LUs that exist on the external storage system must be stopped, and changes must be made to the mapping of the storage system so that the original LUs are presented directly to the IBM FlashSystem Family machine and not to the hosts. IBM FlashSystem discovers the external LUs and recognizes them as *unmanaged* external storage back-end devices (MDisks).
2. The unmanaged MDisks are imported to the IBM FlashSystem image mode volumes and placed in a migration storage pool. This storage pool is now a logical container for the externally attached LUs. Each volume has a one-to-one mapping with an external LU. From a data perspective, the image mode volume represents the SAN-attached LUs exactly as they were before the import operation. The image mode volumes are on the same physical drives of the storage system, and the data remains unchanged.
3. Your hosts are configured for IBM FlashSystem attachment, and image-mode volumes are mapped to them. After the volumes are mapped, the hosts discover their volumes and are ready to continue working with them so that I/O can be resumed.
4. Image-mode volumes are migrated to the internal storage of IBM FlashSystem by using the volume mirroring feature. Mirrored copies are created online so that a host can still access and use the volumes during the mirror synchronization process.
5. After the mirror operations are complete, the image mode volumes are removed (deleted), and external storage system can be disconnected and decommissioned or reused elsewhere.

The GUI of the IBM FlashSystem family provides a storage migration wizard, which simplifies the migration task. The wizard features intuitive steps that guide users through the entire process.

Note: The IBM FlashSystem 5010 and IBM FlashSystem 5030 systems do not support external virtualization for any other purpose other than data migration.

Summary

Storage virtualization is a fundamental technology that enables the realization of flexible and reliable storage solutions. It helps enterprises to better align their IT architecture with business requirements, simplify their storage administration, and facilitate their IT departments efforts to meet business demands.

IBM Spectrum Virtualize running on the IBM FlashSystem family is a mature, 10th-generation virtualization solution that uses open standards and complies with the SNIA storage model. All the products are appliance-based storage, and use in-band block virtualization engines that move the control logic (including advanced storage functions) from a multitude of individual storage devices to a centralized entity in the storage network.

IBM Spectrum Virtualize can improve the usage of your storage resources, simplify storage management, and improve the availability of business applications.

1.16 Business continuity

In today's online, highly connected, and fast-paced world, we expect that today's IT systems provide HA and continuous operations, and that they can be quickly recovered in the event of a disaster. Yet today's IT environment also features an ever-growing time to market pressure, with more projects to complete, more IT problems to solve, and a steep rise in time and resource limitations.

Thankfully, today's IT technology also features unprecedented levels of functions, features, and lowered cost. In many ways, it is easier than ever before to find IT technology that can address today's business concerns. This section describes some IBM FlashSystem storage solutions that can be applied to today's business continuity requirements.

1.16.1 Business continuity with HyperSwap

The HyperSwap HA feature in the IBM Spectrum Virtualize software enables business continuity during hardware failure, power failure, connectivity failure, or disasters, such as fire or flooding. The HyperSwap feature is available on the SVC and IBM FlashSystem products running IBM Spectrum Virtualize software.

The HyperSwap feature provides HA volumes that are accessible through two sites at up to 300 km apart. A fully independent copy of the data is maintained at each site. When data is written by hosts at either site, both copies are synchronously updated before the write operation is completed. The HyperSwap feature automatically optimizes itself to minimize data that is transmitted between sites and to minimize host read and write latency.

HyperSwap includes the following key features:

- ▶ Works with SVC and IBM FlashSystem products running IBM Spectrum Virtualize software.
- ▶ Uses intra-cluster synchronous RC (MM) capabilities along with existing change volume and access I/O group technologies.
- ▶ Makes a host's volumes accessible across two I/O groups in a clustered system by using the MM relationship in the background. They look like a single volume to the host.
- ▶ Works with the standard multipathing drivers that are available on a wide variety of host types, with no additional host support that is required to access the HA volume.

The following references provide you with further details about HyperSwap implementation use cases and guidelines:

- ▶ *IBM Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317
- ▶ *High Availability for Oracle Database with IBM PowerHA SystemMirror and IBM Spectrum Virtualize HyperSwap*, REDP-5459
- ▶ *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597

1.16.2 Business continuity with three-site replication

A three-site replication solution was made available in limited deployments for Version 8.3.1, where data is replicated from the primary site to two alternative sites, and the remaining two sites are aware of the difference between themselves. This solution ensures that in the event of a disaster at any one of the sites, the remaining two sites can establish a consistent_synchronized RC relationship among themselves with minimal data transfer, that is, within the expected RPO.

IBM Spectrum Virtualize V8.4 expands the three-site replication model to include HyperSwap, which improves data availability options in three-site implementations. Systems that are configured in a three-site topology have high DR capabilities, but a disaster might take the data offline until the system can be failed over to an alternative site. HyperSwap allows active-active configurations to maintain data availability, eliminating the need to failover if communications are disrupted. This solution provides a more robust environment, allowing up to 100% uptime for data, and recovery options inherent to DR solutions.

To better assist with three-site replication solutions, IBM Spectrum Virtualize 3-Site Orchestrator coordinates replication of data for DR and HA scenarios between systems.

IBM Spectrum Virtualize 3-Site Orchestrator is a command-line based application that runs on a separate Linux host that configures and manages supported replication configurations on IBM Spectrum Virtualize products.

Figure 1-43 shows the two supported topologies for the three-site replication coordinated solutions.

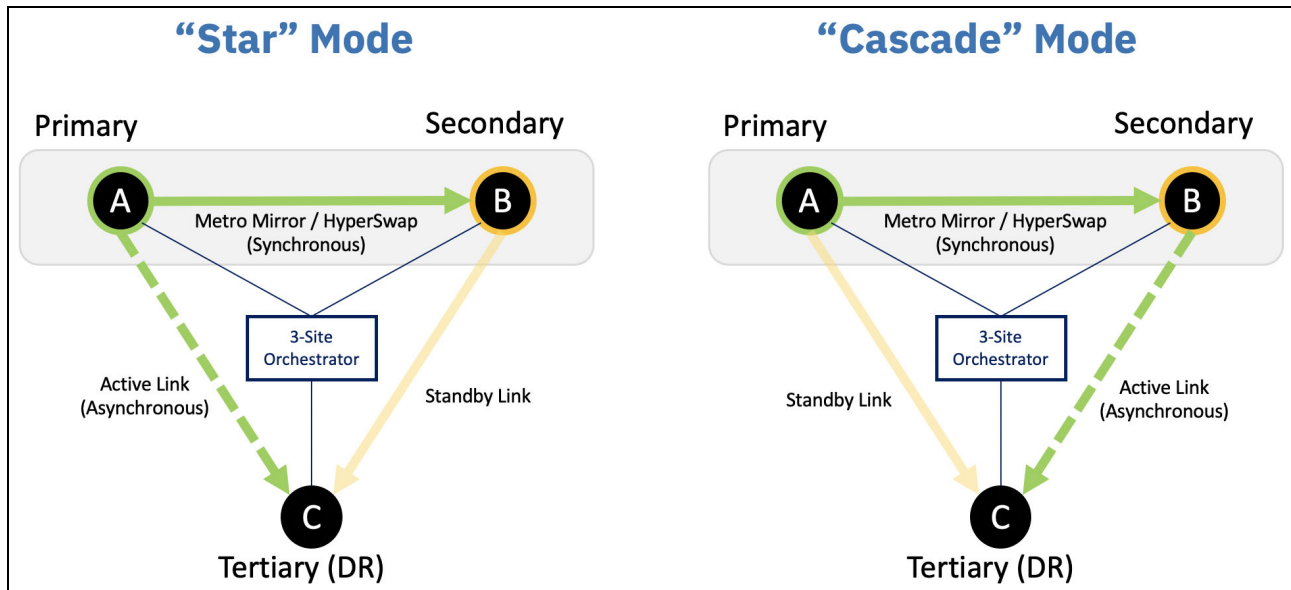


Figure 1-43 "Star" and "Cascade" modes in a three-site solution

For more information about this type of implementation, see *Spectrum Virtualize 3-Site Replication*, SG24-8474.

1.17 Licensing

All IBM FlashSystem functional capabilities are provided through IBM Spectrum Virtualize software, and each platform is licensed as described in the following sections.

1.17.1 Licensing IBM FlashSystem 9200, IBM FlashSystem 9200R, and IBM FlashSystem 7200

The IBM FlashSystem 9200 system has the same licensing scheme as the IBM FlashSystem 9200R system and the IBM FlashSystem 7200 system. They have all-inclusive licensing for all functions except encryption, which is a country-limited feature code, and external virtualization.

Any externally virtualized storage requires the External Virtualization license per storage capacity unit (SCU) that is based on the tier of storage that is available on the external storage system. In addition, if you use FlashCopy and Remote Mirroring on an external storage system, you must purchase a per-terabyte license to use these functions.

The SCU is defined in terms of the category of the storage capacity, as listed in Table 1-19 on page 67.

Table 1-19 SCU category definitions

License	Drive class	SCU ratio
SCM	SCM devices	SCU equates to 1.00 TiB usable of Category 1 storage.
Flash	All flash devices, other than SCM drives	SCU equates to 1.18 TiB usable of Category 2 storage.
Enterprise	10 K or 15 K RPM drives	SCU equates to 2 TiB usable of Category 3 storage.
NL	NL SATA drives	SCU equates to 4.00 TiB usable of Category 4 storage.

1.17.2 Licensing IBM FlashSystem 5100

The base license that is provided with the system includes its basic functions. However, there are also extra licenses that can be purchased to expand the capabilities of the system. Administrators are responsible for purchasing extra licenses and configuring the systems within the license agreement, which includes configuring the settings of each licensed function.

IBM FlashSystem 5100 licenses (enclosure-based)

IBM FlashSystem 5100 systems employ a license scheme that uses certain licensed functions that are based on the number of enclosures that are indicated in the license. The system supports the following licensed functions:

► External virtualization

The system does not require a license for its own control and expansion enclosures, but a license is required for each enclosure of any external systems that are being virtualized. Data can be migrated from existing storage systems to a system that uses the external virtualization function within 45 days of purchase of the system without purchasing a license.

After 45 days, any ongoing use of the external virtualization function requires a license for each enclosure in each external system. The system does not require an external virtualization license for external enclosures that are being used only to provide MDisks for a quorum disk and are not providing any capacity for volumes.

► Remote mirroring

The remote mirroring function sets up a relationship between two volumes so that updates that are made by an application to one volume are mirrored on the other volume. This function is licensed per enclosure. You can use the remote mirroring functions on the total number of enclosures that are licensed.

The total number of enclosures must include the enclosures on external storage systems that are licensed for virtualization and the number of control and expansion enclosures that are part of your local system. The remote mirroring option must be acquired for both the primary (local) and secondary (remote) systems. If the IBM FlashSystem 5100 system is mirrored to a system that is not an IBM FlashSystem 5100 system, the other system must have the appropriate and applicable license for remote mirroring.

- ▶ **Compression**

The compression function requires a separately orderable license that is set on a per enclosure basis. One license is required for each control or expansion enclosure and each enclosure in any external storage systems that use virtualization. With the compression function, data is compressed as it is written to disk, saving extra capacity for the system.

- ▶ **FlashCopy**

The FlashCopy function also requires a license to use, but it does not require any input on the system. For auditing purposes, retain the license agreement for proof of compliance.

In addition to these enclosure-based licensed functions, the system also supports encryption through a key-based license.

If you use a trial license, the system warns you when the trial is about to expire at regular intervals. If you do not purchase and activate the license on the system before the trial license expires, all configurations that use the trial licenses are suspended.

Encryption license (key-based)

Encryption is enabled on IBM FlashSystem 5100 systems by obtaining the Encryption Enablement feature. This feature enables encryption at system level and externally virtualized storage subsystems.

The encryption feature uses a key-based license that is activated by an authorization code. The authorization code is sent with the IBM FlashSystem 5100 Licensed Function Authorization documents that you receive after purchasing the license.

The Encryption USB Flash Drives (Four Pack) feature or an external key manager such as the IBM Security Key Lifecycle Manager are required for encryption keys management.

1.17.3 Licensing IBM FlashSystem 5030 and IBM FlashSystem 5010

The base license that is provided with the system includes its basic functions. However, extra licenses can be purchased to expand the capabilities of the system. Administrators are responsible for purchasing extra licenses and configuring the systems within the license agreement, which includes configuring the settings of each licensed function.

IBM FlashSystem 5000 licenses (key-based)

The IBM FlashSystem 5010 and IBM FlashSystem 5030 systems use key-based licensing in which an authorization code is used to activate licensed functions on the system. The authorization code is sent with the IBM FlashSystem 5000 Licensed Function Authorization documents that you receive after purchasing the license. These documents contain the authorization codes that are required to obtain keys (also known as *DFSA license keys*) for each licensed function that you purchased for your system. For each license that you purchase, a separate document with an authorization code is sent to you.

Each function is licensed to an IBM FlashSystem 5000 control enclosure. It covers the entire system (control enclosure and all attached expansion enclosures) if it consists of one I/O group. If the IBM FlashSystem 5030 system consists of two I/O groups, two keys are required.

The following functions need a license key before they can be activated on the system:

► Easy Tier

Easy Tier automatically and dynamically moves frequently accessed data to flash (solid-state) drives in the system, which results in flash drive performance without manually creating and managing storage tier policies. Easy Tier makes it easy and economical to deploy flash drives in the environment. In this dynamically tiered environment, data movement is seamless to the host application, regardless of the storage tier in which the data is stored.

► Remote Mirroring

The Remote Mirroring (also known as remote copy (RC)) function enables you to set up a relationship between two volumes so that updates that are made by an application to one volume are mirrored on the other volume.

The license settings apply to only the system on which you are configuring license settings. For RC partnerships, a license also is required on any remote systems that are in the partnership.

► FlashCopy upgrade

The FlashCopy upgrade extends the base FlashCopy function that is shipped with the product. The base version of FlashCopy limits the system to 64 target volumes. With the FlashCopy upgrade license activated on the system, this limit is removed. If you reach the limit that is imposed by the base function before activating the upgrade license, you cannot create more FlashCopy mappings.

To help evaluate the benefits of these new capabilities, Easy Tier and RC licensed functions can be enabled at no additional charge for a 90-day trial. Trials are started from the IBM FlashSystem management GUI and do not require any IBM intervention. When the trial expires, the function is automatically disabled unless a license key for that function is installed onto the machine.

If you use a trial license, the system warns you at regular intervals when the trial is about to expire. If you do not purchase and activate the license on the system before the trial license expires, all configurations that use the trial licenses are suspended.

Encryption license (key-based)

Encryption is enabled on IBM FlashSystem 5030 through the acquisition of the Encryption Enablement feature. This feature enables encryption on the entire IBM FlashSystem family system and externally virtualized storage subsystems.

Note: Encryption hardware feature is available only on the IBM FlashSystem 5030 (not on the IBM FlashSystem 5010).

This encryption feature uses a key-based license and is activated with an authorization code. The authorization code is sent with the IBM FlashSystem 5000 Licensed Function Authorization documents that you receive after purchasing the license.

The Encryption USB flash drives (Four Pack) feature or IBM Security Key Lifecycle Manager are required for encryption keys management.



Planning

This chapter describes the steps that are required to plan the installation and configuration of IBM FlashSystem systems in your storage network. Not all features that are described in this chapter are available and supported on all IBM FlashSystem systems. To learn which product features that are relevant to your IBM FlashSystem system are supported, see 1.3, “IBM FlashSystem family” on page 4.

This chapter is *not* intended to provide in-depth information about the described topics; it provides only general guidelines. For an enhanced analysis, see *IBM FlashSystem 9200 and 9100 Best Practices and Performance Guidelines*, SG24-8448, *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521, and *IBM FlashSystem 9100 Architecture, Performance, and Implementation*, SG24-8425.

Note: Make sure that the planned configuration is reviewed by IBM or an IBM Business Partner before implementation. Such a review can both increase the quality of the final solution and prevent configuration errors that might impact the solution delivery.

This chapter includes the following topics:

- ▶ 2.1, “General planning rules” on page 72
- ▶ 2.2, “Planning for availability” on page 73
- ▶ 2.3, “Physical installation planning” on page 73
- ▶ 2.4, “Planning for system management” on page 74
- ▶ 2.5, “Connectivity planning” on page 76
- ▶ 2.6, “Fibre Channel SAN configuration planning” on page 76
- ▶ 2.7, “IP SAN configuration planning” on page 83
- ▶ 2.9, “Back-end storage configuration” on page 88
- ▶ 2.10, “Internal storage configuration” on page 90
- ▶ 2.11, “Storage pool configuration” on page 92
- ▶ 2.12, “Volume configuration” on page 96
- ▶ 2.13, “Host attachment planning” on page 98
- ▶ 2.14, “Planning copy services” on page 100
- ▶ 2.15, “Data migration” on page 103
- ▶ 2.16, “Performance monitoring with IBM Storage Insights” on page 104
- ▶ 2.17, “Configuration backup procedure” on page 106

2.1 General planning rules

To maximize the benefit from a system, installation planning must include several important steps. These steps ensure that the system provides the best possible performance, reliability, and ease of management for your application needs.

The general rule of planning is to define your goals and then plan a solution that makes you able to reach these goals.

Consider the following points when planning a system:

- ▶ Collect and document information about application servers (hosts) that you want to attach to the system and their data:
 - Amount of data in use for each host and growth plans.
 - Data profile: Compressibility and deduplicability.
 - Host traffic profile: Percentage of reads and writes, percentage of sequential/random access patterns, and data block size.
 - Host performance requirements: Input/output operations per second (IOPS) and bandwidth.
- ▶ Perform capacity and performance sizing of a system:
 - If any external back-end systems are going to be virtualized, assess their capacity and performance capabilities.
 - Calculate the number of drives or IBM FlashCore Module (FCM) drives that are needed to satisfy your capacity requirements by considering your data compression ratios and accounting for future growth.
 - Verify that the capacity assessment results satisfy your performance requirements.

Note: Contact your IBM sales representative or IBM Business Partner to perform these calculations.

- ▶ Assess your recovery point objective (RPO) / recovery time objective (RTO) requirements and plan for high availability (HA) and Remote Copy (RC) functions. Decide whether you require a dual-site or three-site deployment, and decide whether you must implement RC and determine its type (synchronous or asynchronous). Review the extra configuration requirements that are imposed.
- ▶ Define the number of input/output (I/O) groups (control enclosures) and expansion enclosures. The number of necessary enclosures depends on the solution type, overall performance, and capacity requirements.
- ▶ Plan for host attachment interfaces, protocols, and storage area network (SAN). Consider the number of ports, bandwidth requirements, and HA.
- ▶ Perform configuration planning by defining the number of internal storage arrays and external storage arrays that will be virtualized. Define a number and the type of pools, the number of volumes, and the capacity of each of the volumes.
- ▶ Define a naming convention for the system nodes, volumes, and other storage objects.
- ▶ Plan a management IP network and management users' authentication system.
- ▶ Plan for the physical location of the equipment in the rack.
- ▶ Verify that your planned environment is a supported configuration.

Note: Use [IBM System Storage Interoperation Center \(SSIC\)](#) to check compatibility.

- ▶ Verify that your planned environment does not exceed system configuration limits.

Note: For more information about your platform and code version, see [Configuration Limits and Restrictions](#).

- ▶ Review the planning aspects that are described in the following sections of this chapter.

2.2 Planning for availability

When planning the deployment of the IBM FlashSystem family solution, avoid creating single points of failure (SPOFs). Plan your system availability according to the requirements of your solution. Depending on your availability needs, consider the following aspects:

- ▶ Single-site or multi-site configuration

Multi-site configurations increase solution resiliency, and can be the basis of disaster recovery (DR) solutions. Systems can be configured as a multi-site solution with sites working in active-active mode. Both synchronous and asynchronous data replication are supported by multiple inter-site link options. With IBM Spectrum Virtualize V8.4, three-site replication deployments are supported.

- ▶ Physical separation of system building blocks

A dual-rack deployment might increase the availability of your system if your back-end storage, SAN, and local area network (LAN) infrastructure also do not use a single-rack placement scheme. You can further increase system availability by ensuring that enclosures are powered from different power circuits and in different fire protection zones.

- ▶ Quorum disk placement

For a deployment with multiple I/O groups, plan for a quorum device on an external back-end system or an IP quorum application. IP quorum applications must be deployed on hosts that do not depend on storage that is provisioned by a system. Multiple IP quorum application deployment is recommended.

2.3 Physical installation planning

You must consider several key factors when you plan the physical site of a system. The physical site must have the following characteristics:

- ▶ Sufficient rack space must exist to install controller and disk enclosures.
- ▶ The site must meet the power, cooling, and environmental requirements.

For more information about power and environmental requirements, see the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, to see the IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and expand **Planning** → **Planning for hardware** → **Physical installation planning**, and then select **Connections for control enclosures** and **SAS expansion enclosure requirements**.

Your system order includes a printed copy of the *Quick Installation Guide*, which also provides information about environmental and power requirements.

Create a cable connection table that follows your environment's documentation procedure to track the following connections that are required for the setup:

- ▶ Power
- ▶ Serial-attached Small Computer System Interface (SCSI) (SAS)
- ▶ Ethernet
- ▶ Fibre Channel (FC)

When planning for power, plan for a separate independent power source for each of the two redundant power supplies of a system enclosure.

Distribute your expansion enclosures between control enclosures and SAS chains, as described in 13.1.4, "Enclosure SAS cabling" on page 801. For more information, see the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, to see the IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and expand **Installing** → **Connecting the components** → **Connecting 2U expansion enclosures to the control enclosure**.

When planning SAN cabling, make sure that your physical topology adheres to zoning rules and recommendations.

The physical installation and initial setup of IBM FlashSystem 9100 and IBM FlashSystem 9200 is performed by an IBM System Services Representative (IBM SSR).

IBM FlashSystem 7200, IBM FlashSystem 5100, IBM FlashSystem 5030, and IBM FlashSystem 5010 are classified as Customer Setup Units (CSUs), and the physical installation and initial setup is the responsibility of the customer. IBM can be contracted to perform these services for a fee.

2.4 Planning for system management

Each system's node has a *technician port*. It is a dedicated 1 gigabits per second (Gbps) Ethernet port. The initialization of a system and its basic configuration is performed by using this port. After the initialization is complete, the technician port must remain disconnected from a network and used only to service the system.

On the IBM FlashSystem 5010, as opposed to other platforms, the technician port is not dedicated. On those systems, after the initial configuration, it is converted to a regular Ethernet port that can be connected to the network and used for management tasks and to serve I/O to hosts with internet Small Computer Systems Interface (iSCSI).

For management, each system node requires at least one Ethernet connection. The cable must be connected to port 1, which is a 10 Gbps Ethernet port (it does not negotiate speeds below 1 Gbps). For increased availability, an optional management connection may be configured over Ethernet port 2.

For configuration and management, you must allocate an IP address to each node canister, which is referred to as the *service IP address*. Both IPv4 and IPv6 are supported.

In addition to a service IP address on each node, each system has a *cluster management IP address*. The cluster management IP address cannot be the same as any of the defined service IP addresses. The cluster management IP can automatically fail over between cluster nodes if there are maintenance actions or a node failure.

For example, a system that consists of two control enclosures requires a minimum of five unique IP addresses: one for each node and one for the system as a whole.

Ethernet ports 1 and 2 are not reserved only for management. They may be also used for iSCSI or IP replication traffic if they are configured to do so. However, management and service IP addresses cannot be used for host or back-end storage communication.

System management is performed by using an embedded GUI that is running on the nodes; the command-line interface (CLI) is also available. To access the management GUI, point a web browser to the cluster management IP address. To access the management CLI, point a Secure Shell (SSH) client to a cluster management IP and use the default SSH protocol port (22/TCP).

By connecting to a service IP address with a browser or SSH client, you can access the *Service Assistant Interface*, which may be used for maintenance and service tasks.

When you plan your management network, note that the IP Quorum applications and Transparent Cloud Tiering (TCT) are communicating with a system through the management ports. For more information about cloud backup requirements, see 10.3, “Transparent Cloud Tiering” on page 621.

2.4.1 User password creation options

IBM Spectrum Virtualize V8.4 has a new password policy support feature that allows system administrators to set security requirements. These requirements are related to password creation and expiration, timeout for inactivity, and actions after failed logon attempts.

Password policy support allows administrators to set security rules that are based on your organization's security guidelines and restrictions. The system supports the password and security-related rules that are described in the following subsections.

Password creation rules

Administrator can set and manage the following rules for all passwords that are created on the system:

- ▶ Specify password length requirements for all users.
- ▶ Require passwords to use uppercase and lowercase characters.
- ▶ Require passwords to contain special characters.
- ▶ Prevent users from reusing recent passwords.
- ▶ Require users to change their password on next login under any of these conditions:
 - Their password expired.
 - An administrator created accounts with temporary passwords.
- ▶ Password history checking can be enabled.
- ▶ The minimum required password age can be set to prevent bypassing the password history restriction by rapidly changing passwords multiple times.

A new policy does not apply retroactively to existing passwords.

Password expiration and account locking rules

The administrator can create the following rules for password expiration:

- ▶ Set a password expiration limit.

- ▶ Set a password to expire immediately.
- ▶ Set number of failed login attempts before the account is locked.
- ▶ Set a period for locked accounts.
- ▶ Automatic log out for inactivity.
- ▶ Locking superuser account access.

Note: Systems that support a dedicated technician port can lock the superuser account. The superuser account is the default user that can complete installation, initial configuration, and other service-related actions on the system. If the superuser account is locked, service tasks cannot be completed.

For more information about implementing these features, see Chapter 4, “IBM Spectrum Virtualize GUI” on page 155.

2.5 Connectivity planning

An IBM FlashSystem system offers a wide range of connectivity options to back-end storage and hosts, such as FC technologies (“traditional” SCSI FC and Non-Volatile Memory Express over Fibre Channel (FC-NVMe) (also known as NVMe over Fabric (NVMe-oF))), IP network technologies (iSCSI and iSCSI Extensions for Remote Direct Memory Access (RDMA) (iSER)) and SAS technologies. The connection options and capabilities depend on the hardware configuration.

Table 2-1 lists the communication types that can be used for communicating between system nodes, hosts, and back-end storage systems. All types can be used concurrently.

Table 2-1 Communication options

Communication type	System to host	System to back-end storage	Node to node (intra-cluster)	System to system (replication)
SCSI FC	Yes	Yes	Yes	Yes
FC-NVMe	Yes	No	No	No
iSCSI	Yes	Yes	No	No ^a
iSER	Yes	No	Yes	No ^a
SAS	Yes ^b	Yes ^c		

- a. Replication traffic can be sent over an IP network with native IP replication, which can be configured on both onboard 10 Gb Ethernet (GbE) ports and optional 25 GbE ports.
- b. SAS host attachment is available only on IBM FlashSystem 5010 and IBM FlashSystem 5030.
- c. Back-end storage attachment is supported for data migration only.

2.6 Fibre Channel SAN configuration planning

Each node canister may be equipped with one, two, or three 4-port 16 Gbps or 32 Gbps FC adapters (the maximum number of adapters depends on the system hardware type) that are used for SCSI FC and FC-NVMe attachment.

2.6.1 Physical topology

The switch configuration for a fabric must comply with the switch manufacturer's configuration rules, which can impose restrictions. For example, a switch manufacturer might limit the number of supported switches or ports in a SAN fabric. Operating outside of the switch manufacturer's rules is not supported.

In an environment where you have a fabric with mixed port speeds (8 Gb, 16 Gb, and 32 Gb), the best practice is to connect the system to the switch operating at the highest speed.

The connections between the system's enclosures (node-to-node traffic) and between a system and the virtualized back-end storage require the best available bandwidth. For optimal performance and reliability, ensure that paths between the system nodes and storage systems do not cross inter-switch links (ISLs). If you use ISLs on these paths, make sure that sufficient bandwidth is available. SAN monitoring is required to identify faulty ISLs.

No more than three ISL hops are permitted among nodes that are in the same system but in different I/O groups. If your configuration requires more than three ISL hops for nodes that are in the same system but in different I/O groups, contact your IBM Support Center.

Direct connection of the system FC ports to host systems or between nodes in the system without using an FC switch is supported. For more information, see the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, for IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and expand **Planning** → **Planning your network and storage network** → **Planning for a direct-attached configuration**.

For the planning and topology requirements for HyperSwap configurations, see *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597 and *IBM Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317.

For the planning and topology requirements for three-site replication configurations, see *Spectrum Virtualize 3-Site Replication*, SG24-8474.

2.6.2 Zoning

A SAN fabric must have four distinct zone classes:

Inter-node zones	For communication between nodes in the same system
Storage zones	For communication between the system and back-end storage
Host zones	For communication between the system and hosts
Inter-system zones	For remote replication

Figure 2-1 shows the system zoning classes.

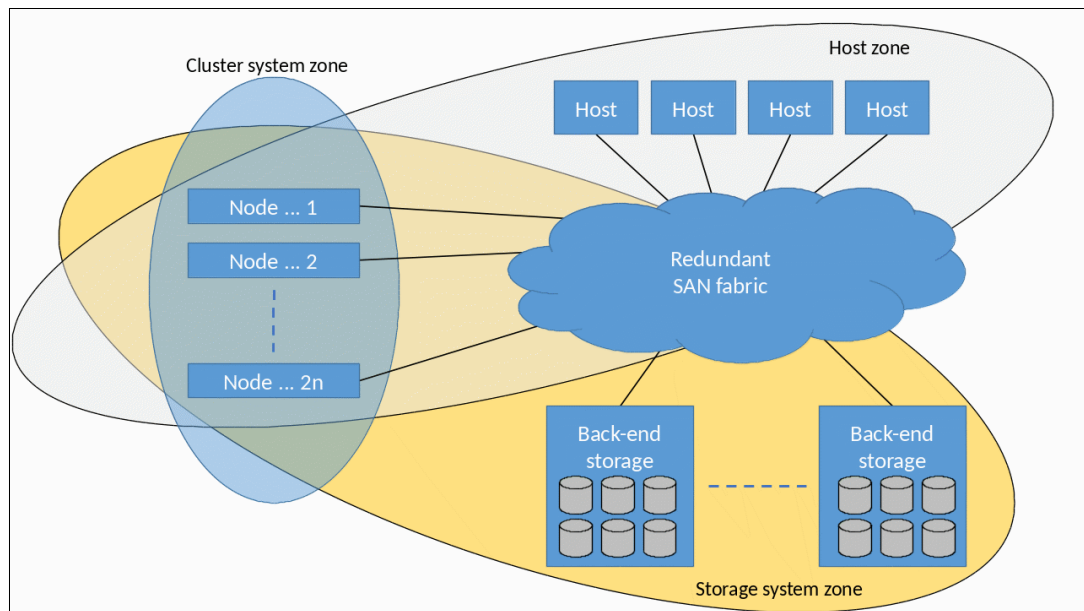


Figure 2-1 System zoning

The fundamental rules of system zoning are described in the rest of this section. However, you must review the latest zoning guidelines and requirements when designing zoning for the planned solution by reviewing the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, for the IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and expand **Configuring** → **Configuration details** → **SAN configuration and zoning rules summary**.

2.6.3 N_Port ID Virtualization

N_Port ID Virtualization (NPIV) is a method for virtualizing a physical FC port that is used for host I/O. By default, all new systems work in NPIV mode (the Target Port Mode attribute is set to Enabled).

NPIV mode creates a virtual worldwide port name (WWPN) for every system physical FC port. This WWPN is available only for host connection. During node maintenance, restart, or failure, the virtual WWPN from that node is transferred to the same port of the other node in the I/O group.

For more information about NPIV mode and how it works, see Chapter 7, “Hosts” on page 405.

Ensure that the FC switches give each physically connected system port the ability to create four more NPIV ports.

When performing zoning configuration, virtual WWPNs are used only for host communication, that is, “system to host” zones must include virtual WWPNs, and internode, intersystem, and back-end storage zones must use the WWPNs of physical ports. Ensure that equivalent ports (with the same port ID) are on the same fabric and in the same zone.

For more information about other host zoning requirements, see the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, for the IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and expand **Configuring** → **Configuration details** → **Zoning details** → **Zoning requirements for N_Port ID Virtualization**.

2.6.4 Inter-node zone

The purpose of intracluster or inter-node zones is to enable traffic between all node canisters within the clustered system. This traffic consists of heartbeats, cache synchronization, and other data that nodes must exchange to maintain a healthy cluster state.

Traffic between nodes in one control enclosure is sent over a Peripheral Component Interconnect Express (PCIe) connection over an enclosure backplane. However, for redundancy, you must configure an inter-node SAN zone even if you have a single I/O group system. For a system with multiple I/O groups, all traffic between control enclosures must pass through a SAN.

A system node cannot have more than 16 fabric paths to another node in the same system.

2.6.5 Back-end storage zones

Create a separate zone for each back-end storage subsystem that is virtualized. Switch zones that contain back-end storage system ports must not have more than 40 ports. A configuration that exceeds 40 ports is not supported.

All nodes in a system must connect to the same set of back-end storage system ports on each device.

If the edge devices contain more stringent zoning requirements, follow the storage system rules to further restrict the system zoning rules.

Note: Cisco Smart Zoning and Brocade Peer Zoning are supported, which let you put target ports and multiple initiator ports in a single zone for easy of management but act the same as though each initiator and target are configured in isolated zones. Using these zoning techniques are supported for both host attachment and for storage virtualization. As a best practice, use normal zones when configuring ports for clustering or for replication because these functions require the port to be both an initiator and a target.

For more information connecting back-end storage systems, see the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, for IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and expand **Configuring** → **Configuration details** → **External storage system configuration details (Fibre Channel)** and **Configuring** → **Configuring and servicing storage systems** → **External storage system configuration with Fibre Channel connections**.

2.6.6 Host zones

A host must be zoned to an I/O group to access volumes that are presented by this I/O group.

The preferred zoning policy is *single initiator zoning*. To implement it, create a separate zone for each host bus adapter (HBA) port, and place exactly one port from each node in each I/O group that the host accesses in this zone. For deployments with more than 64 hosts that are defined in the system, this host zoning scheme is mandatory.

Note: Cisco Smart Zoning and Brocade Peer Zoning are supported, which let you put target ports and multiple initiator ports in a single zone for easy of management but act the same as though each initiator and target are configured in isolated zones. Using these zoning techniques are supported for both host attachment and for storage virtualization. As a best practice, use normal zones when configuring ports for clustering or for replication because these functions require the port to be both an initiator and a target.

For smaller installations, you may have up to 40 FC ports (including both host HBA ports and the system's virtual WWPNs) in a host zone if the zone contains similar HBAs and operating systems (OSs). A valid zone can be 32 host ports plus eight system ports.

FC-NVMe applies more limits to the host zone configuration:

- ▶ Zone up to four host ports to detect up to four ports on a node, and zone the same or more host ports to detect an extra four ports on the second node of the I/O group.
- ▶ Zone a total maximum of 16 hosts to detect a single I/O group.

Consider the following rules for zoning hosts over either SCSI or FC-NVMe:

- ▶ For any volume, the number of paths through the SAN from the host to a system must not exceed eight. For most configurations, four paths to an I/O group are sufficient.

In addition to zoning, you can use a *port mask* to control the number of host paths. For more information, see 3.4.5, "Configuring the local Fibre Channel port masking" on page 134.

- ▶ Balance the host load across the system's ports. For example, zone the first host with ports 1 and 3 of each node in I/O group, zone the second host with ports 2 and 4, and so on. To obtain the best overall performance of the system, the load of each port should be equal. Assuming that a similar load is generated by each host, you can achieve this balance by zoning approximately the same number of host ports to each port.
- ▶ Spread the load across all system ports. Use all ports that are available on your machine.
- ▶ Balance the host load across HBA ports. If the host has more than one HBA port per fabric, zone each host port with a separate group of system ports.

All paths must be managed by the multipath driver on the host side. Make sure that the multipath driver on each server can handle the number of paths that is required to access all volumes that are mapped to the host.

2.6.7 Zoning considerations for Metro Mirror and Global Mirror

The SAN configurations that use inter-cluster Metro Mirror (MM) and Global Mirror (GM) relationships have the following extra switch zoning requirements:

- ▶ If two ISLs are connecting the sites, split the ports from each node between the ISLs, that is, exactly one port from each node must be zoned across each ISL.
- ▶ Local clustered system zoning continues to follow the standard requirement for all ports on all nodes in a clustered system to be zoned to one another.
- ▶ Review the latest requirements and recommendations in the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, for the IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and select **Configuring** → **Configuration details** → **Zoning details** → **Zoning constraints for Metro Mirror and Global Mirror**.

When designing zoning for a geographically dispersed solution, consider the effect of the cross-site links on the performance of the local system.

Using mixed port speeds for intercluster communication can lead to port congestion, which can negatively affect the performance and resiliency of the SAN. Therefore, it is not supported.

Note: If you limit the number of ports that are used for remote replication to two ports on each node, you can limit the effect of a severe and abrupt overload of the intercluster link on system operations.

If all node ports (N_Ports) are zoned for intercluster communication and the intercluster link becomes severely and abruptly overloaded, the local FC fabric can become congested so that no FC ports on the local system can perform local intracluster communication, which can result in cluster consistency disruption.

For more information about how to avoid such situations, see 2.6.8, “Port designation recommendations” on page 81.

For more information about zoning best practices, see *IBM FlashSystem 9200 and 9100 Best Practices and Performance Guidelines*, SG24-8448, and *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521.

2.6.8 Port designation recommendations

If you have enough available FC ports on the system, designate different types of traffic to different ports. This configuration provides a level of protection against malfunctioning devices and workload spikes that might otherwise impact the system.

Intra-cluster communication must be protected because it is used for heartbeat and metadata exchange between all nodes of all I/O groups of the cluster.

In solutions with multiple I/O groups, upgrade nodes beyond the standard four FC port configuration. This upgrade provides an opportunity to dedicate ports to local node traffic, which separates them from other cluster traffic on the remaining ports.

Isolating remote replication traffic to dedicated ports is beneficial because it ensures that any problems that affect the cluster-to-cluster interconnect do not affect all ports on the local cluster.

To isolate both node-to-node and system-to-system traffic, use the port designations that are shown in Figure 2-2.

Card/Port	4 port	8 port	12 port	SAN Fabric
Adapter 1 Port 1	Host+Storage	Host+Storage	Host+Storage	A
Adapter 1 Port 2	Host+Storage	Host+Storage	Host+Storage	B
Adapter 1 Port 3	Intracuster+Replication	Intracuster	Intracuster	A
Adapter 1 Port 4	Intracuster+Replication	Intracuster	Intracuster	B
Adapter 2 Port 1		Host+Storage	Host+Storage	A
Adapter 2 Port 2		Host+Storage	Host+Storage	B
Adapter 2 Port 3		Intracuster or Replication	Replication or Host+Storage	A
Adapter 2 Port 4		Intracuster or Replication	Replication or Host+Storage	B
Adapter 3 Port 1			Host+Storage	A
Adapter 3 Port 2			Host+Storage	B
Adapter 3 Port 3			Intracuster	A
Adapter 3 Port 4			Intracuster	B
Adapter 4 Port 1				A
Adapter 4 Port 2				B
Adapter 4 Port 3				A
Adapter 4 Port 4				B
localfcportmask	1100	11001100 OR 00001100	110000001100	
remotefcportmask	1100	00000000 OR 11000000	000011000000	
<p>Host refers to host objects defined in the system.</p> <p>Storage refers to controller objects defined in the system if external storage is being used.</p> <p>Replication refers to nodes which are part of a different cluster.</p> <p>Intracuster refers to nodes within the same cluster.</p> <p>The "+" indicates that both types are should to be used</p> <p>The word "or" indicates that one of the options must be selected. If using replication, preference should be given to replication.</p>				

Figure 2-2 Port masking configuration

When planning masking, consider the following examples:

- ▶ A system with a single control enclosure (I/O group) and without replication: No port dedication and masking are required. Inter-node traffic is sent over a backplane.
- ▶ A HyperSwap system with two control enclosures: Dedicate ports for inter-node traffic and apply an FC mask.
- ▶ A standard topology system with four I/O groups: The masking setup depends on the storage configuration, so more planning is required.

To achieve traffic isolation, use a combination of SAN zoning and *local and partner port masking*. For more information about how to send port masks, see Chapter 3, "Initial configuration" on page 107.

Alternative port mappings that spread traffic across HBAs might allow adapters to come back online after a failure. However, they do not prevent a node from going offline temporarily to restart and attempt to isolate the failed adapter and then rejoin the cluster. Also, the mean time between failures (MTBF) of the adapter is not significantly shorter than that of the non-redundant node components. The approach that is presented here accounts for all these considerations with the idea that increased complexity can lead to migration challenges in the future, so a simpler approach is better.

2.7 IP SAN configuration planning

Each system node is equipped with four onboard 10 Gbps Ethernet network interface ports. They can operate with link speeds of 1 Gbps and 10 Gbps. Any of them can be used for host I/O with the iSCSI protocol, external storage virtualization with iSCSI, and for native IP replication. Also, ports 1 and 2 may be used for managing the system.

Each node may also be configured with one, two, or three 2-port 25 Gbps RDMA-capable Ethernet adapters. The maximum number of adapters depends on the system hardware type. Adapters can auto-negotiate link speeds 1 - 25 Gbps. All their ports may be used for host I/O with iSCSI or iSER, external storage virtualization with iSCSI, node-to-node traffic, and for IP replication.

With IBM Spectrum Virtualize V8.4, support for 10 Gbps Finisar small form factor pluggable (SFP) (Finisar FTLX8574D3BCL) on the Mellanox and Chelsio 25 Gbps Ethernet adapters is introduced.

Note: At the time of writing, only the 10 Gbps Finisar SFP is supported on the 25 GbE adapters. In all other instances, connecting a 10 Gbps switch to a 25 Gbps interface is supported only through a SCORE request. For more information, contact your IBM representative.

You can set virtual local area network (VLAN) settings to separate network traffic for Ethernet transport. The system supports VLAN configurations for the system, host attachment, storage virtualization, and IP replication traffic. VLANs can be used with priority flow control (PFC) (IEEE 802.1Qbb).

All ports may be configured with an IPv4 address, an IPv6 address, or both. Each application of a port needs a separate IP. For example, port 1 of every node can be used for management, iSCSI, and IP replication, but three unique IP addresses are required.

If node Ethernet ports are connected to different isolated networks, then a different subnet must be used for each network.

2.7.1 iSCSI and iSER protocols

The iSCSI protocol is a block-level access protocol that encapsulates SCSI commands into TCP/IP packets. Therefore, iSCSI uses an IP network rather than requiring the FC infrastructure.

The iSER is a network protocol that extends iSCSI to use RDMA. RDMA is provided by either the internet Wide Area RDMA Protocol (iWARP) or RDMA over Converged Ethernet (RoCE). It permits data to be transferred directly into and out of SCSI buffers, providing faster connection and processing time than traditional iSCSI connections.

iSER requires optional 25 Gbps RDMA-capable Ethernet cards. RDMA links work only between RoCE ports or between iWARP ports: from a RoCE node canister port to a RoCE port on a host, or from an iWARP node canister port to an iWARP port on a host. So, there are two types of 25 Gbps adapters that are available for a system, and they cannot be interchanged without a similar RDMA type change on the host side.

Either iSCSI or iSER works for standard iSCSI communications, that is, ones that do not use RDMA.

The 25 Gbps adapters come with SFP28 fitted, which can be used to connect to switches that use OM3 optical cables.

For more information about the Ethernet switches and adapters that are supported by iSER adapters, see [SSIC](#).

With IBM Spectrum Virtualize V8.4, support for 10 Gbps Finisar SFP (Finisar FTLX8574D3BCL) on the Mellanox and Chelsio 25 Gbps Ethernet adapters is introduced.

Note: At the time of writing, only the 10 Gbps Finisar SFP is supported on the 25 Gbps Ethernet adapters. In all other instances, connecting a 10 Gbps switch to a 25 Gbps interface is supported only through a SCORE request. For more information, contact your IBM representative.

2.7.2 Priority flow control

PFC is an Ethernet protocol that you can use to select the priority of different types of traffic within the network. With PFC, administrators can reduce network congestion by slowing or pausing certain classes of traffic on ports, thus providing better bandwidth for more important traffic. The system supports PFC on various supported Ethernet-based protocols on three types of traffic classes: system (node-to-node), host attachment, and back-end storage traffic.

You can configure a priority tag for each of these traffic classes. The priority tag can be any value 0 - 7. You can set identical or different priority tag values to all these traffic classes. You can also set bandwidth limits to ensure quality of service (QoS) for these traffic classes by using the Enhanced Transmission Selection (ETS) setting on the network.

To use PFC and ETS, ensure that the following tasks are completed:

- ▶ Configure a VLAN on the system to use PFC capabilities for the configured IP version.
- ▶ Ensure that the same VLAN settings are configured on the all entities, including all switches between the communicating end points.
- ▶ On the switch, enable Data Center Bridging Exchange (DCBx). DCBx enables switch and adapter ports to exchange parameters that describe traffic classes and PFC capabilities. For these steps, check your switch documentation for details.
- ▶ For each supported traffic class, configure the same priority tag on the switch. For example, if you plan to have a priority tag setting of 3 for storage traffic, ensure that the priority is also set to 3 on the switch for that traffic type.
- ▶ If you are planning on using the same port for different types of traffic, ensure that ETS settings are configured on the network.

For more information, see the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, for the IBM FlashSystem 9200 related information, see [IBM FlashSystem 9200 documentation](#) and expand **Configuring** → **Configuring priority flow control**.

2.7.3 RDMA clustering

An IBM FlashSystem system may use 25 Gbps cards for node-to-node traffic. A dual-site HyperSwap configuration can also use the cards for an inter-site link.

A minimum of two dedicated RDMA-capable ports are required for node-to-node RDMA communications to ensure best performance and reliability. These ports must be configured for inter-node traffic only and cannot be used for host attachment, virtualization of Ethernet-attached external storage, or IP replication traffic.

Note: RDMA clustering is not supported on IBM FlashSystem 5010 or IBM FlashSystem 5030.

The following limitations apply to a configuration of ports that are used for RDMA-clustering:

- ▶ Only IPv4 addresses are supported.
- ▶ Only the default value of 1500 is supported for the maximum transmission unit (MTU).
- ▶ Port masking is not supported on RDMA-capable Ethernet ports. Due to this limitation, do not exceed the maximum of four ports for node-to-node communications.
- ▶ Node-to-node communications that use RDMA-capable Ethernet ports are not supported in a network configuration that contain more than two hops in the fabric of switches.
- ▶ Some environments might not include a stretched layer 2 subnet. In such scenarios, a layer 3 network such as in standard topologies or long-distance RDMA node-to-node HyperSwap configurations is applicable. To support the layer 3 Ethernet network, the unicast discovery method can be employed for RDMA node-to-node communication. This method relies on unicast-based fabric discovery rather than multicast discovery. To configure unicast discovery, see the man pages for the `addnodediscoverysubnet`, `rmnodediscoverysubnet`, or `l1nodediscoverysubnet` commands.

For more information, see the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, for the IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and expand **Configuring** → **Configuration details** → **Configuration details for using RDMA-capable Ethernet ports for node-to-node communications**.

Note: Before you configure a system that uses RDMA-capable Ethernet ports for node-to-node communications in a standard or HyperSwap topology system, contact your IBM representative.

2.7.4 iSCSI back-end storage attachment

An IBM FlashSystem system supports the virtualization of external storage systems that are attached through iSCSI. Onboard 10 Gbps Ethernet ports or optional 25 Gbps Ethernet ports may be used. The 25 GbE network interface controllers (NICs) work in plain iSCSI mode without using any RDMA capabilities.

Consider the following items when planning for iSCSI virtualization:

- ▶ A one-to-one mapping of source ports to target ports is required.
- ▶ Direct attachment between the system and external storage systems is not supported, and requires Ethernet switches between the system and the external storage.

- ▶ To avoid a SPOF, a dual-switch configuration is recommended. For full redundancy, a minimum of two paths between each initiator node and target node must be configured with each path going through a separate switch.
- ▶ Extra paths can be configured to increase throughput if both initiator and target nodes support more ports.

All planning and implementation aspects of external storage virtualization with iSCSI are described in detail in *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327.

2.7.5 IP network host attachment

You can attach the system to iSCSI or iSER hosts by using the Ethernet ports of the systems.

For each Ethernet port on a node, a maximum of one IPv4 address and one IPv6 address can be designated for iSCSI or iSER I/O. You can configure the internet Storage Name Service (iSNS) to facilitate a scalable configuration and management of iSCSI storage devices.

The same ports can be used for iSCSI and iSER host attachment concurrently, but a single host can establish either an iSCSI or iSER session, but not both.

iSCSI or iSER hosts connect to the system through IP addresses, which are assigned to the Ethernet ports of the node. If the node fails, the address becomes unavailable and the host loses communication with the system through that node. To allow hosts to maintain access to data, the node-port IP addresses for the failed node are transferred to the partner node in the I/O group. The partner node handles requests for both its own node-port IP addresses and also for node-port IP addresses on the failed node. This process is known as *node-port IP failover*. In addition to node-port IP addresses, the iSCSI name and iSCSI alias for the failed node are also transferred to the partner node. After the failed node recovers, the node-port IP address and the iSCSI name and alias are returned to the original node.

Note: The cluster name and node name form parts of the iSCSI name. Changing any of them requires reconfiguration all iSCSI hosts that communicate with the system.

iSER supports only one-way authentication through the Challenge Handshake Authentication Protocol (CHAP). iSCSI supports two types of CHAP authentication: one-way authentication (iSCSI target authenticating iSCSI initiators) and two-way (mutual) authentication (iSCSI target authenticating iSCSI initiators, and vice versa).

For more information about iSCSI host attachment, see *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327.

Make sure that iSCSI initiators, host iSER adapters, and Ethernet switches that are attached to the system are supported by using [SSIC](#).

2.7.6 Native IP replication

Two systems can be linked over native IP links that are connected directly or by Ethernet switches to perform RC functions. RC over native IP provides a less expensive alternative to using FC configurations.

IP replication is supported on both onboard 10G bps Ethernet ports and optional 25 Gbps Ethernet ports. However, when configured over 25 Gbps ports, it does not use RDMA capabilities, and it does not provide a performance improvement compared to 10 Gbps ports.

As a best practice, use a different port for iSCSI host I/O and IP partnership traffic. Also, use a different VLAN ID for iSCSI host I/O and IP partnership traffic.

Specific intersite link requirements must be met when you are planning to use IP partnership for RC. These requirements are described in the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, for the IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and select **Configuring** → **Configuring IP partnerships** → **Intersite link planning**. Also, see Chapter 10, “Advanced Copy Services” on page 553.

2.7.7 Firewall planning

After you have your IP network planned, set up the appropriate firewall rules for each data flow.

For a list of mandatory and optional network flows that are required for operating, see the IBM Documentation information relevant to your IBM FlashSystem platform. For example, for the IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and expand **Planning** → **Planning for hardware** → **Physical installation planning** → **IP address allocation and usage**.

2.8 Planning topology

This section describes the planning topology.

2.8.1 High availability

IBM FlashSystem system support two dual-site topologies: Standard topology, which includes synchronous or asynchronous replication, and HyperSwap. The key attributes of HyperSwap are listed in Table 2-2.

Table 2-2 *HyperSwap attributes*

Item	HyperSwap
Minimum number of I/O groups that are required.	2.
Independent copies of data that are maintained.	2 (Four if volume mirroring to two pools in each site is configured.)
Cache that is retained if only one site is online.	Yes.
Stale consistent data is retained during resynchronization for DR.	Yes.
Ability to use MM, GM, or GM together with an HA solution.	Yes.
Maximum HA volume count.	2000.
Licensing.	Requires a Remote Mirroring license.

The HyperSwap topology uses extra system resources to support a full independent cache on each site, enabling full performance even if one site is lost.

For more information, see the following publications:

- ▶ *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597
- ▶ *IBM Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317

2.8.2 Three-site replication

IBM FlashSystem systems support a three-site replication topology, which includes three-site replication with HyperSwap or three-site replication with MM configurations.

With the three-site replication topology, data is replicated from the primary site or production site to two alternative sites. This feature ensures that if a disaster situation occurs at any one of the sites, the remaining two sites can establish a consistent replication operation with minimal data transfer. The RC relationships are synchronous or asynchronous, depending on which site failed.

The three-site replication topology places three I/O groups at three different sites. It can ensure the availability of a minimum of two copies of data always.

Note: Make sure that the planned configuration is reviewed by IBM or an IBM Business Partner before implementation. Such a review can increase both the quality of the final solution and prevent configuration errors that might impact solution delivery.

For more information, see *Spectrum Virtualize 3-Site Replication*, SG24-8474.

2.9 Back-end storage configuration

External back-end storage systems (also known as *controllers*) provide their logical volumes (LUs), which are detected by a system as managed disks (MDisks) and can be used in storage pools to provision their capacity to system's hosts.

Note: IBM FlashSystem 5010 and IBM FlashSystem 5030 support external virtualization for migration purposes only.

The back-end storage subsystem configuration must be planned for all external storage systems that are attached. Apply the following general guidelines:

- ▶ Most of the supported FC-attached storage controllers must be connected through an FC SAN switch. However, a limited number of systems (including IBM FlashSystem 900 and members of the Storwize, IBM FlashSystem 5000, IBM FlashSystem 7000, and IBM FlashSystem 9000 family) can be direct-attached by using FC.
- ▶ Connect all back-end storage ports to the SAN switch up to a maximum of 16 and zone them to all of the system to maximize bandwidth. The system is designed to handle many paths to the back-end storage.

- ▶ The cluster can be connected to a maximum of 1024 worldwide node names (WWNNs). The general practice is that:
 - EMC DMX/SYMM, all HDS, and SUN/HP HDS clones use one worldwide node name (WWNN) per port. Each port appears as a separate controller to the system.
 - IBM, EMC CLARiiON, and HP use one WWNN per subsystem. Each port appears as a part of a subsystem with multiple ports, with up to a maximum of 16 ports (WWPNs) per WWNN.

However, if you plan for a configuration that might be limited by the WWNN maximum, verify the WWNN versus WWPN policy with the back-end storage vendor.

- ▶ When defining a controller configuration, avoid hybrid configurations and automated tiering solutions. Create LUs for provisioning to the system from a homogeneous disk arrays or solid-state drive (SSD) arrays.
- ▶ Do not provision all available drives on the back-end storage capacity as a single LU. A best practice is to create one LU for eight hard disk drives (HDDs) or SSDs for the back-end system.
- ▶ If your back-end storage system is not supported by the round-robin path policy, ensure that the number of MDisks per storage pool is a multiple of the number of storage ports that are available. This approach ensures sufficient bandwidth for the storage controller, and an even balance across storage controller ports.
- ▶ An IBM FlashSystem system must have exclusive access to every LU that is provisioned to it from a back-end controller. Any specific LU cannot be presented to more than one system. Presenting the same back-end LU to a system and a host is not allowed.
- ▶ Data reduction (compression and deduplication) on the back-end controller is supported only with a limited set of IBM Storage systems.

In general, configure back-end controllers as though they are used as stand-alone systems. However, there might be specific requirements or limitations as to the features that are usable in the specific back-end storage system. For more information about the requirements that are specific to your back-end controller, see the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, for the IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and expand **Configuring** → **Configuring and servicing storage systems**.

The system's large cache and advanced cache management algorithms also allow it to improve the performance of many types of underlying disk technologies. Because hits to the cache can occur in the upper (the system itself) and the lower (back-end controller) level of the overall solution, the solution as a whole can use the larger amount of cache wherever it is. Therefore, the system's cache also provides more performance benefits for back-end storage systems with extensive cache banks.

However, the system cannot increase the throughput potential of the underlying disks in all cases. The performance benefits depend on the underlying back-end storage technology and the workload characteristics, including the degree to which the workload exhibits hotspots or sensitivity to cache size or cache algorithms.

2.10 Internal storage configuration

For general-purpose storage pools with various I/O applications, follow the storage configuration wizard recommendations in the GUI. For specific applications with known I/O patterns, use the CLI to create arrays that suit your needs.

Distributed redundant array of independent disks (DRAID) configurations create large-scale internal MDisks. There are different types of DRAIDs. DRAID 5 can contain as few as four drives initially, and DRAID 6 can contain as few as six drives initially. They both can be expanded up to and contain a maximum of 128 drives. However, DRAID 1 can contain only 2 - 6 drives initially, and can be expanded up to 16 drives of the same capacity.

An array-level recommendation for all types of internal storage except storage-class memory (SCM) is DRAID 6, which outperforms other available RAID levels in most applications while providing fault tolerance and high rebuild speeds.

In specific IBM FlashSystem configurations, for example, small SCM or flash arrays, the newly introduced DRAID 1 feature is suggested to allow for high I/O performance due to all member drive participation in the I/O and the optimized I/O path for multi-core CPUs. It also adds fast rebuilt times on smaller arrays due to the distributed rebuild area.

With IBM Spectrum Virtualize V8.4, up to 12 SCM drives are supported in IBM FlashSystem enclosures. DRAID 1 is recommended for best performance.

DRAID 1 is the only DRAID that can be configured without a rebuild area, supports arrays with a minimum of two member drives, and is limited to 16 member drives (after expansion). Initially, start with six or less member drives, so based on the anticipated capacity (current and future), consider whether to start with a DRAID 1 array or plan for a DRAID 6 array (which can expand even more).

DRAID 1 is recommended as the default in the following scenarios:

- ▶ Two member drives array with no rebuild area
- ▶ Three to six member drives with one rebuild area

Important:

- ▶ DRAID 1 is not recommended with two member drives (and no rebuild area) for HDDs of any size.
- ▶ DRAID 1 is not recommended with two member drives (and no rebuild area) for SSDs (either SAS, FCM, or NVMe) larger than 20 TB of physical capacity.
- ▶ DRAID 1 is not recommended with two member drives (and no rebuild area) for SCMs larger than 8 TB of physical capacity.
- ▶ DRAID 1 is not recommended with three to six member drives (and one rebuild area) for HDDs larger than 8 TB of physical capacity.
- ▶ DRAID 1 supports only a single rebuild area per 3 - 16 member drives.

Due to their mirrored nature, DRAID 1 arrays can use only half of the array's capacity for data. DRAID 6 can achieve better capacity utilization ratios.

At the time of writing, DRAID 1 is supported only on the existing IBM FlashSystem 9200 (AG8/UG8) and IBM FlashSystem 7200 (824/U7C) platforms. Traditional RAID (TRAID) 1 is still supported on IBM FlashSystem 5100, IBM FlashSystem 5030, and IBM FlashSystem 5010.

Note: With IBM Spectrum Virtualize V8.4, you no longer can create arrays with a 128 KB strip size. DRAID 1 arrays with 128 KB strip size are not supported. When you determine the array configuration for your system, plan to create arrays of 256 KB strip size only.

Figure 2-3 provides some planning guidance for the recommended DRAID configuration based on the number of array member drives.

Number of drives	RAID Config	Usable Capacity	I/O Amplification for 70:30 (lower number is better)
★ 2	DRAID-1 2+No Spare	50%	1.3 - Poor redundancy
★ 3	DRAID-1 2+S	33%	1.3 - Performance optimized
★ 4	DRAID-1 3+S	37%	1.3 - Performance optimized
▲ 4	DRAID-5 2+P+S	50%	1.9 - Capacity optimized
★ 5	DRAID-1 4+S	40%	1.3 - Performance optimized
▲ 5	DRAID-5 3+P+S	60%	1.9 - Capacity optimized
★ 6	DRAID-1 5+S	41%	1.3 - Performance optimized
▲ 6	DRAID-6 3+P+Q+S	50%	2.5 - Best redundancy
7	DRAID-1 6+S	42%	1.3 - Performance optimized
★ 7	DRAID-6 4+P+Q+S	57%	2.5 - Best redundancy
8	DRAID-1 7+S	43%	1.3 - Performance optimized
★ 8	DRAID-6 5+P+Q+S	62%	2.5 - Best redundancy

★ #1 recommended configuration
 ▲ #2 recommended configuration

Figure 2-3 Distributed RAID planning guidance

For more information about internal storage configuration, see *IBM FlashSystem 9200 and 9100 Best Practices and Performance Guidelines*, SG24-8448 and *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521.

Summary of supported array types and RAID levels

IBM FlashSystem systems support FCM NVMe drives, industry standard NVMe drives, SCM drives with NVMe architecture, and SAS drives that are within expansion enclosures. The type and level of arrays vary depending on the type of drives in the I/O group.

Table 2-3 summarizes the supported levels. For storage arrays with fewer than seven drives, DRAID 1 is recommended because it offers enhanced resiliency over DRAID 6 arrays. DRAID 6 is recommended for storage arrays with seven or more drives because it can handle two concurrent drive failures.

Table 2-3 Summary of supported drives, array types, and RAID levels

Drive type	Non-DRAIDs	DRAIDs		
	RAID 1	DRAID 1	DRAID 5	DRAID 6
Industry standard NVMe drives or SAS drives (expansion enclosure)	x	x	x	x
FCM NVMe drives		x	x	x
SCM		x	x	x

2.11 Storage pool configuration

The storage pool is at the center of the many-to-many relationship between the internal drive arrays or externally virtualized logical unit numbers (LUNs), which are represented as *MDisks*, and the volumes. It acts as a container of physical disk capacity from which chunks of MDisk space, which is known as *extents*, are allocated to form volumes that are presented to hosts.

The system supports two types of pools: *standard pools* and *Data Reduction Pools (DRP)*. The type is configured when a pool is created and it cannot be changed later. The type of the pool determines the set of features that is available on the system:

- ▶ A feature that can be implemented only with standard pools is VMware vSphere integration with VMware vSphere Virtual Volumes (VVOLs).
- ▶ Features that can be implemented only with DRPs are:
 - Automatic capacity reclamation with SCSI UNMAP (This feature returns capacity that is marked as no longer used by a host back to storage pool.)
 - DRP compression (in-flight data compression)
 - DRP deduplication
 - FlashCopy with redirect-on-write (RoW)

Note: FlashCopy with RoW is usable only for volumes with supported deduplication without mirroring relationships and within the same pool and I/O group. Automatic mode selection (RoW/copy-on-write (CoW)) is based on these conditions.

In addition to providing data reduction options, DRP amplifies the I/O and CPU workload, which should account for during performance sizing and planning.

Also, self-compressing drives (FCM drives) still perform compression independently of the pool type.

IBM Spectrum Virtualize V8.4 introduces the *Comprestimation Always On* feature, where the continuous compression of all virtual disks (VDisks) is provided so that compressibility estimations are always available. This feature is *on* by default.

Another base storage pool parameter is the extent size. There are two implications of a storage pool extent size:

- ▶ The maximum volume, MDisks, and managed storage capacity depend on the extent size. The bigger the extent that is defined for the specific pool, the larger is the maximum size of this pool, the maximum MDisk size in the pool, and the maximum size of a volume that is created in the pool.
- ▶ The volume sizes must be a multiple of the extent size of the pool in which the volume is defined. Therefore, the smaller the extent size, the better control that you have over the volume size.

The system supports extent sizes of 16 - 8192 mebibytes (MiB). The extent size is a property of the storage pool, and it is set when the storage pool is created.

Note: The base pool parameters, pool type, and extent size are set during pool creation and cannot be changed later. If you need to change the extent size or pool type, all volumes must be migrated from a storage pool and then the pool itself must be deleted and re-created.

When planning pools, the encryption is defined on a pool level and the encryption setting cannot be changed after a pool is created. If you create an unencrypted pool, there is no way to encrypt it later. Your only option is to delete it and re-create it as encrypted.

When planning storage pool layout, consider the following aspects:

- ▶ Pool reliability, availability, and serviceability (RAS):
 - The storage pool is a failure domain. If one array or external MDisk is unavailable, the pool and all volumes in it goes offline.
 - The number and size of storage pools affects system availability. Using a larger number of smaller pools reduces the failure domain if one of the pools goes offline. However, increasing the number of storage pools affects the storage use efficiency, and the number is subject to the configuration maximum limit.
 - You cannot migrate volumes between storage pools with different types or extent sizes. However, you can use volume mirroring to create copies between storage pools.
- ▶ Pool performance:
 - Do not mix same-tier arrays or MDisks with different performance characteristics in one pool. For example, do not use DRAID 6 arrays of six tier 1 SSDs and DRAID 6 arrays of 24 tier 1 SSDs in the same pool. This technique is the only way to ensure consistent performance characteristics of volumes that are created from the pool.

Arrays with different tiers in one pool may be used because their performance differences become beneficial when you use the Easy Tier function.
 - Create multiple storage pools if you must isolate specific workloads to separate storage.
 - Ensure that performance sizing was done for selected pool type and feature set.

2.11.1 Child pools

Instead of being created directly from MDisks, child pools are created from existing usable capacity that is assigned to a parent pool. As with parent pools, volumes can be created that specifically use the usable capacity that is assigned to the child pool. Child pools are similar to parent pools with similar properties and can be used for volume copy operation.

When a standard child pool is created, the usable capacity for a child pool is reserved from the usable capacity of the parent pool. The usable capacity for the child pool must be smaller than the usable capacity in the parent pool. After the child pool is created, the amount of usable capacity that is specified for the child pool is no longer reported as usable capacity of its parent pool.

When a data reduction child pool is created, the usable capacity for the child pool is the entire usable capacity of the data reduction parent pool without limit. After a data reduction child pool is created, the usable capacity of the child pool and the usable capacity of the parent pool are reported as the same.

A number of administration tasks benefit from being able to define and work with a part of a pool. For example, the system supports VVOLs, which are used in VMware vCenter and vSphere APIs for Storage Awareness (VASA) applications. Before a child pool can be used for virtual volumes for these applications, the system must be enabled for virtual volumes.

Consider the following general guidelines when you create or work with a child pool:

- ▶ The management GUI displays only the capacity details for child and migration pools.
- ▶ Child pools can be created and changed with the CLI or GUI.
- ▶ When using child pools with standard pools, you can specify a warning threshold that alerts you when the used capacity of the child pool is reaching its upper limit. Use this threshold to ensure that access is not lost when the used capacity of the child pool is close to its usable capacity.
- ▶ On systems with encryption enabled, standard child pools can be created to migrate existing volumes in a non-encrypted pool to encrypted child pools. When you create a standard child pool after encryption is enabled, an encryption key is created for the child pool even when the parent pool is not encrypted. You can then use volume mirroring to migrate the volumes from the non-encrypted parent pool to the encrypted child pool. Encrypted data reduction child pools can be created only if the parent pool is encrypted. The data reduction child pool inherits an encryption key from the parent pool.
- ▶ Ensure that any child pools that are associated with a parent pool have enough usable capacity for the volumes that are in the child pool before removing MDisks from a parent pool. The system automatically migrates all extents that are used by volumes to other MDisks in the parent pool to ensure that data is not lost.
- ▶ You cannot shrink the usable capacity of a child pool below its used capacity. The system also resets the warning level when the child pool is shrunk and issues a warning if the level is reached when the usable capacity is shrunk.
- ▶ The system supports migrating a copy of volumes between child pools within the same parent pool or migrating a copy of a volume between a child pool and its parent pool. Migrations between a source and target child pool with different parent pools are not supported. However, you can migrate a copy of the volume from the source child pool to its parent pool. The volume copy can then be migrated from the parent pool to the parent pool of the target child pool. Finally, the volume copy can be migrated from the target parent pool to the target child pool.

Child pools can be assigned to an ownership group. An *ownership group* defines a subset of users and objects within the system. You can create ownership groups to further restrict access to specific resources that are defined in the ownership group. Only users with Security Administrator roles can configure and manage ownership groups.

Ownership can be defined explicitly or it can be inherited from the user, user group, or from other parent resources, depending on the type of resource. Ownership of child pools must be assigned explicitly, and they do not inherit ownership from other parent resources. New or existing volumes that are defined in the child pool inherit the ownership group that is assigned for the child pool.

For more information about ownership groups, see Chapter 11, “Ownership groups” on page 723.

2.11.2 The storage pool and cache relationship

The system uses cache partitioning to limit the potential negative effects that a poorly performing storage controller can have on the clustered system. The cache partition allocation size is based on the number of configured storage pools. This design protects against an individual overloaded back-end storage system from filling the system write cache and degrading the performance of the other storage pools. Table 2-4 lists the limits of the write-cache data that can be used by a single storage pool.

Table 2-4 Limits of the cache data

Number of storage pools	Upper limit
1	100%
2	66%
3	40%
4	30%
5 or more	25%

No single partition can occupy more than its upper limit of write cache capacity. When the maximum cache size is allocated to the pool, the system starts to limit incoming write I/Os for volumes that are created from the storage pool. The host writes are limited to the destage rate on a one-out-one-in basis.

Only writes that target the affected storage pool are limited. The read I/O requests for the throttled pool continue to be serviced normally. However, because the system is offloading cache data at the maximum rate that the back-end storage can sustain, read response times are expected to be affected.

All I/O that is destined for other (non-throttled) storage pools continues as normal.

2.12 Volume configuration

When planning a volume, consider the required performance, availability, and capacity. Every volume is assigned to an I/O group that defines which pair of system nodes services I/O requests to the volume.

Note: No fixed relationship exists between I/O groups and storage pools.

When a host sends I/O to a volume, it can access the volume with either of the nodes in the I/O group but each volume has a *preferred node*. Many of the multipathing driver implementations that the system supports use this information to direct I/O to the preferred node. The other node in the I/O group is used only if the preferred node is not accessible.

During volume creation, the system selects the node in the I/O group that has the fewest volumes to be the preferred node. After the preferred node is chosen, it can be changed manually, if required.

Strive to distribute volumes evenly across available I/O groups and nodes within the system.

For more information about volume types, see Chapter 6, “Volumes” on page 299.

2.12.1 Planning for image mode volumes

Use image mode volumes to present to hosts data that is written to the back-end storage before it was virtualized. An image mode volume directly corresponds to the MDisk from which it is created.

Image mode volumes are a useful tool in storage migration and during system implementation to a working environment.

2.12.2 Planning for fully allocated volumes

A fully allocated volume presents to mapped hosts the same capacity that the volume uses in the storage pool. No data reduction is performed on a pool level. However, if a fully allocated volume is provisioned from a pool with data reducing storage, such as self-compressing drives (FCM drives), the data is still compressed on a drive level.

Fully allocated volumes provide the best performance because they do not cause I/O amplification, and they require less CPU time compared to other volume types.

2.12.3 Planning for thin-provisioned volumes

A thin-provisioned volume presents a different capacity to mapped hosts than the capacity that the volume uses in the storage pool. Space is not allocated on a thin-provisioned volume if an incoming host write operation contains all zeros.

Using the thin-provisioned volume feature that is called *zero detect*, you can reclaim unused allocated disk space (zeros) when you convert a fully allocated volume to a thin-provisioned volume by using volume mirroring.

DRPs enhance capacity efficiency for thin-provisioned volumes by monitoring the host's capacity usage. When the host indicates that the capacity is no longer needed, the capacity is released and can be reclaimed by the DRP to be redistributed automatically. Standard pools cannot reclaim capacity.

Note: Avoid using thin-provisioned volumes on a data-reducing back end like self-compressing drives when implementing DRP.

2.12.4 Planning for compressed volumes

With compressed volumes, data is compressed as it is written to disk, which saves more space. When data is read to hosts, the data is decompressed.

Compression is available through data reduction support as part of the system. If you want volumes to use compression as part of data reduction support, compressed volumes must belong to DRPs.

If you use compressed volumes over a pool with self-compressing drives, the drive still attempts compression because it cannot be disabled on the drive level. However, there is no performance impact due to the algorithms that FCM uses to manage compression.

Before implementing compressed volumes, perform data analysis to discover your average compression ratio and ensure that performance sizing was done for compression.

IBM Spectrum Virtualize V8.4 introduces the Comprestimation Always On feature, which ensures the continuous comprestimation of all VDisks is provided so that compressibility estimations are always available. This feature is *on* by default.

Special considerations must be taken when implementing compression on IBM FlashSystem 5030, which does not have compression accelerator hardware and uses the canister's CPU for compression and decompression. Therefore, strict performance planning and sizing is required.

Note: If you use compressed volumes over FCM drives, the compression ratio on a drive level must be assumed to be 1:1 to avoid array overprovisioning and running out of space.

2.12.5 Planning for deduplicated volumes

Deduplication can be configured for volumes that use different capacity saving methods, such as thin provisioning. Deduplicated volumes must be created in DRPs for added capacity savings. Deduplication is a type of data reduction that eliminates duplicate copies of data. Deduplication of user data occurs within a DRP and only between volumes or volume copies that are marked as deduplicated.

With deduplication, the system identifies unique chunks of data that is called *signatures* to determine whether new data is written to the storage. Deduplication is a hash-based solution, which means chunks of data are compared to their signatures rather than to the data itself. If the signature of the new data matches an existing signature that is stored on the system, then the new data is replaced with a reference. The reference points to the stored data instead of writing the data to storage. This process saves the capacity of the back-end storage by not writing new data to storage, and it might improve the performance of read operations to data that has an existing signature.

The same data pattern can occur many times, and deduplication decreases the amount of data that must be stored on the system. A part of every hash-based deduplication solution is a repository that supports looking up matches for incoming data. The system contains a database that maps the signature of the data to the volume and its virtual address. If an incoming write operation does not have a signature that is stored in the database, then a duplicate is not detected and the incoming data is stored on back-end storage.

To maximize the space that is available for the database, the system distributes this repository between all nodes in the I/O groups that contain deduplicated volumes. Each node carries a distinct portion of the records that are stored in the database. If nodes are removed or added to the system, the database is redistributed between the nodes to ensure full use of the available memory.

Before implementing deduplication, perform data analysis to estimate the deduplication savings and make sure that system performance sizing was done for deduplication.

2.13 Host attachment planning

The system supports the attachment of a various host hardware types running different OSs with FC SAN or IP SAN. For a list of instructions that is specific to your host setup, see the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, for the IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and expand **Configuring** → **Host attachment**.

Note: With IBM Spectrum Virtualize V8.4, FC-NVMe host attachment with HyperSwap configurations is supported.

2.13.1 Queue depth

Typically, hosts issue subsequent I/O requests to storage systems without waiting for the completion of previous ones. The number of outstanding requests is called *queue depth*. Sending multiple I/O requests in parallel (asynchronous I/O) provides significant performance benefits compared to sending them one-by-one (synchronous I/O). However, if the number of queued requests exceeds the maximum that is supported by the storage controller, you experience performance degradation.

For more information about how to calculate the correct host queue depth for your environment, see the IBM Documentation information that is relevant to your IBM FlashSystem platform. For example, for the IBM FlashSystem 9200 related information, go to [IBM FlashSystem 9200 documentation](#) and expand **Configuring** → **Host attachment**.

2.13.2 Microsoft Offloaded Data Transfer

If your Windows hosts are configured to use Microsoft Offloaded Data Transfer (ODX) to offload the copy workload to the storage controller, consider the benefits of this technology against the extra load on the storage controllers. The benefits and effects of enabling ODX are especially prominent in Microsoft Hyper-V environments with ODX enabled.

2.13.3 SAN boot support

The system supports SAN boot or startup for selected configurations of hosts running AIX, Microsoft Windows, and other OSs. To check whether your configuration is supported for SAN boot, see the [SSIC](#).

2.13.4 Planning for large deployments

Each I/O group can have up to 512 host objects defined. This limit is the same whether hosts are attached by using FC, iSCSI, or a combination of both. To allow more than 512 hosts to access the storage, you must divide them into groups of 512 hosts or less and map each group to a single I/O group only. With this approach, you can configure up to 2048 host objects on a system with four I/O groups (eight nodes).

For best performance, split each host group into two sets. For each set, configure the preferred access node for volumes that are presented to the host set to one of the I/O group nodes. This approach helps to evenly distribute load between the I/O group nodes.

Note: A volume can be mapped only to a host that is associated with the I/O group to which the volume belongs.

2.13.5 Planning for SCSI UNMAP

UNMAP is a set of SCSI primitives that hosts use to indicate to a SCSI target that space that is allocated to a range of blocks on a target storage volume is no longer required. With this command, the storage controller takes measures and optimizes the system so that the space can be reused for other purposes.

The IBM FlashSystem supports end-to-end UNMAP compatibility, which means that a command that is issued by a host is processed and sent to the back-end storage device or drive.

UNMAP processing can be controlled with two separate settings:

- ▶ First setting advertises UNMAP support to hosts.
- ▶ Second setting controls whether IBM FlashSystem sends **UNMAP** commands to back-end storage (drives and external controllers).

Host UNMAP support is enabled by default on FlashSystem 9100 and 9200 and disabled by default on all other IBM FlashSystem family systems.

Thorough planning is required if you want to switch host UNMAP support on. Enabling it will allow you to fully benefit from capacity reclamation features in Data Reduction Pools, but host UNMAP requests might overload IBM FlashSystem back-end if it has spinning disks, especially NL-SAS drives, causing serious performance problems.

Back-end UNMAP is enabled by default on all IBM FlashSystem platforms, and it is a best practice to keep it turned on for most use cases.

2.14 Planning copy services

IBM FlashSystem systems offer a set of copy services, such as IBM FlashCopy (snapshots) and RC, in synchronous and asynchronous modes. For more information about copy services, see Chapter 10, “Advanced Copy Services” on page 553.

2.14.1 FlashCopy guidelines

With the FlashCopy function, you can perform a point-in-time (PiT) copy of one or more volumes. The FlashCopy function creates a PiT or time-zero (T0) copy of data that is stored on a source volume to a target volume by using a CoW and copy-on-demand mechanism.

While the FlashCopy operation is performed, the source volume is stopped briefly to initialize the FlashCopy bitmap, and then I/O can resume. Although several FlashCopy options require the data to be copied from the source to the target in the background, which can take time to complete, the resulting data on the target volume is presented so that the copy appears to complete immediately.

The FlashCopy function operates at the block level below the host OS and cache, so those levels must be flushed by the OS for a FlashCopy copy to be consistent.

When you use the FlashCopy function, observe the following guidelines:

- ▶ Both the FlashCopy source and target volumes should use the same preferred node.
- ▶ If possible, keep the FlashCopy source and target volumes on separate storage pools.

With IBM Spectrum Virtualize V8.4, a FlashCopy with RoW mechanism is available with DRPs. FlashCopy with RoW uses the DRP internal deduplication referencing capabilities to reduce overheads by creating references instead of copying the data. It provides for better performance and reduces back-end I/O amplification for FlashCopies and snapshots.

Note: FlashCopy with RoW is usable only for volumes with supported deduplication without mirroring relationships and within the same pool and I/O group. Automatic mode selection (RoW/CoW) is based on these conditions.

For more information about planning for the FlashCopy function, see *IBM FlashSystem 9200 and 9100 Best Practices and Performance Guidelines*, SG24-8448 and *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521.

2.14.2 Planning for Metro Mirror and Global Mirror

MM is a copy service that provides a continuous, synchronous mirror of one volume to a second volume. The systems can be up to 300 km (186.4 miles) apart. Because the mirror is updated synchronously, no data is lost if the primary system becomes unavailable. MM is typically used for DR purposes, where it is important to avoid any data loss.

GM is a copy service that is similar to MM, but copies data asynchronously. You do not have to wait for the write to the secondary system to complete. For long distances, performance is improved compared to MM. However, if a failure occurs, you might lose data.

GM uses one of two methods to replicate data. Multicycling GM is designed to replicate data while adjusting for bandwidth constraints. It is appropriate for environments where it is acceptable to lose a few minutes of data if a failure occurs.

For environments with higher bandwidth, non-cycling GM can be used so that less than a second of data is lost if a failure occurs. GM also works well when sites are more than 300 kilometers (186.4 miles) apart.

When copy services are used, all components in the SAN must sustain the workload that is generated by application hosts and the data replication workload. Otherwise, the system can automatically stop copy services relationships to protect your application hosts from increased response times.

While planning RC services, consider the following aspects:

► Copy services topology

One or more clusters can participate in a copy services relationship. One typical and simple use case is DR, where one site is active and another performs only a DR function. In such a case, the solution topology is simple, with one cluster per site and uniform replication direction for all volumes. However, multiple other topologies are possible that you can use to design a solution that optimally fits your set of requirements.

► GM versus MM

Decide which type of copy services you are going use. This decision should be requirement-driven. With MM, you prevent any data loss during a system failure, but it has more stringent requirements, especially regarding intercluster link bandwidth and latency, and remote site storage performance. Also, MM incurs a performance penalty because writes are not confirmed to host until a data reception confirmation is received from the remote site.

With GM, you can relax constraints on the system requirements at the cost of using asynchronous replication, which enables the remote site to lag behind the local site. The choice of the replication type has major effects on all other aspects of the copy services planning.

Using GM and MM between the same two clustered systems is supported. Also, the RC type may be changed from one to another one.

For native IP replication, use the RC mode of Multicycling GM (or Global Mirror with Change Volumes (GMCV)).

► Intercluster link

The local and remote clusters can be connected by an FC or IP network. Each of the technologies has its own requirements concerning supported distance, link speeds, bandwidth, and vulnerability to frame or packet loss.

When planning the intercluster link, consider the peak performance that is required. This consideration is especially important for MM configurations.

The bandwidth between sites must be sized to meet the peak workload requirements. When planning the inter-site link, consider the initial sync and any future resync workloads. It might be worthwhile to secure more link bandwidth for the initial data synchronization.

If the link between the sites is configured with redundancy so that they can tolerate single failures, you must size the link so that the bandwidth and latency requirements are met even during single failure conditions.

When planning the inter-site link, note whether it is dedicated to the inter-cluster traffic or is going to be used to carry any other data. Sharing the link with other traffic might affect the link's ability to provide the required bandwidth for data replication.

► Volumes and consistency groups

Determine whether volumes can be replicated independently. Some applications use multiple volumes and require that the order of writes to these volumes is preserved in the remote site. Notable examples of such applications are databases.

If an application requires that the write order is preserved for the set of volumes that it uses, create a consistency group for these volumes.

2.15 Data migration

Data migration is an important part of an implementation, so you must prepare a detailed data migration plan. You might need to migrate your data for one of the following reasons:

- ▶ Redistribute a workload within a clustered system across back-end storage subsystems.
- ▶ Move a workload on to newly installed storage.
- ▶ Move a workload off old or failing storage ahead of decommissioning it.
- ▶ Move a workload to rebalance a changed load pattern.
- ▶ Migrate data from an older disk subsystem.
- ▶ Migrate data from one disk subsystem to another one.

Because multiple data migration methods are available, choose the method that best fits your environment, OS platform, type of data, and the application's service-level agreement (SLA).

Data migration methods can be divided into three classes:

- ▶ Based on the host OS, for example, by using the system's logical volume manager (LVM)
- ▶ Based on specialized data migration software
- ▶ Based on the system data migration features

For more information about system data migration tools, see Chapter 8, "Storage migration" on page 485 and Chapter 10, "Advanced Copy Services" on page 553.

With data migration, apply the following guidelines:

- ▶ Choose the data migration method that best fits your OS platform, type of data, and SLA.
- ▶ Choose where you want to place your data after migration in terms of the storage tier, pools, and back-end storage.
- ▶ Check whether enough free space is available in the target storage pool.
- ▶ To minimize downtime during the migration, plan ahead of time all of the required changes, including zoning, host definition, and volume mappings.
- ▶ Prepare a detailed operation plan so that you do not overlook anything at data migration time. Especially for a large or critical data migration, have the plan peer-reviewed and formally accepted by an appropriate technical design authority within your organization.
- ▶ Perform and verify a backup before you start any data migration.
- ▶ You might want to use the system as a data mover to migrate data from a non-virtualized storage subsystem to another non-virtualized storage subsystem. In this case, you might have to add checks that relate to the specific storage subsystem that you want to migrate.

Be careful when you are using slower disk subsystems for the secondary volumes for high-performance primary volumes because the system's cache might not be able to buffer all the writes. Flushing cache writes to slower back-end storage might impact performance of your hosts.

- ▶ Consider storage performance. The migration workload might be much higher than expected during normal operations of the system. If there is already application data on the system to which you are migrating, the application performance might suffer if the system is overloaded. Consider using host or volume level throttles when performing migration on a production environment.

2.16 Performance monitoring with IBM Storage Insights

IBM Storage Insights is integral to monitoring and ensuring the continued availability of the system.

Available at no additional charge, the cloud-based IBM Storage Insights product provides a single dashboard that provides a clear view of all your IBM block storage. You can make better decisions by seeing trends in performance and capacity.

With storage health information, you can focus on areas needing attention and when IBM support is needed, IBM Storage Insights simplifies uploading logs, speeds resolution with online configuration data, and provides an overview of open tickets all in one place.

IBM Storage Insights provides a unified view of IBM systems. By using it, you can see all of your IBM storage inventory as a live event feed so that you know what is going on with your storage.

IBM Storage Insights provides advanced customer service with an event filter that provides the following functions:

- ▶ The ability for you and support to view support tickets and open and close them, and to track trends.
- ▶ With the auto log collection capability, you can collect the logs and send them to IBM before IBM Support starts looking into the problem. This feature can reduce the time to solve the case by as much as 50%.

Figure 2-4 shows the architecture of the IBM Storage Insights application, the supported products, and the three main teams who can benefit from the use of the tool.

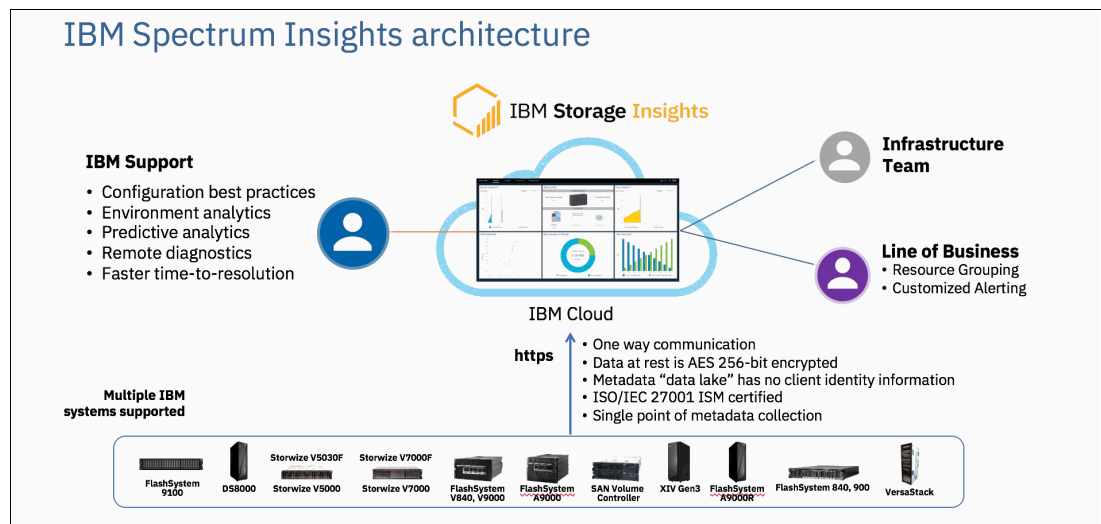


Figure 2-4 IBM Storage Insights architecture

IBM Storage Insights provides a lightweight data collector that is deployed on a Linux, Windows, or AIX server or a guest in a virtual machine (VM) (for example, a VMware guest).

The data collector streams performance, capacity, asset, and configuration metadata to your IBM Cloud instance.

The metadata flows in one direction, that is, from your data center to IBM Cloud over HTTPS. In the IBM Cloud, your metadata is protected by physical, organizational, access, and security controls. IBM Storage Insights is ISO/IEC 27001 Information Security Management certified.

To monitor storage systems, you must provide a username and password to log in to the storage systems. The role or user group that is assigned to the username must have the appropriate privileges to monitor the data that is collected. As of IBM Spectrum Virtualize V8.3.1.2 and SI/ IBM Spectrum Control V5.3.7 or later, data collection can be done with the Monitor (least privileged) role.

Figure 2-5 shows the data flow from systems to the IBM Storage Insights cloud.

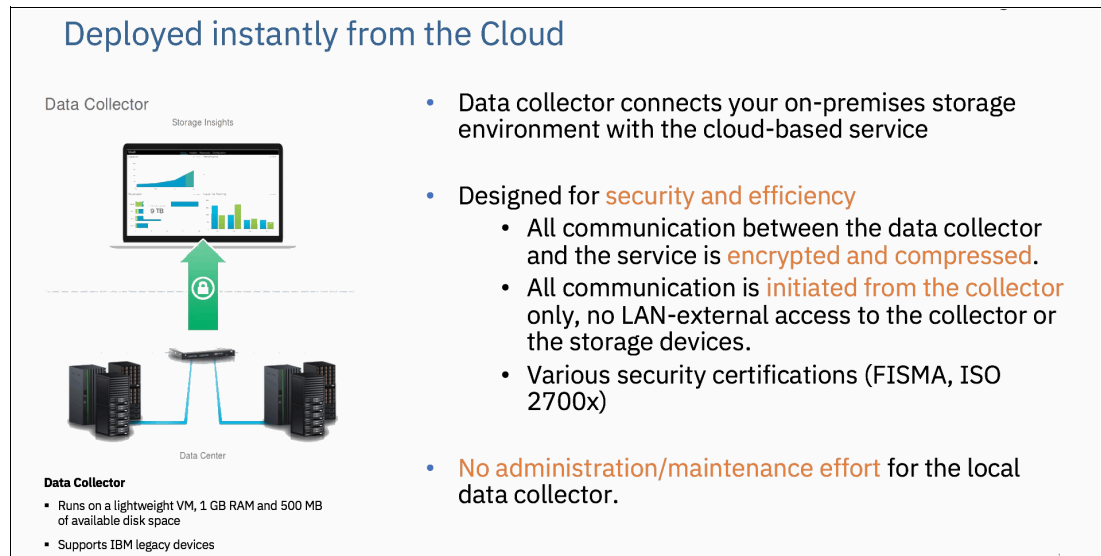


Figure 2-5 Data flow from the storage systems to the IBM Storage Insights cloud

Metadata about the configuration and operations of storage resources is collected, such as:

- ▶ Name, model, firmware, and type of storage system
- ▶ Inventory and configuration metadata for the storage system's resources, such as volumes, pools, disks, and ports
- ▶ Capacity values, such as capacity, unassigned space, used space, and the compression ratio
- ▶ Performance metrics, such as read and write data rates, I/O rates, and response times

The application data that is stored on the storage systems cannot be accessed by the data collector.

Access to the metadata that is collected is restricted to the following users:

- ▶ The customer who owns the dashboard.
- ▶ The administrators who are authorized to access the dashboard, such as the customer's operations team.
- ▶ The IBM Cloud team that is responsible for the day-to-day operation and maintenance of IBM Cloud instances.
- ▶ IBM Support for investigating and closing service tickets.

For more information about IBM Storage Insights and to sign up and register for the free service, see the following resources:

- ▶ [Fact Sheet](#)
- ▶ [Demonstration](#)
- ▶ [Security Guide](#)
- ▶ [Registration](#)

For more information, see 13.12, “IBM Storage Insights monitoring” on page 865.

2.17 Configuration backup procedure

Save the configuration before and after any major configuration changes on the system. Saving the configuration is a crucial part of management, and various methods can be applied to back up your system configuration. A best practice is to implement an automatic configuration backup by using the configuration backup command. Make sure that you save the configuration to a host system that does not depend on the storage that is provisioned from a system whose configuration is backed up.

For more information, see 13.4, “Configuration backup” on page 808.



Initial configuration

This chapter describes the initial configuration of the IBM FlashSystem 9100, IBM FlashSystem 9200, IBM FlashSystem 7200, IBM FlashSystem 5100, IBM FlashSystem 5030 and IBM FlashSystem 5010 systems, and the IBM Storwize V5100 and IBM Storwize V7000 Gen3 systems. It provides step-by-step instructions about how to do the initial setup and defines the base settings of the system, which are done during the implementation phase before volumes are created and provisioned.

This chapter includes the following topics:

- ▶ 3.1, “Prerequisites” on page 108
- ▶ 3.2, “System initialization” on page 109
- ▶ 3.3, “System setup” on page 113
- ▶ 3.4, “Base configuration” on page 123
- ▶ 3.5, “Configuring management access” on page 139

3.1 Prerequisites

Note: IBM FlashSystem 9100 and IBM FlashSystem 9200 are installed by an IBM System Services Representative (IBM SSR). You must provide all the necessary information to the IBM SSR by filling out the planning worksheets, which can be found in [IBM FlashSystem 9200 documentation](#) by selecting **Planning** → **Planning worksheets (customer task)**.

After the IBM SSR completes their task, continue the setup by following the instructions in 3.3, “System setup” on page 113.

Before initializing and setting up the system, ensure that the following prerequisites are met:

- ▶ The physical components fulfill all the requirements and are correctly installed, including:
 - The control enclosures are physically installed in the racks.
 - The Ethernet and Fibre Channel (FC) cables are connected.
 - The expansion enclosures, if available, are physically installed and attached to the control enclosures that will use them.
 - The system control enclosures and optional expansion enclosures are powered on.
- ▶ The web browser that is used for managing the system is supported by the management GUI. For the list of supported browsers, see [Supported Browsers](#).
- ▶ You have the required information, which can be found in IBM Documentation, including:
 - The IPv4 (or IPv6) addresses that are assigned for the system’s management interfaces:
 - The unique cluster IP address, which is the address that is used for the management of the system.
 - Unique service IP addresses, which are used to access node service interfaces. You need one address for each node (two per control enclosure).
 - The IP subnet mask for each subnet that is used.
 - The IP gateway for each subnet that is used.
 - The licenses that might be required to use particular functions (depending on the system type):
 - Remote Copy (RC).
 - External virtualization.
 - IBM FlashCopy.
 - Compression.
 - Encryption.
 - Information that is used by a system when performing Call Home functions, such as:
 - The company name and system installation address.
 - The name, email address, and phone number of the storage administrator whom IBM can contact if necessary.
 - (optional) The Network Time Protocol (NTP) server IP address.

- (optional) The Simple Mail Transfer Protocol (SMTP) server IP address, which is necessary only if you want to enable Call Home or want to be notified about system events through email.
- (optional) The IP addresses for Remote Support Proxy Servers, which are required only if you want to use them with the Remote Support Assistance feature.

3.2 System initialization

This section provides step-by-step instructions about how to create the system cluster.

To start the initialization procedure, connect a desktop PC or a Notebook to the technician port. The *technician port* is a dedicated 1 Gb Ethernet (GbE) port at the rear of each of the nodes in the control enclosure. On all platforms except IBM FlashSystem 5010, it can be used only to initialize or service the system. It cannot be connected to an Ethernet switch because it supports only a direct connection, and it remains disconnected after the initial setup is done.

On IBM FlashSystem 5010, the technician port is enabled initially, but after the setup wizard is complete, the port is switched to internet Small Computer Systems Interface (iSCSI) host attachment mode. However, to re-enable the onboard Ethernet port 2 on a system to be used as the technician port, run the command that is shown in Example 3-1.

Example 3-1 Reenabling the onboard Ethernet port 2 as the technician port

```
IBM_IBM FlashSystem 5010:superuser>satask chserviceip -techport enable -force
```

The location of the technician port is shown in Figure 3-1.

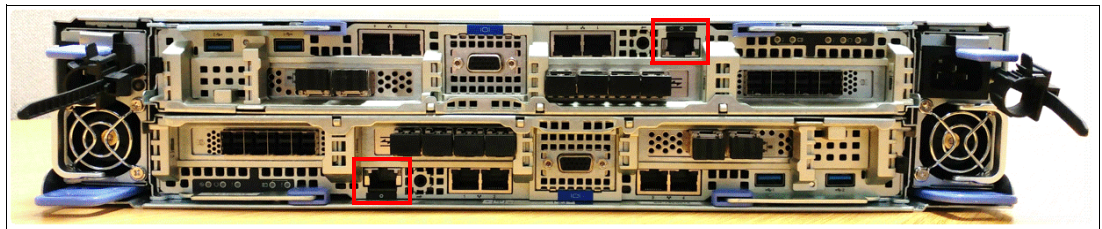


Figure 3-1 Location of the technician port

The location of the technician port on Storwize V7000 Gen2 is shown in Figure 3-2.

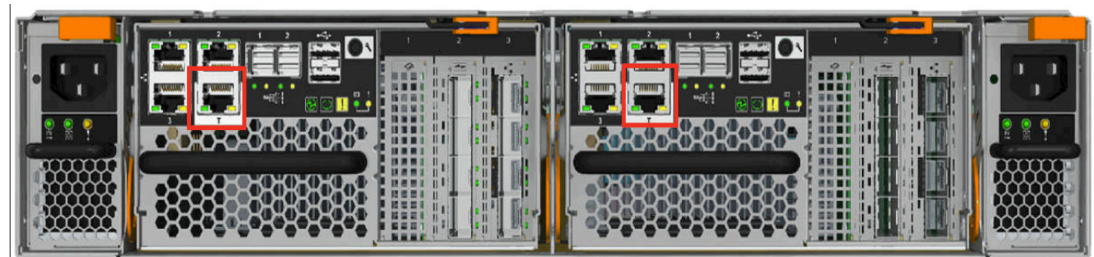


Figure 3-2 Location of the technician port on Storwize V7000 Gen2

The location of the technician port on IBM FlashSystem 5030 is shown in Figure 3-3.



Figure 3-3 Location of the technician port on IBM FlashSystem 5030

The location of the technician port on IBM FlashSystem 5010 is shown in Figure 3-4.



Figure 3-4 Location of the technician port on IBM FlashSystem 5010

The technician port runs an IPv4 DHCP server, and it can assign an address to any device that is connected to this port. Ensure that your PC or Notebook Ethernet adapter is configured to use a DHCP client if you want the IP to be assigned automatically. If you prefer not to use DHCP, you can set a static IP on the Ethernet port from the 192.168.0.x/24 subnet, for example, 192.168.0.2 with the netmask 255.255.255.0.

The default IP address of a technician port on a node canister is 192.168.0.1. Do not use this IP address for your PC or Notebook.

Note: Ensure that the technician port is not connected to the organization's network. No Ethernet switches or hubs are supported on this port.

3.2.1 System initialization process

Before a system is initialized, each node canister of a new system remains in the *candidate* state and cannot process I/O. During initialization, the nodes in one enclosure are joined in a *cluster*, which is later configured to process data. If your systems have more than one control enclosure, all the other ones except the first one must not be initialized. The remaining control enclosures are added to the cluster by using a cluster management interface (GUI or command-line interface (CLI)) after the first one is set up.

You must specify IPv4 or an IPv6 system management addresses, which are assigned to Ethernet port 1 on each node and used to access the management GUI and CLI. After the system is initialized, you can specify other IP addresses.

Note: Do not perform the system initialization procedure on more than one node canister of one control enclosure. After initialization completes, use the management GUI or CLI to add control enclosures to the system.

To do the initialization of a new system, complete the following steps:

1. Connect your PC or Notebook to a technician port of any canister of the control enclosure. Ensure that you obtained a valid IPv4 address with DHCP.
2. Open a supported web browser and go to `http://install`. The browser is automatically redirected to the System Initialization wizard. You can also use the IP address `http://192.168.0.1` if you are not automatically redirected.

Note: During the system initialization, you are prompted to accept untrusted certificates because the system certificates are self-signed. If you are directly connected to the service interface, there is no doubt about the identity of the certificate issuer, so you can safely accept the certificates.

If the system is not in a state that allows initialization, you are redirected to the Service Assistant interface. Use the displayed error codes to troubleshoot the problem.

3. The Welcome dialog box opens, as shown in Figure 3-5. Click **Next** to start the procedure.

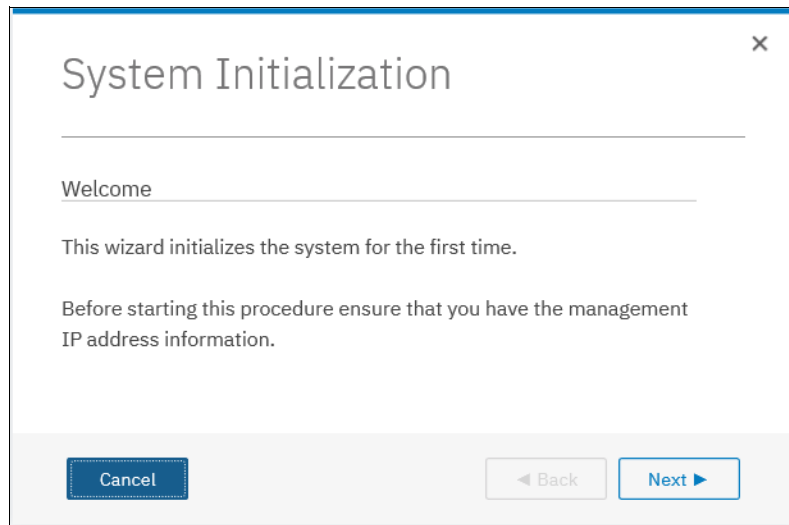


Figure 3-5 System Initialization: Welcome dialog box

4. A window opens in which two options are presented, as shown in Figure 3-6. Select the first option and click **Next**.

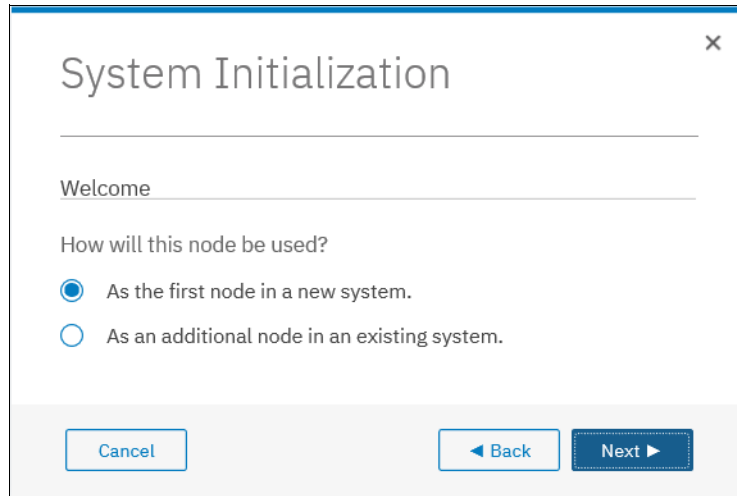


Figure 3-6 System Initialization: Create a system or expand the existing one

If you select **As an additional node in an existing system**, you are prompted to disconnect from the technician port and use the GUI of an existing system to add new nodes.

5. Enter the management IP address information for the new system, as shown in Figure 3-7. Set the IP address, network mask, and gateway.

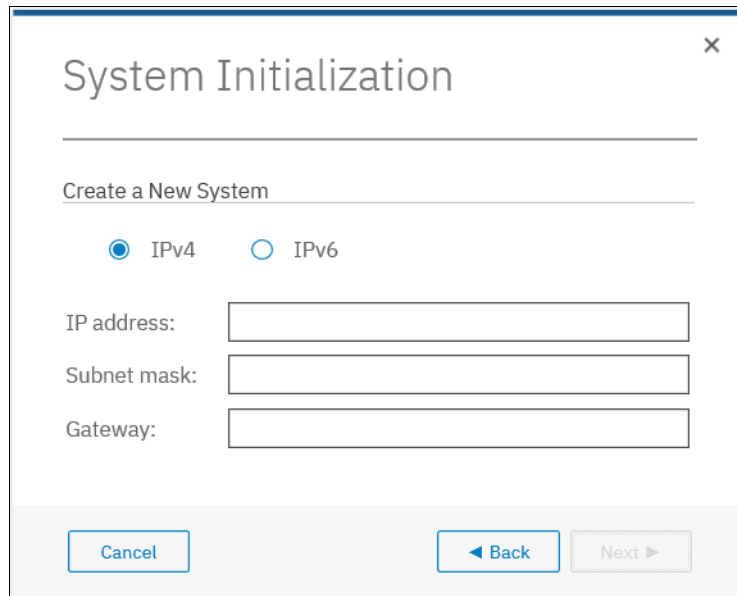


Figure 3-7 System Initialization: Management IP

6. Click **Next**.

7. A window with restart timer opens. When the timeout is reached, you can click **Next** to see the final initialization window, as shown in Figure 3-8. Follow the instructions, and browser is redirected to the management IP address to access the system GUI after you click **Finish**.

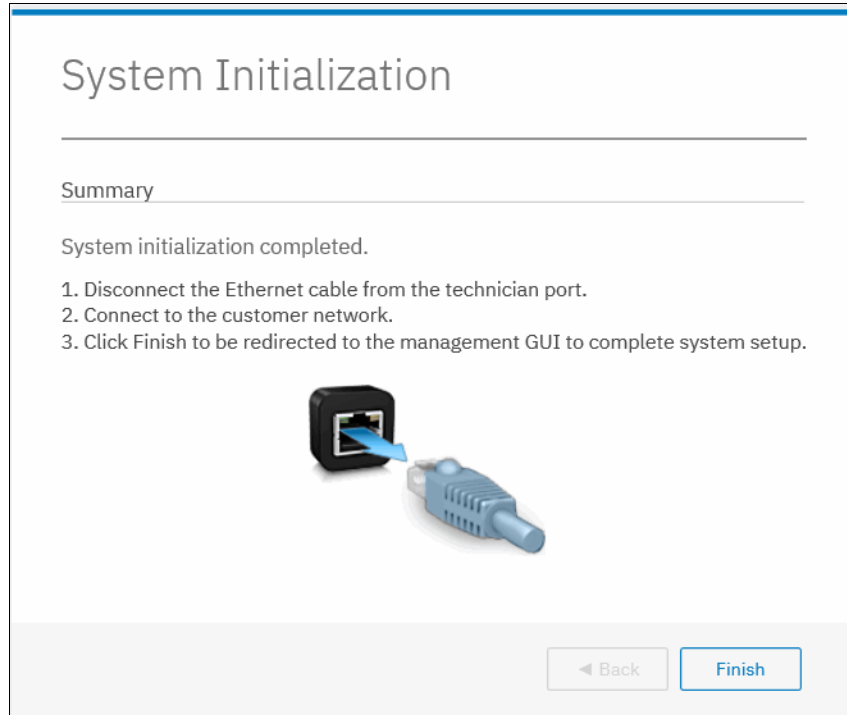


Figure 3-8 System Initialization: Complete

If you cannot connect to a network that has access to the management IP, you can continue the system setup from any other workstation that can reach it.

3.3 System setup

This section provides instructions about how to define the basic settings of the system by using the system setup wizard.

3.3.1 System setup wizard

After the initialization is complete and you are redirected to a management GUI from your PC or Notebook, or you browse to the management IP address of a freshly initialized system from another workstation, you must complete the system setup wizard to define the basic settings of the system.

Note: Experienced users can disable the system setup wizard and complete the configuration manually. However, this method is *not recommended* for most use cases.

- ▶ To disable the system setup wizard on a new system, run the following command:

```
chsystem -easysetup no
```

- ▶ During the setup wizard, you are prompted to change the default superuser password. If the wizard is bypassed, the system blocks the configuration functions until it is changed. All attempts at configuration return the following error:

```
CMMVC9473E The command failed because the superuser password must be changed before the system can be configured
```

- ▶ All configuration settings that are done by using the system setup wizard can be changed later by using the system GUI or CLI.

The first time that you connect to the management GUI, you are prompted to accept untrusted certificates because the system certificates are self-signed. If your company policy requests certificates that are signed by a trusted certificate authority (CA), you can install them after you complete the system setup. For more information about how to perform this task, see 3.5.1, “Configuring secure communications” on page 139.

To complete the system setup wizard, complete the following steps:

1. Log in to system GUI. Until the wizard is complete, you may use only *superuser* account, as shown in Figure 3-9. Click **Sign in**.

Note: The default password for the superuser account is `passw0rd` (with the number zero and not the capital letter O). The default password must be changed by using the system setup wizard or after the first CLI login. The new password cannot be set to the default one.

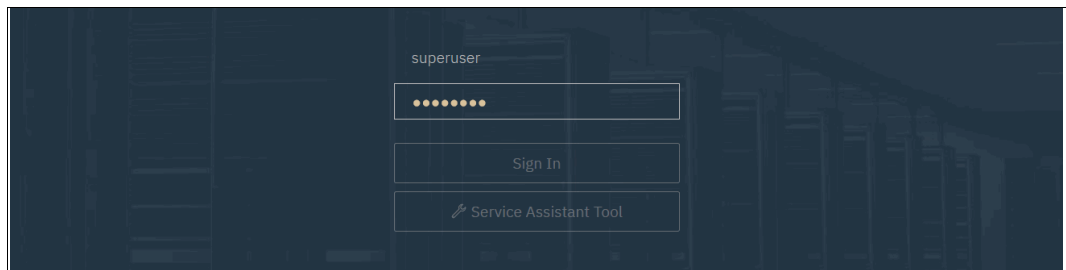


Figure 3-9 System Setup: Logging in for the first time

2. The welcome window opens, as shown in Figure 3-10. Verify the prerequisites and click **Next**.

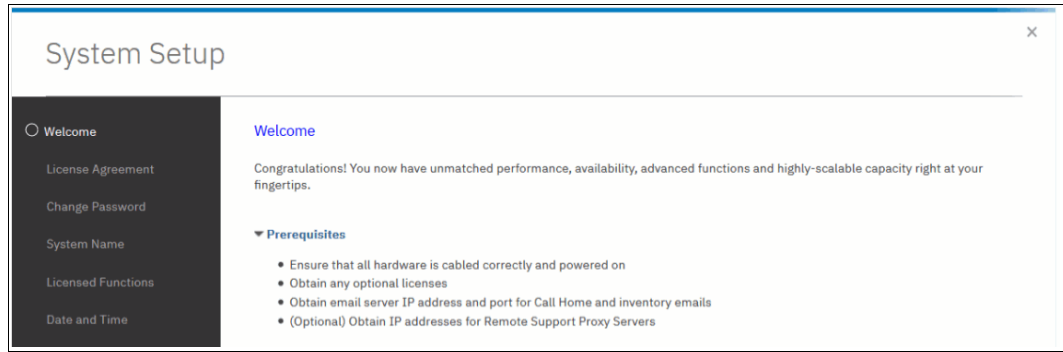


Figure 3-10 System Setup: Welcome

3. Carefully read the license agreement, select **I agree with the terms in the license agreement** if you want to continue the setup, as shown in Figure 3-11, and click **Next**.

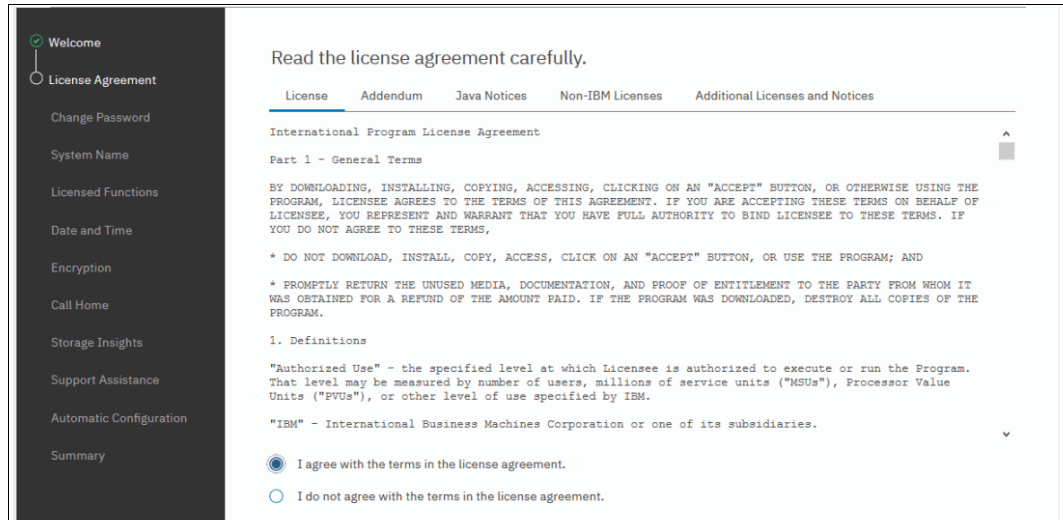


Figure 3-11 System Setup: License agreement

4. Enter a new password for the superuser, as shown in Figure 3-12. A valid password is 6 - 64 characters and cannot begin or end with a space. Also, the password cannot be set to match the default password. For more information, see 3.5.2, “Configuring password policies” on page 142. Click **Apply** and then **Next**.

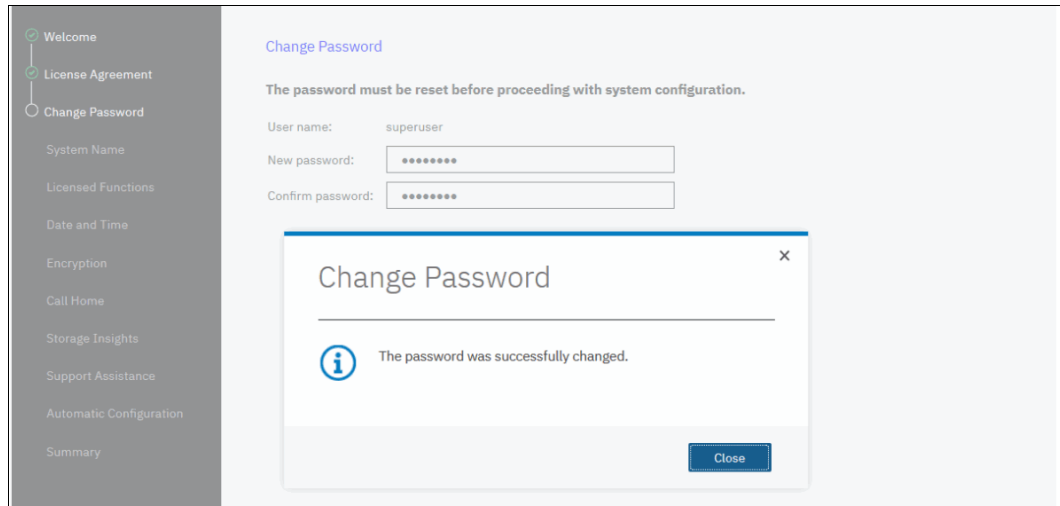


Figure 3-12 System Setup: Changing the password for the superuser

Note: All configuration changes that are done with the system setup wizard are applied immediately, including the password change.

5. Enter the name that you want to give the new system, as shown on Figure 3-13. Click **Apply** and then **Next**.

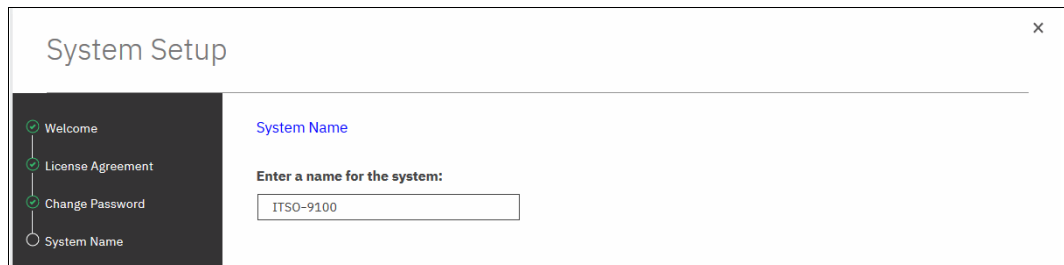


Figure 3-13 System Setup: Setting the system name

Avoid using an underscore (`_`) in a system name. While permitted here, it is not allowed in domain name server (DNS) shortnames and fully qualified domain names (FQDNs), so such naming might cause confusion and access issues. The following characters can be used: A - Z, a - z, 0 - 9, and - (hyphen).

Note: In a 3-Site replication solution, to prepare the IBM Spectrum Virtualize clusters at Master, AuxNear, and AuxFar sites to work, make sure that the system name is unique for all three clusters. The system names must remain different through the life of the 3-Site configuration.

If required, the system name can be changed by running the `chsystem -name <new_system_name>` command.

- Enter the number of licensed enclosures or licensed capacity for each function, as shown on Figure 3-14.

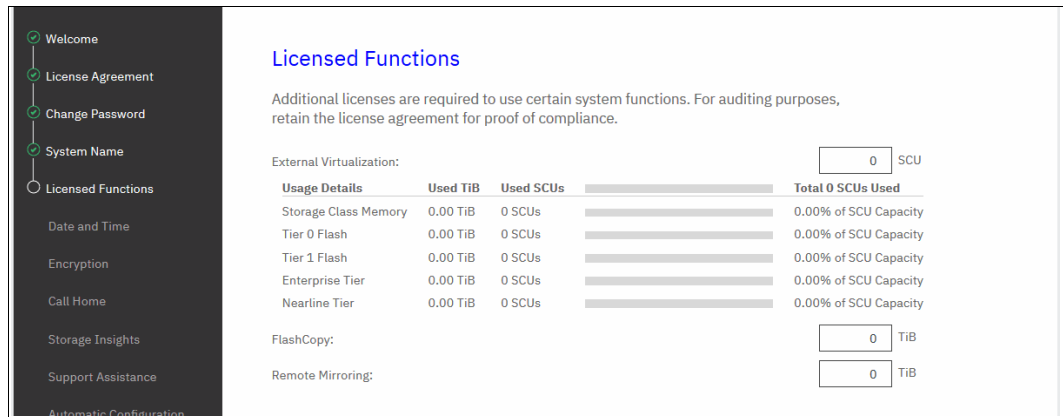


Figure 3-14 System Setup: Setting the system licenses

Note: IBM FlashSystem 5010 and IBM FlashSystem 5030 work with Licensed Internal Code (LIC). All licenses are controller-based. There is no capacity or enclosure licensing.

The IBM FlashSystem 5100 system follows an enclosure-based licensing scheme that allows the use of certain licensed functions on the number of enclosures (control and expansion) that is indicated in the license.

IBM FlashSystem 7200, IBM FlashSystem 9100, and IBM FlashSystem 9200 systems use differential and capacity-based licensing. For external virtualization, differential licensing offers different pricing rates for different types of storage and is based on the number of storage capacity units (SCUs) that are purchased. For other licensed functions, the system supports capacity-based licensing.

Make sure that the numbers you enter here match the numbers in your license authorization papers. For more information, see 1.17, “Licensing” on page 66.

When done, click **Apply** and then **Next**.

Note: Encryption uses a key-based licensing scheme, and it is activated later in the wizard.

- Enter the date and time settings. In the example that is shown in Figure 3-15, the date and time are set by using an NTP server. Generally, use an NTP server so that all of your storage area network (SAN) and storage devices have a common timestamp. This practice facilitates troubleshooting and prevents time stamp-related errors if you use a key server as an encryption key provider.

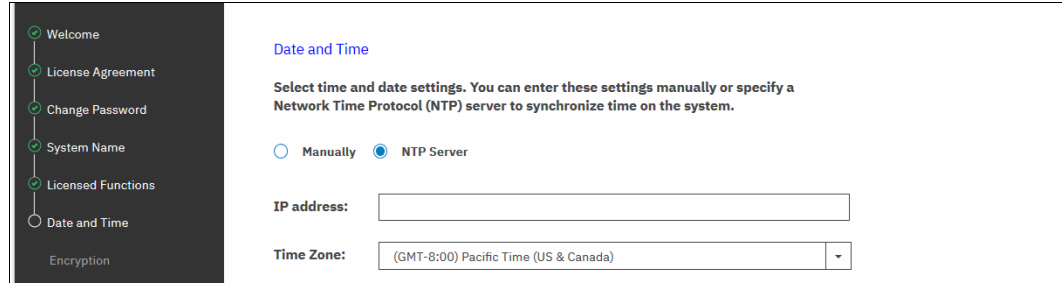


Figure 3-15 System Setup: Setting the date and time

If you choose to manually enter these settings, you are prompted to input the date, time, and time zone, or you can take those settings from your web browser. You cannot use a 24-hour clock system here, but you can switch to it later by using the system GUI.

When the data is set, click **Apply** and then **Next**.

- Select whether the encryption feature was purchased for this system, as shown in Figure 3-16.

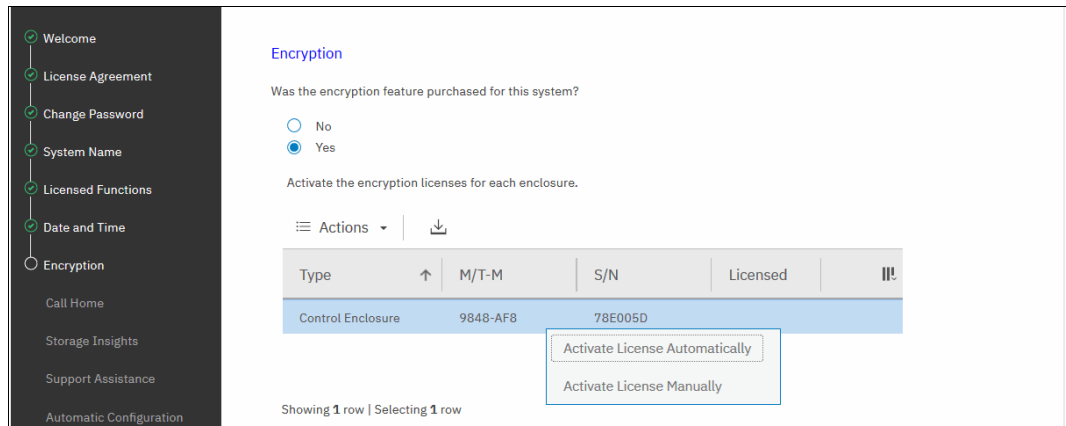


Figure 3-16 System Setup: Encryption activation

If encryption is not planned at this moment, select **No** and click **Next**. You can enable this feature later, as described in Chapter 12, “Encryption” on page 735.

If you purchased the encryption feature, you are prompted to activate your license manually or automatically. The encryption license is key-based and required for each control enclosure.

You can use automatic activation if the PC or Notebook that you use to connect to the GUI and run the system setup wizard has internet access. If no internet connection is available, use manual activation and follow the instructions. For more information, see Chapter 12, “Encryption” on page 735.

After the encryption license is activated, you see a green check mark for each enclosure, as shown in Figure 3-17 on page 119. After all the control enclosures show that encryption is licensed, click **Next**.

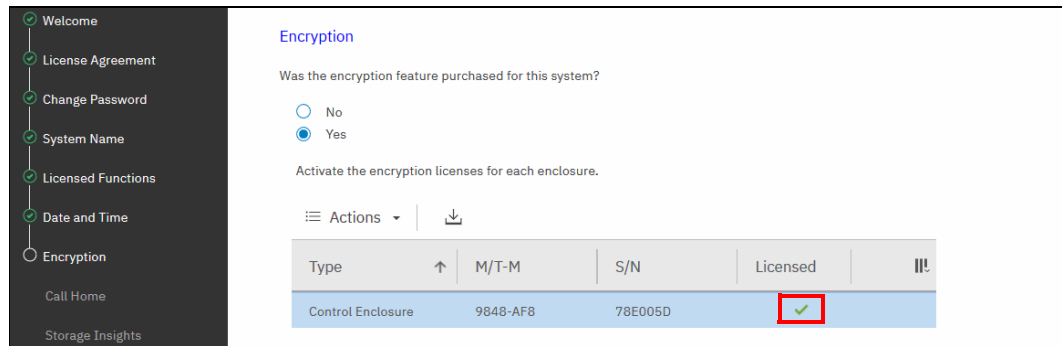


Figure 3-17 System Setup: Encryption licensed

- Set up the Call Home functions, as shown in Figure 3-18. With Call Home enabled, IBM automatically opens problem reports and contacts you to verify whether replacement parts are required.

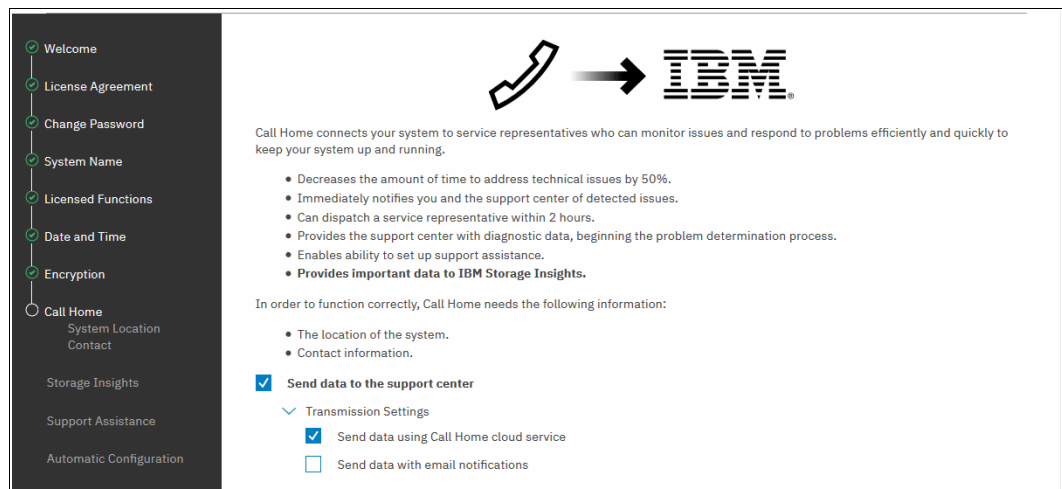


Figure 3-18 System Setup: Call Home methods

Note: It is a best practice to configure Call Home and keep it enabled if your system is under warranty or if you have a hardware maintenance agreement.

On IBM FlashSystem 9100 and IBM FlashSystem 9200 systems, an IBM SSR configures Call Home during installation. You need to check only whether all the entered data is correct.

The system supports two methods of sending Call Home notifications to IBM:

- Cloud Call Home
- Call Home with email notifications

Cloud Call Home is the default and preferred option for a system to report event notifications to IBM Support. With this method, the system uses RESTful application programming interfaces (APIs) to connect to an IBM centralized file repository that contains troubleshooting information that is gathered from customers. This method requires no extra configuration.

The system may also be configured to use email notifications for this purpose. If this method is selected, you are prompted to enter the SMTP server IP address.

If both methods are enabled, cloud Call Home is used, and the email notifications method is kept as a backup.

For more information about setting up Call Home, including Cloud Call Home, see Chapter 13, “Reliability, availability, and serviceability, monitoring and logging, and troubleshooting” on page 793.

If either of these methods is selected, the system location and contact information must be entered. This information is used by IBM to provide technical support. All fields in the form must be populated. In this step, the system also verifies that it can contact the Cloud Call Home servers, as shown in Figure 3-19.

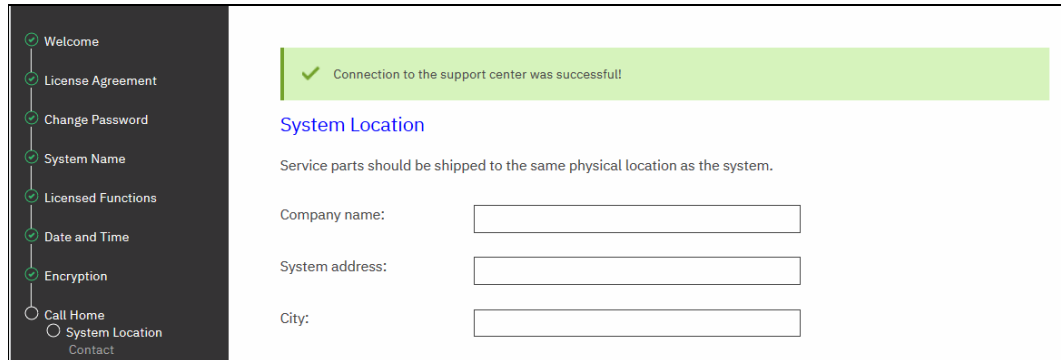


Figure 3-19 System Setup: System location

After clicking **Next**, you can provide business-to-business contact information that IBM Support uses to contact a person who manages this machine if it is necessary, as shown in Figure 3-20.

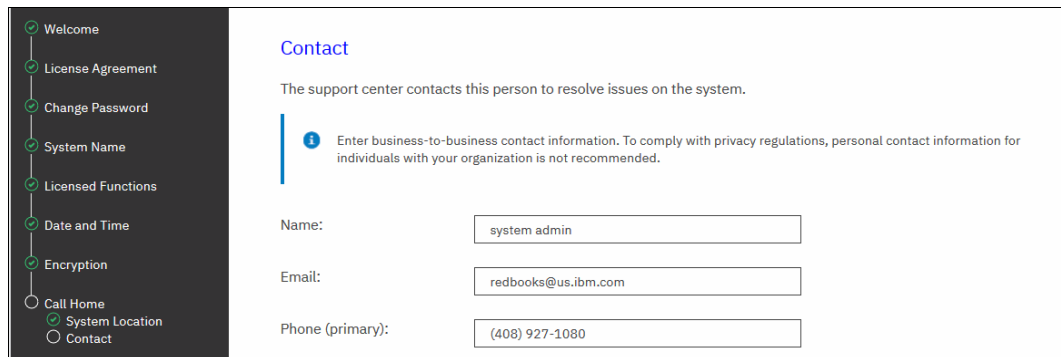


Figure 3-20 System Setup: Contact information

If the **Email notifications** option was selected, you are prompted to enter the details for the email servers to be used for Call Home. Figure 3-21 on page 121 shows an example. You can click **Ping** to verify that the email server is reachable over the network. Click **Apply** and then **Next**.

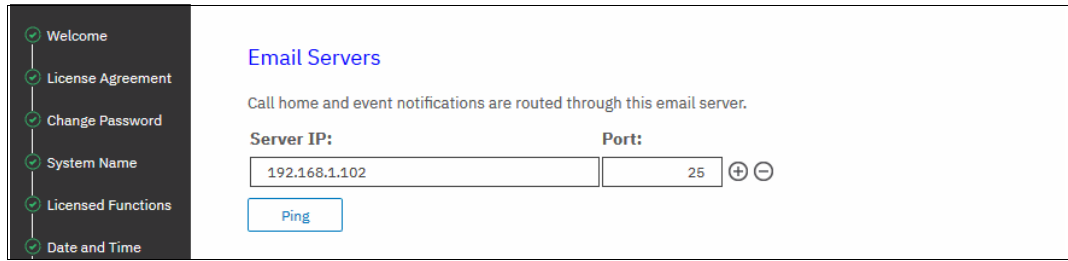


Figure 3-21 System Setup: Email servers

10. IBM FlashSystem family systems may be used with IBM Storage Insights, which is an IBM cloud storage monitoring and management tool. During this setup phase, the system tries to contact the IBM Storage Insights web service. If it is available, you are prompted to sign up, as shown in Figure 3-22.

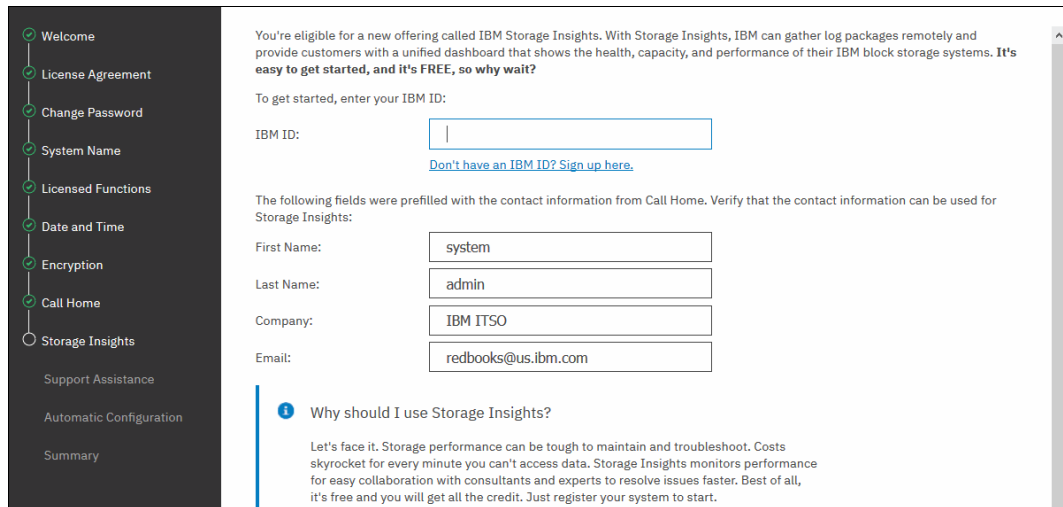


Figure 3-22 System Setup: IBM Storage Insights

If a connection cannot be established, you are prompted to add the system that you are currently working on to the IBM Storage Insights setup manually, as shown in Figure 3-23.

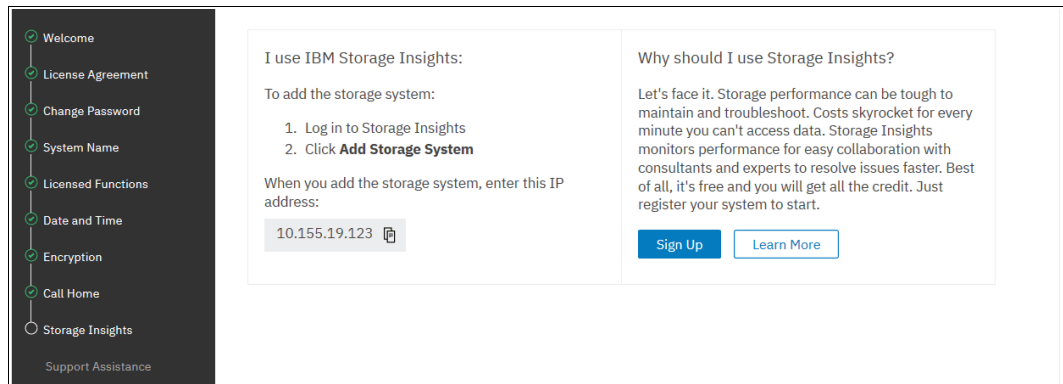


Figure 3-23 System Setup: IBM Storage Insights

For more information about IBM Storage Insights, see Chapter 13, “Reliability, availability, and serviceability, monitoring and logging, and troubleshooting” on page 793.

11. After you click **Next**, if you enabled at least one Call Home method, the Support Assistance configuration window opens, as shown in Figure 3-24. The Support Assistance function requires Call Home, so if it is disabled, Support Assistance cannot be used.

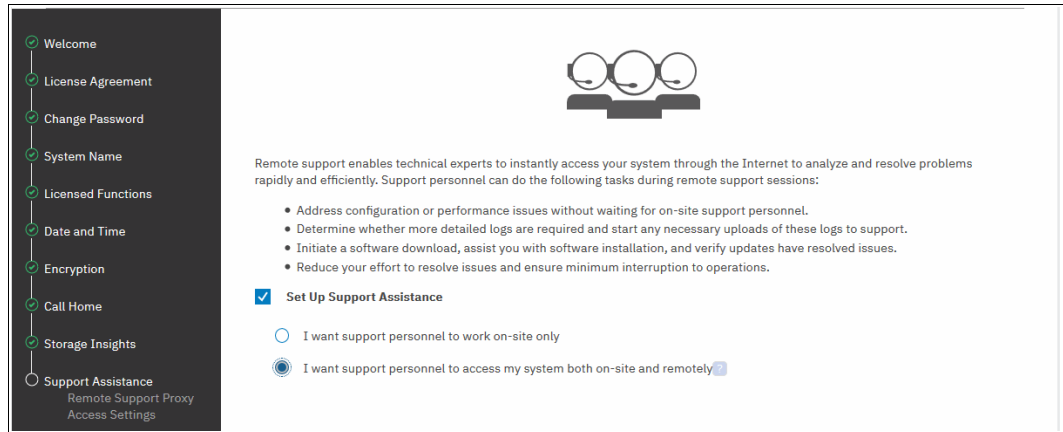


Figure 3-24 System Setup: Support Assistance

With the Support Assistance feature, you allow IBM Support to perform maintenance tasks on your system while an IBM SSR is onsite. The IBM SSR can log in locally with your permission and a special user ID and password so that a superuser password does not need to be shared with the IBM SSR.

You can also enable Support Assistance with remote support to allow IBM Support personnel to log in remotely to the machine with your permission through a secure tunnel over the internet.

For more information about the Support Assistance feature, see Chapter 13, “Reliability, availability, and serviceability, monitoring and logging, and troubleshooting” on page 793.

If you allow remote support, you are given the IP addresses and ports of the remote support centers and an opportunity to provide proxy server details (if required) to allow the connectivity, as shown in Figure 3-25. Also, you can allow remote connectivity at any time or only after obtaining permission from the storage administrator.

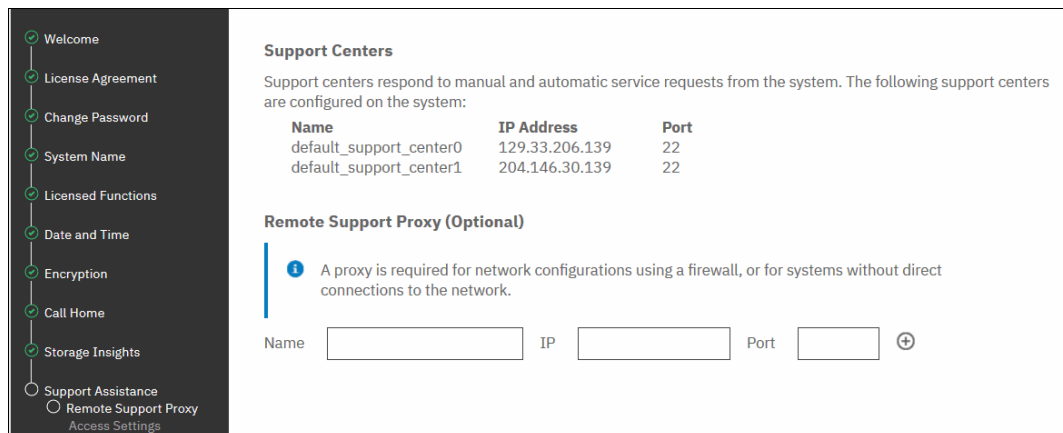


Figure 3-25 System Setup: Support Centers

12. As the last initial system setup step, you are prompted to perform automatic configuration for the system that you will use as FC-attached back-end storage for IBM SAN Volume Controller (SVC).

If you plan to use the system in stand-alone mode (not behind an SVC), leave **Automatic Configuration** turned off, as shown in Figure 3-26. If your solution design later changes and the system becomes an SVC back end, you can run automatic configuration later by using the GUI.

If you turn on automatic configuration, after the system setup completes, the system redirects you to the Automatic Configuration for Virtualization wizard, which is described in 3.4.6, “Automatic configuration for IBM SAN Volume Controller back-end storage” on page 136.

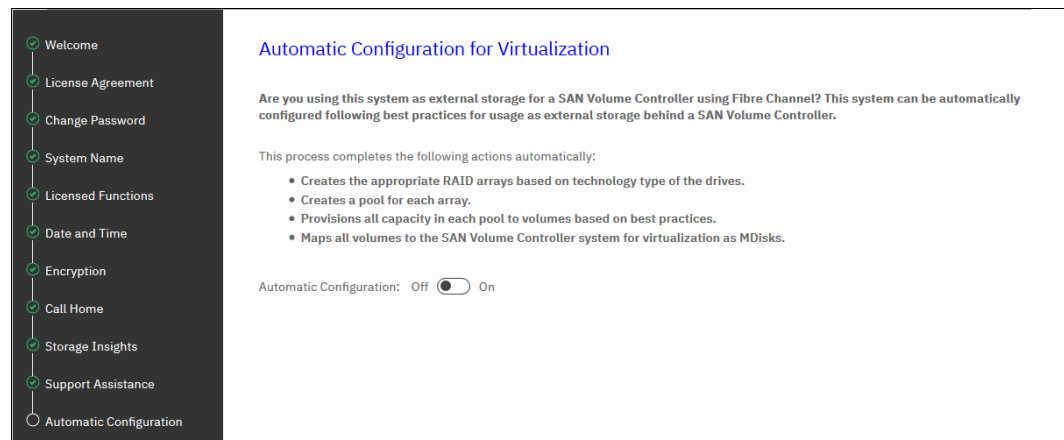


Figure 3-26 System Setup: Automatic configuration for IBM SAN Volume Controller

13. On the Summary page, the settings that were set by the system setup wizard are shown. If corrections are needed, you may return to a previous step by clicking **Back**. Otherwise, click **Finish** to be redirected to a system GUI.

After the wizard completes, your system consists only of the control enclosure that includes the node canister that you used to initialize the system and its partner, and the expansion enclosures that are attached to them. If you have other control and expansion enclosures, you must add them to complete the system setup. For more information about how to add a control or expansion enclosure, see 3.4.2, “Adding an enclosure” on page 127.

If you have no more enclosures to add to this system, the system setup process is complete. All the mandatory steps of the initial configuration are done. If required, you can configure other global functions, such as system topology, user authentication, or local port masking, before configuring the volumes and provisioning them to hosts.

3.4 Base configuration

Tasks that are listed in this section are used to define global system configuration settings. Often, they are performed during system setup. However, they also can be performed any time later, such as when the system is expanded or the system environment is reconfigured.

3.4.1 Configuring Remote Direct Memory Access clustering

Up to four control enclosures may be joined in an IBM HyperSwap or a standard topology cluster. This subsection describes the configuration steps that must be performed if a system is designed for IP-based Remote Direct Memory Access (RDMA) node-to-node traffic.

For FC SAN clustering, no special configuration is required on the system, but the SAN must be set up as described in Chapter 2, “Planning” on page 71.

Prerequisites

Before RDMA clustering is configured, ensure that the following prerequisites are met:

- ▶ 25 gigabits per second (Gbps) RDMA-capable Ethernet cards are installed in each node.
- ▶ RDMA-capable adapters in all nodes use the same technology, such as RDMA over Converged Ethernet (RoCE) or internet Wide Area RDMA Protocol (iWARP).
- ▶ RDMA-capable adapters are installed in the same slots across all the nodes of the system.
- ▶ Ethernet cables between each node are connected correctly.
- ▶ The network configuration does not contain more than two hops in the fabric of switches. The router must *not* be placed between nodes that use RDMA-capable Ethernet ports for node-to-node communication.
- ▶ The negotiated speeds on the local and remote adapters are the same.
- ▶ The local and remote port (RPORT) virtual local area network (VLAN) identifiers are the same. All the ports that are used for node-to node communication must be assigned to one VLAN ID, and ports that are used for host attachment must have a different VLAN ID. If you plan to use VLAN to create this separation, you must configure VLAN support on the all the Ethernet switches in your network before you define the RDMA-capable Ethernet ports on nodes in the system. On each switch in your network, set the VLAN to Trunk mode and specify the VLAN ID for the RDMA-ports that will be in the same VLAN.
- ▶ A minimum of two dedicated RDMA-capable Ethernet ports are required for node-to-node communications to ensure best performance and reliability. These ports must be configured for inter-node traffic only and must not be used for host attachment, virtualization of Ethernet-attached external storage, or IP replication traffic.
- ▶ A maximum of four RDMA-capable Ethernet ports per node are allowed for node-to-node communications.

Configuration process

To enable RDMA clustering, IP addresses must be configured on each port of each node that is used for node-to-node communication. Complete the following steps:

1. Connect to a Service Assistant of a node by going to https://<node_service_IP>/service and clicking **Change Node IP**, as shown in Figure 3-27 on page 125.

Collect Logs	<div style="border: 1px solid red; padding: 2px;">The operation completed successfully.</div>		
Manage System	Node IP Address 1	Node IP Address 2	Node IP Address 3
Recover System	Port ID: 1	Port ID: 2	Port ID: 3
Re-install Software	RDMA Type:	RDMA Type:	RDMA Type:
Update Manually	Port Speed: 10Gb/s	Port Speed: 10Gb/s	Port Speed: 10Gb/s
Configure Node	Link State: Active	Link State: Active	Link State: Active
Change Service IP	State: Unconfigured	State: Unconfigured	State: Unconfigured
Change Node IP	Node IP Address:	Node IP Address:	Node IP Address:
Change Node Discovery Subnet	Subnet Mask:	Subnet Mask:	Subnet Mask:
Ethernet Connectivity	Gateway:	Gateway:	Gateway:
Configure CLI Access	VLAN:	VLAN:	VLAN:
Restart Service	Node IP Address 4	Node IP Address 5	Node IP Address 6
	Port ID: 4	Port ID: 5	Port ID: 6
	RDMA Type:	RDMA Type: RoCE	RDMA Type: RoCE
	Port Speed: 10Gb/s	Port Speed: 25Gb/s	Port Speed: 25Gb/s
	Link State: Active	Link State: Active	Link State: Active
	State: Unconfigured	State: Unconfigured	State: <input type="button" value="Modify"/> <input type="button" value="Unconfigure"/>
	Node IP Address:	Node IP Address:	Node IP Address:
	Subnet Mask:	Subnet Mask:	Subnet Mask:
	Gateway:	Gateway:	Gateway:
	VLAN:	VLAN:	VLAN:

Figure 3-27 Node IP address setup for Remote Direct Memory Access clustering

Figure 3-27 shows that ports 1 - 4 do not show any RDMA type, so they cannot be used for node-to-node traffic. Ports 5 and 6 show RDMA type RoCE, so they can be used.

2. Hover your cursor over a tile with a port and click **Modify** to set the IP address, netmask, gateway address, and VLAN ID for a port. The IP address for each port must be unique and cannot be used anywhere else on the system. The VLAN ID for ports that are used for node-to-node traffic must be the same on all nodes. When the required information is entered, click **Save** and verify that the operation completed successfully, as shown in Figure 3-28. Repeat this step for all ports that you intend to use for node-to-node traffic, with a minimum of two and a maximum of four ports per node.

Collect Logs	<div style="border: 1px solid red; padding: 2px;">The operation completed successfully.</div>		
Manage System	Node IP Address 1	Node IP Address 2	Node IP Address 3
Recover System	Port ID: 1	Port ID: 2	Port ID: 3
Re-install Software	RDMA Type:	RDMA Type:	RDMA Type:
Update Manually	Port Speed: 10Gb/s	Port Speed: 10Gb/s	Port Speed: 10Gb/s
Configure Node	Link State: Active	Link State: Active	Link State: Active
Change Service IP	State: Unconfigured	State: Unconfigured	State: Unconfigured
Change Node IP	Node IP Address:	Node IP Address:	Node IP Address:
Change Node Discovery Subnet	Subnet Mask:	Subnet Mask:	Subnet Mask:
Ethernet Connectivity	Gateway:	Gateway:	Gateway:
Configure CLI Access	VLAN:	VLAN:	VLAN:
Restart Service	Node IP Address 4	Node IP Address 5	Node IP Address 6
	Port ID: 4	Port ID: 5	Port ID: 6
	RDMA Type:	RDMA Type: RoCE	RDMA Type: RoCE
	Port Speed: 10Gb/s	Port Speed: 25Gb/s	Port Speed: 25Gb/s
	Link State: Active	Link State: Active	Link State: Active
	State: Unconfigured	State: Configured	State: Configured
	Node IP Address:	Node IP Address: 10.0.99.11	Node IP Address: 192.168.59.10
	Subnet Mask:	Subnet Mask: 255.255.255.0	Subnet Mask: 255.255.255.0
	Gateway:	Gateway: 10.0.99.0	Gateway: 192.168.59.1
	VLAN:	VLAN:	VLAN:

Figure 3-28 Node IP addresses configured

To list and change the node IP configuration by using the CLI, run the **sainfo lsnodeip** and **satask chnodeip** commands, as shown in Example 3-2.

Example 3-2 Setting the IP addresses for node-to-node connectivity

```

IBM_IBM FlashSystem:ITS0-FS9100:superuser>sainfo lsnodeip
port_id    rdma_type port_speed vlan link_state state      node_IP_address
1          10Gb/s    active    unconfigured
2          10Gb/s    active    unconfigured
3          10Gb/s    active    unconfigured
4          10Gb/s    active    unconfigured
5          RoCE     25Gb/s    active    unconfigured
6          RoCE     25Gb/s    active    unconfigured
IBM_IBM FlashSystem:ITS0-FS9100:superuser>satask chnodeip -ip 10.0.99.12 -gw
10.0.99.1 -mask 255.255.255.0 -port_id 5
IBM_IBM FlashSystem:ITS0-FS9100:superuser>satask chnodeip -ip 192.168.59.11 -gw
192.168.59.1 -mask 255.255.255.0 -port_id 6
IBM_IBM FlashSystem:ITS0-FS9100:superuser>sainfo lsnodeip
port_id    rdma_type port_speed vlan link_state state      node_IP_address
1          10Gb/s    active    unconfigured
2          10Gb/s    active    unconfigured
3          10Gb/s    active    unconfigured
4          10Gb/s    active    unconfigured
5          RoCE     25Gb/s    active    configured  10.0.99.12
6          RoCE     25Gb/s    active    configured  192.168.59.11

```

- Some environments might not include a stretched layer 2 subnet. In such scenarios, a layer 3 network such as in standard topologies or long-distance RDMA node-to-node HyperSwap configurations is applicable. To support the layer 3 Ethernet network, use the unicast discovery method for RDMA node-to-node communication. This method relies on unicast-based fabric discovery rather than multicast discovery.

To configure unicast discovery, see the information about the **satask addnodediscoverysubnet**, **satask rmnodediscoverysubnet**, or **sainfo lsnodediscoverysubnet** commands in [Command-line Interface](#). You can also configure discovery subnets by using the Service Assistant interface menu option **Change Node Discovery Subnet**, as shown in Figure 3-29.

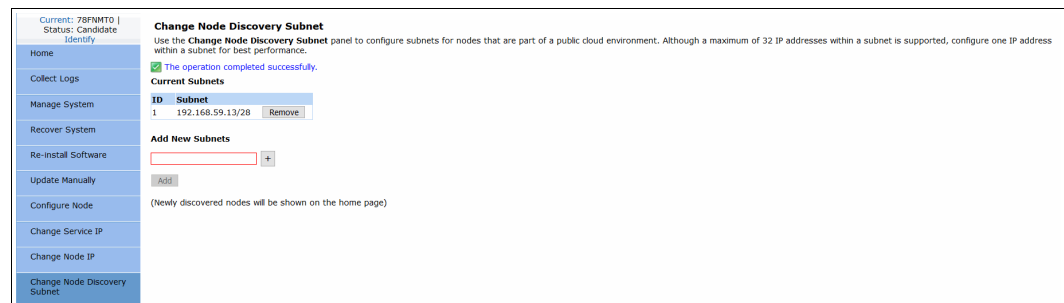


Figure 3-29 Setting the node discovery subnet

- After the IP addresses are configured on all nodes in a system, run the **sainfo 1snodeipconnectivity** command or use the Service Assistant GUI menu **Ethernet Connectivity** to verify that the partner nodes are visible on the IP network, as shown in Figure 3-30. If necessary, troubleshoot connection problems by running the **ping** and **sainfo traceroute** commands.

The screenshot shows the 'Ethernet Connectivity' section of a GUI. It includes a sidebar with navigation options like 'Home', 'Collect Logs', 'Manage System', etc. The main area displays a table with columns: Cluster ID, Status, Error Data, Port, ID, Type, IP Address, VLAN, and WWNN. Two rows of data are visible, both showing 'Connected:RoCE' status for different ports and IP addresses.

Cluster ID	Status	Error Data	Port	ID	Type	IP Address	VLAN	WWNN
000002032C411602	Connected:RoCE		Local 5	5	RoCE	192.168.59.10		
			Remote 5	5	RoCE	192.168.59.11		500507680C008B01
000002032C411602	Connected:RoCE		Local 6	6	RoCE	10.0.99.11		
			Remote 6	6	RoCE	10.0.99.12		500507680C008B01

Figure 3-30 Node-to-node Ethernet connectivity

When all the nodes that are joined to the cluster are connected, the enclosure may be added to the cluster.

3.4.2 Adding an enclosure

This procedure is the same whether you are configuring the system for the first time or expanding it later. When performed by using the system GUI, the same steps are used for adding expansion or control enclosures.

Before beginning this process, ensure that the new control enclosure is correctly installed and cabled to the existing system. For FC node-to-node communication, verify that correct the SAN zoning is set. For node-to-node communication over RDMA-capable Ethernet ports, ensure that the IP addresses are configured and a connection between nodes can be established.

To add an enclosure to the system, complete the following steps:

- In the GUI, select **Monitoring** → **System**. When a new enclosure is detected by a system, the **Add Enclosure** button appears on the System - Overview window next to System Actions, as shown in Figure 3-31.



Figure 3-31 Add Enclosure button

Note: If the **Add Enclosure** button does not appear, review the installation instructions to verify that the new enclosure is connected and set up correctly.

2. Click **Add Enclosure**, and a list of available candidate enclosures opens, as shown in Figure 3-32. To light the Identify light-emitting diode (LED) on a selected enclosure, select **Actions** → **Identify**. When the required enclosure (or enclosures) is chosen, click **Next**.

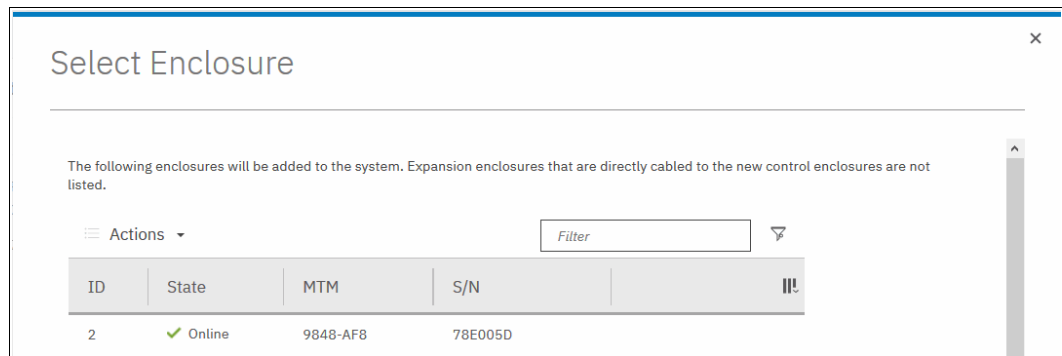


Figure 3-32 Selecting the control enclosure to add

3. Review the summary in the next window and click **Finish** to add the expansion enclosure or the control enclosure and all expansions that are attached to it to the system.

Note: When a new control enclosure is added, the software version running on its nodes is upgraded or rolled back to match the system software version. This process can take up to 30 minutes or more, and the enclosure is added only when this process completes.

4. After the control enclosure is successfully added to the system, a success message appears. Click **Close** to return to the System Overview window and check that the new enclosure is visible and available for management.

To perform the same procedure by using a CLI, complete the following steps. For more information about the detailed syntax for each command, go to [Command-line Interface](#).

1. When adding control enclosures, check for unpopulated I/O groups by running `lsiogrp`. Each control enclosure has two nodes, so it forms an I/O group. Example 3-3 shows that only `io_grp0` has nodes, so a new control enclosure can be added to `io_grp1`.

Example 3-3 Listing the I/O groups

```
IBM_IBM FlashSystem:ITS0-FS9100:superuser>lsiogrp
id name          node_count vdisk_count host_count site_id site_name
0  io_grp0         2           0           0           0
1  io_grp1         0           0           0           0
2  io_grp2         0           0           0           0
3  io_grp3         0           0           0           0
4  recovery_io_grp 0           0           0
```

2. To list control enclosures that are available to add, run the `lscontrolenclosurecandidate` command, as shown in Example 3-4 on page 129. To list the expansion enclosures, run the `lseclosure` command. Expansions that have the `managed` parameter set to `no` are available for addition.

Example 3-4 Listing the candidate control enclosures

```
IBM_IBM FlashSystem:ITS0-FS9100:superuser>lscontrolenclosurecandidate
serial_number product_MTM machine_signature
78E005D      9848-AF8    4AD2-EA69-8B5E-D0C0
```

3. Add a control enclosure by running the **addcontrolenclosure** command, as shown in Example 3-5. The command triggers only the process, which starts in background and can take up to 30 minutes or more.

Example 3-5 Adding a control enclosure

```
IBM_IBM FlashSystem:ITS0-FS9100:superuser>addcontrolenclosure -iogrp 1 -sernum
78E005D
```

4. To add an expansion enclosure, change its status to managed = yes by running the **chenclosure** command, as shown in Example 3-6.

Example 3-6 Adding an expansion enclosure

```
IBM_IBM FlashSystem:ITS0-FS9100:superuser>lsenclosure
id status type      managed IO_group_id IO_group_name product_MTM serial_number
1  online control   yes    0          io_grp0     9848-AF8   78E006A
2  online expansion no     0          io_grp0     9848-AFF   78CBVF5
IBM_IBM FlashSystem:ITS0-FS9100:superuser>chenclosure -managed yes 2
```

3.4.3 Changing the system topology to HyperSwap

Note: HyperSwap over IP is not supported by IBM FlashSystem 5030.

The HyperSwap function is a high availability (HA) feature that provides dual-site, active-active access to a volume. You can create an HyperSwap topology system configuration where each I/O group in the system is physically on a different site. When these configurations are used with HyperSwap volumes, they can be used to maintain access to data on the system if site-wide outages occur.

If your solution is designed to use the HyperSwap function, use the guidance in this section to configure a cluster for a multi-site HyperSwap topology.

For a list of requirements for a HyperSwap configuration with FC or RDMA-capable Ethernet connections, see [IBM Documentation](#) and expand **Configuring** → **Configuration details** → **HyperSwap system configuration details**.

To change the system topology to HyperSwap, complete the following steps:

1. In the GUI, click **Monitoring** → **System** to open the System - Overview window. Click **System Actions** and expand **Modify System Topology**, as shown in Figure 3-33.



Figure 3-33 Starting the Modify System Topology wizard

- The Modify Topology wizard welcome window opens. Click **Next**. You are prompted to change the default site names, as shown in Figure 3-34. The site names can indicate, for example, building locations for each site, or other descriptive information.

Assign Site Names

Enter the names:

Site 1:

Site 2:

Site 3 (quorum):

Figure 3-34 Assigning site names

- Assign I/O groups to sites. Click the marked icons in the center of the window to swap site assignments, as shown in Figure 3-35. Click **Next**.

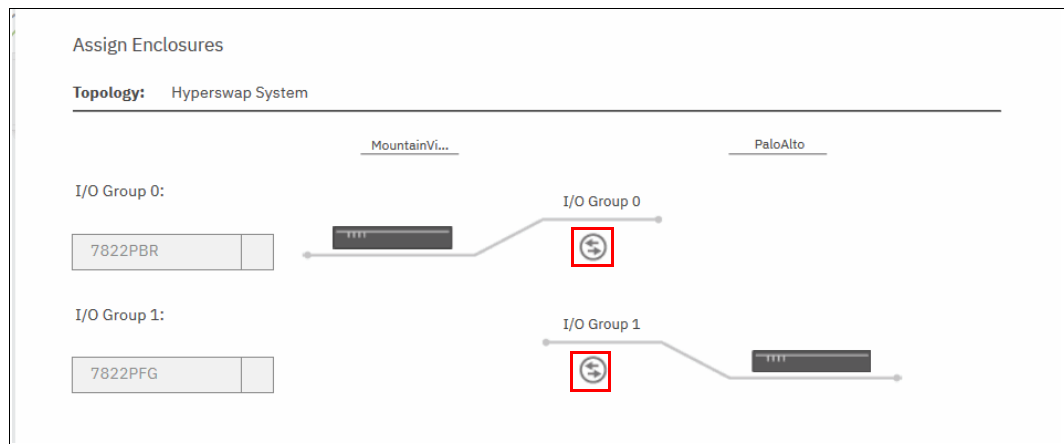


Figure 3-35 Specifying the system topology

- If any host objects or back-end storage controllers are configured, you must assign a site for each of them. Right-click the object and click **Modify Site**, as shown in Figure 3-36.

Assign Hosts to a Site

A host must have a site that is assigned to it before it can be mapped to a HyperSwap volume.

Actions

Name	Status	Site
ITSO-VMHOST-02	Online	
Windows-Host-01	Online	
Windows-Host-02	Online	

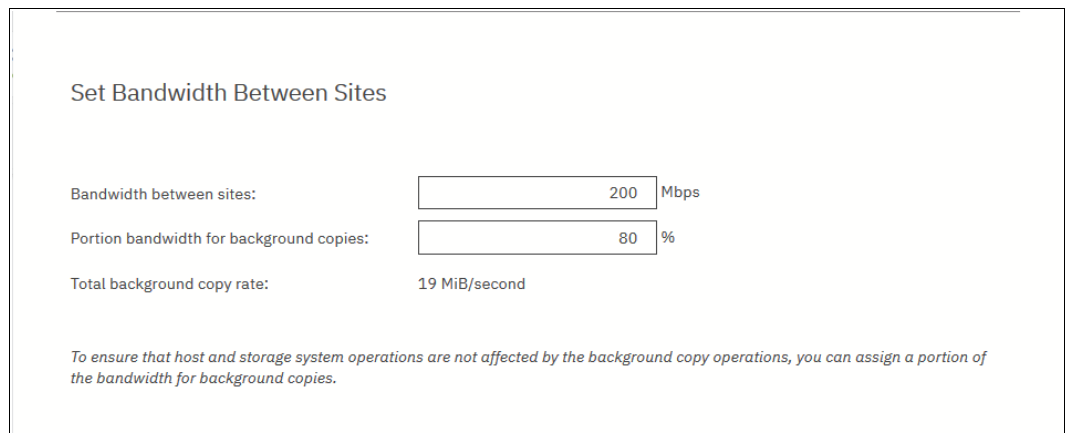
Modify Site

Properties

Figure 3-36 Assigning hosts to sites

5. Set the maximum background copy operations bandwidth between the sites. *Background copy* is the initial synchronization and any subsequent resynchronization traffic for HyperSwap volumes. Use this setting to limit the impact of volume synchronization to host operations. You may also set it higher during the initial setup (when there are no host operations on the volumes yet), and set it lower when the system is in production.

As shown in Figure 3-37, you must specify the total bandwidth between the sites in megabits per second (Mbps) and what percentage of this bandwidth that can be used for background copying. Click **Next**.



Set Bandwidth Between Sites

Bandwidth between sites: Mbps

Portion bandwidth for background copies: %

Total background copy rate: 19 MiB/second

To ensure that host and storage system operations are not affected by the background copy operations, you can assign a portion of the bandwidth for background copies.

Figure 3-37 Setting the bandwidth between the sites

6. Review the summary and click **Finish**. The wizard starts implementing changes to migrate the system to the HyperSwap solution.

When you later add a host or back-end storage controller objects, the GUI prompts you to set an object site during the creation process.

3.4.4 Configuring quorum disks or applications

Quorum devices are required for a system to hold a copy of important system configuration data. An internal drive of an IBM FlashCore Module (FCM), a managed disk (MDisk) from FC-attached external back-end storage, or a special application that is connected over an IP network may work as a quorum device.

One of these items is selected for the *active quorum* role, which is used to resolve failure scenarios where half the nodes on the system become unavailable or a link between enclosures is disrupted. The active quorum determines which nodes can continue processing host operations and to avoid a “split brain” condition, which happens when both halves of the system continue I/O processing independently of each other.

For systems with a single control enclosure, quorum devices are selected automatically. No special configuration actions are required. This function also applies for systems with multiple control enclosures, a standard topology, and virtualizing external storage.

For HyperSwap topology systems, an active quorum device must be on a third, independent site. Due to the costs that are associated with deploying a separate FC-attached storage device on a third site, an IP-based quorum device may be used for this purpose.

On a standard topology system with two or more control enclosures and no external storage, an active quorum device cannot be on an internal drive of an FCM. For such configurations, it is a best practice to deploy an IP-based quorum application.

Creating and installing an IP quorum application

To create and install an IP quorum application, complete the following steps:

1. Select **System** → **Settings** → **IP Quorum** to download the IP quorum application, as shown in Figure 3-38. If you are using IPv6 for management IP addresses, the **Download IPv6 Application** button is available and the IPv4 option is disabled.

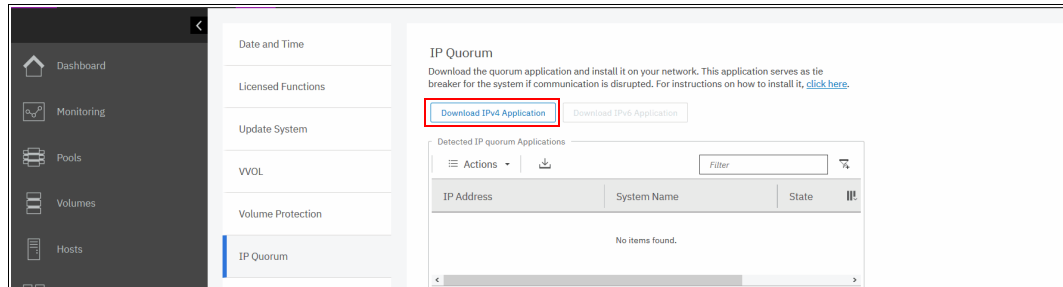


Figure 3-38 Download IPv4 quorum button

2. After you click **Download...**, a window opens, as shown in Figure 3-39. It provides an option to create an IP application that is used for tie-breaking only, or an application that can be used as a tie-breaker and to store recovery metadata.

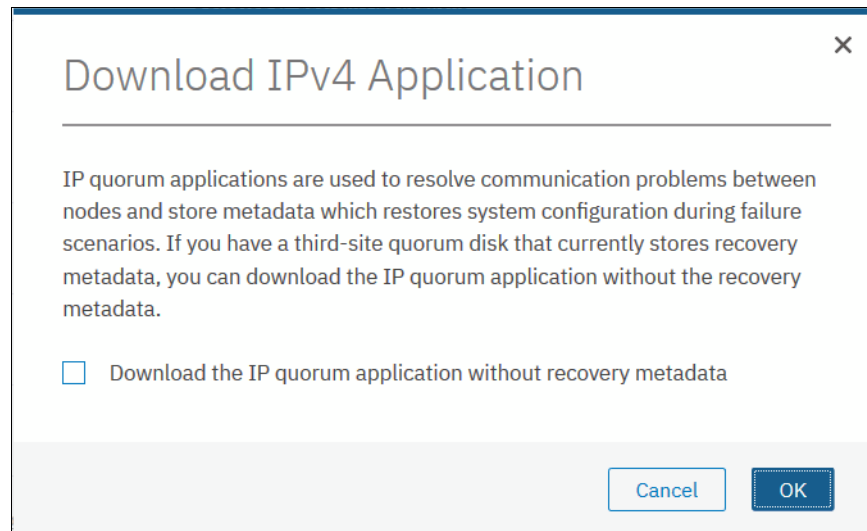


Figure 3-39 Download IP quorum application window

An application that does not store recovery metadata requires less channel bandwidth for a link between the system and the quorum app, which might be a decision-making factor for using a multi-site HyperSwap system.

For a full list of IP quorum app requirements, see [IBM Documentation](#) and expand **Configuring** → **Configuration details** → **Configuring quorum** → **IP quorum application configuration**.

- After you click **OK**, the `ip_quorum.jar` file is created. Save the file and transfer it to a supported AIX, Linux, or Windows host that can establish an IP connection to the service IP address of each system node. Move it to a separate directory and start the app, as shown in Example 3-7.

Example 3-7 Starting the IP quorum application on the Windows operating system

```
C:\IPQuorum>java -jar ip_quorum.jar
=== IP quorum ===
Name set to null.
Successfully parsed the configuration, found 4 nodes.
....
```

Note: Add the IP quorum application to the list of auto-started applications at each start or restart or configure your operating system (OS) to run it as an auto-started service in the background. The server that runs the IP quorum must be in the same subnet as the IBM FlashSystem. You can have a total of five IP quorums.

The IP quorum log file and recovery metadata are stored in the same directory with the `ip_quorum.jar` file.

- Check that the IP quorum application is successfully connected and running by verifying its Online status by selecting **System** → **Settings** → **IP Quorum**, as shown in Figure 3-40.

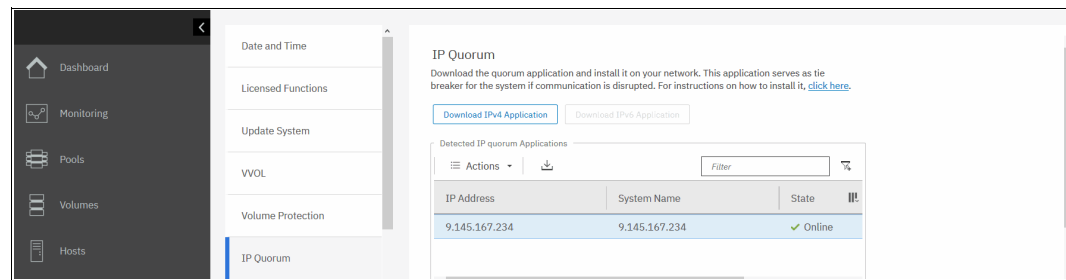


Figure 3-40 IP quorum application that is deployed and connected

Configuring the IP quorum mode

On a standard topology system, only the Standard quorum mode is supported. No additional configuration is required. On a HyperSwap topology, you may configure different tie-breaker scenarios (a tie occurs when exactly half of the nodes that were previously a member of the system are present):

- ▶ If the quorum mode is set to **Standard**, both sites have an equal chance to continue working after the tie-breaker.
- ▶ If the quorum mode is set **Preferred**, during a disruption, the system delays processing tie-breaker operations on non-preferred sites, leaving more time for the preferred site to win. If during an extended period a preferred site cannot contact the IP quorum app (for example, if it is destroyed), a non-preferred site continues working.
- ▶ If the quorum mode is set to **Winner**, the selected site always is the tie-breaker winner. If the winner site is destroyed, the remaining site may continue operating only after manual intervention.

The Preferred and Winner quorum modes are supported only with an IP quorum. For a FC-attached active quorum MDisk, only Standard mode is possible.

To set a quorum mode, select **System** → **Settings** → **IP Quorum** and click **Quorum Setting**. The Quorum Setting window opens, as shown in Figure 3-41.

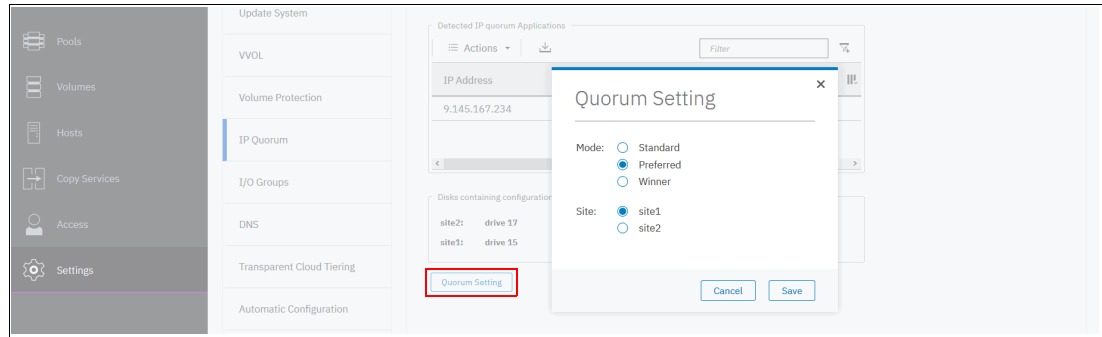


Figure 3-41 Changing the quorum mode

3.4.5 Configuring the local Fibre Channel port masking

With FC port masking, you control the usage of FC ports. By applying a mask, you restrict node-to-node communication or FC RC traffic on selected ports.

To decide whether your system must have port masks configured, see 2.6.8, “Port designation recommendations” on page 81.

To set the FC port mask by using the GUI, complete the following steps:

1. Select **System** → **Network** → **Fibre Channel Ports**. In a displayed list of FC ports, the ports are grouped by a system port ID. Each port is configured identically across all nodes in the system. You can click the arrow next to the port ID to expand a list and see which node ports (N_Port) belong to the selected system port ID and their worldwide port names (WWPNs).
2. Right-click a system port ID that you want to change and select **Modify Connection**, as shown in Figure 3-42.

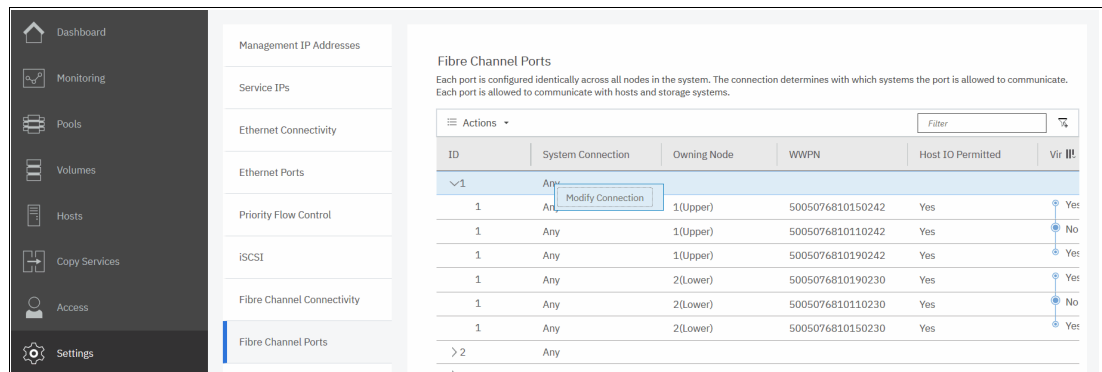


Figure 3-42 Applying a port mask by using a GUI

By default, all system ports can send and receive traffic of any kind:

- ▶ Host traffic
- ▶ Traffic to virtualized back-end storage systems
- ▶ Local system traffic (node-to-node)
- ▶ Partner system (remote replication) traffic

The first two types are always allowed, and you may control them only with SAN zoning. The other two types can be blocked by port masking. In the Modify Connection dialog box, as shown in Figure 3-43, you can choose which type of traffic that a port can send.

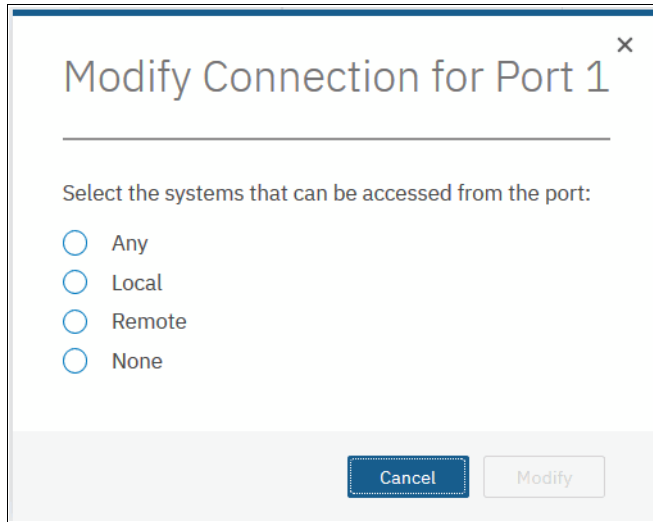


Figure 3-43 Modify Connection dialog box

Any	A port can work with all four types.
Local	Remote replication traffic is blocked on this port.
Remote	Blocks local node-to-node traffic.
None	Both local and remote systems traffic, but there is still system to host and system to back-end storage communication.

Port masks can also be set by using the CLI. Local and remote partner port masks are internally represented as a string of zeros and ones. The last digit in the string represents port one. The previous digits represent ports two, three, and so on. If the digit for a port is set to “1”, the port is enabled for the specific type of communication. If it is set to “0”, the system does not send or receive traffic that is controlled by a mask on the port.

To view the current port mask settings, run the **lssystem** command, as shown in Example 3-8. The output shows that all system ports allow all kinds of traffic.

Example 3-8 Viewing the local port mask

```
IBM_IBM FlashSystem:ITS0-FS9100:superuser>lssystem |grep mask
local_fc_port_mask 1111111111111111111111111111111111111111111111111111111111111111
partner_fc_port_mask 1111111111111111111111111111111111111111111111111111111111111111
```

To set the local or RPORT mask, run the **chsystem** command. Example 3-9 shows the mask setting for a system with four FC ports on each node and that has RC relationships. Masks are applied to allow local node-to-node traffic on ports 1 and 2, and replication traffic on ports 3 and 4.

Example 3-9 Setting a local port mask by running the chsystem command

```
IBM_IBM FlashSystem:ITS0-FS9100:superuser>chsystem -localfcportmask 0011
IBM_IBM FlashSystem:ITS0-FS9100:superuser>chsystem -partnerfcportmask 1100
IBM_IBM FlashSystem:ITS0-FS9100:superuser>lssystem |grep mask
local_fc_port_mask 0000000000000000000000000000000000000000000000000000000000000011
partner_fc_port_mask 0000000000000000000000000000000000000000000000000000000000001100
```

The mask is extended with zeros, and all ports that are not explicitly set in a mask have the selected type of traffic blocked.

Note: When replacing or upgrading your node hardware, consider that the number of FC ports and their arrangement might be changed. If so, make sure that any configured port masks are still valid for the new configuration.

3.4.6 Automatic configuration for IBM SAN Volume Controller back-end storage

If a system is supposed to work as FC-attached back-end storage for SVC, you can enable Automatic Configuration for Virtualization during the initial system setup or anytime later by selecting **Settings** → **System** → **Automatic Configuration**.

Automatic Configuration for Virtualization is intended for a new system. If there are host, pool, or volume objects that are configured, all the user data must be migrated out of the system, and those objects must be deleted.

The Automatic Configuration for Virtualization wizard starts immediately after you complete the initial setup wizard if you set **Automatic Configuration** to **On**. The following steps are performed by it:

1. Add control or expansion enclosures to the system that are not added yet. Click **Add Enclosure** to start the adding process, or click **Skip** to move to the next step. You can turn off the Automatic Configuration for Virtualization wizard at any step by clicking the ... (hamburger) symbol in the upper right, as shown in Figure 3-44.

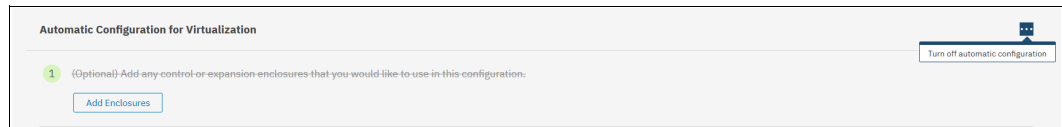


Figure 3-44 Automatic Configuration wizard: Add Enclosure

2. The wizard checks whether the SVC is correctly zoned to the system. By default, newly installed systems run in N_Port ID Virtualization (NPIV) mode (Target Port Mode). The system's virtual (host) WWPNs must be zoned for SVC. On the SVC side, physical WWPNs must be zoned to a back-end system independently of the NPIV mode setting.
3. Create a host cluster object for SVC. Each SVC node has its own worldwide node name (WWNN). Make sure to select all WWNNs that belong to nodes of the same SVC cluster.

Figure 3-45 shows that the system detects an SVC cluster with a single I/O group, so two WWNNs are selected.

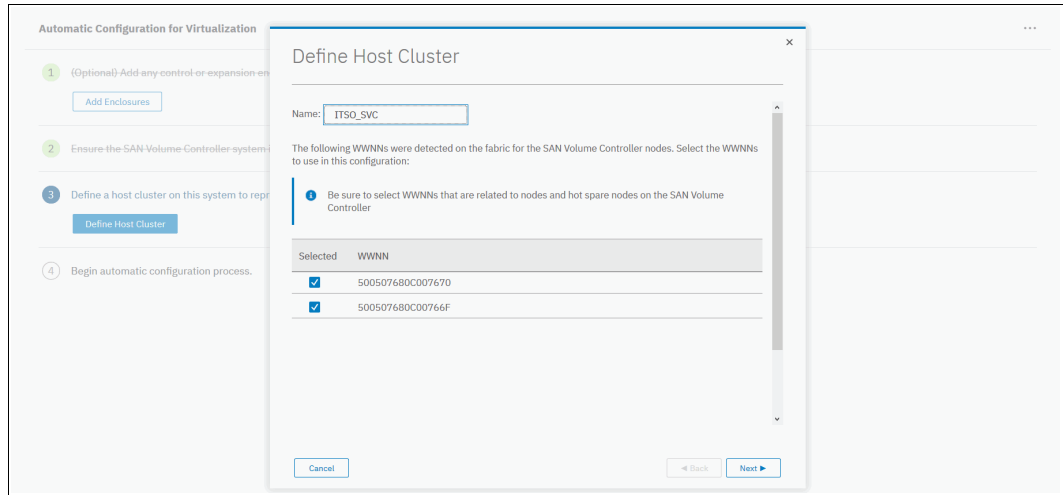


Figure 3-45 Defining a host cluster

- When all nodes of an SVC cluster including the spare one are selected, you can change the host object name for each one, as shown in Figure 3-46. For convenience, name the host objects to match the SVC node names or serial numbers.

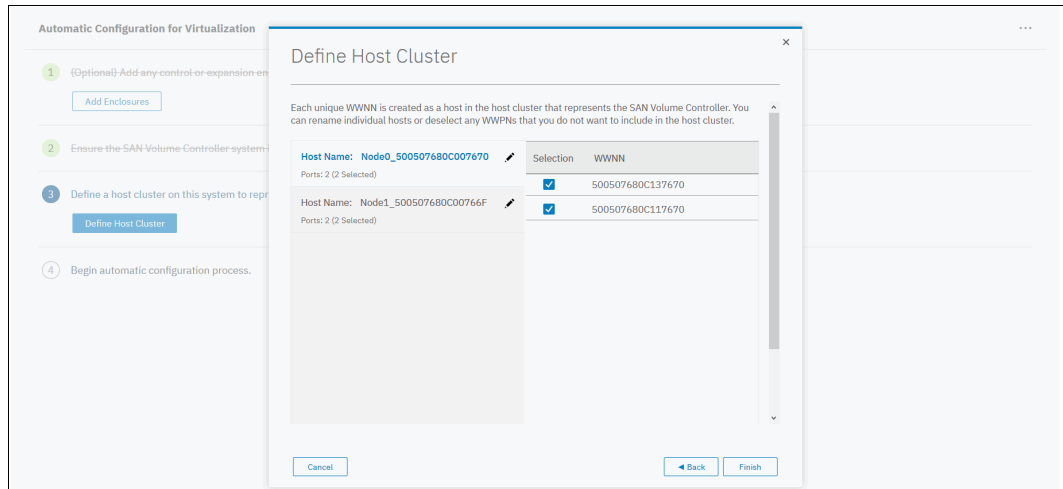


Figure 3-46 Hosts inside an IBM SAN Volume Controller host cluster

- Click **Automatic Configuration** and check the list of internal resources that are used. Click **Cancel** if the list is not correct; otherwise, click **Next**.

- If the system uses compressed drives (FCM drives), you are prompted to enter your expected compression ratio (or total capacity that will be provisioned to SVC), as shown in Figure 3-47. If SVC is using encryption or writes data that is not compressible, set the ratio to 1:1.

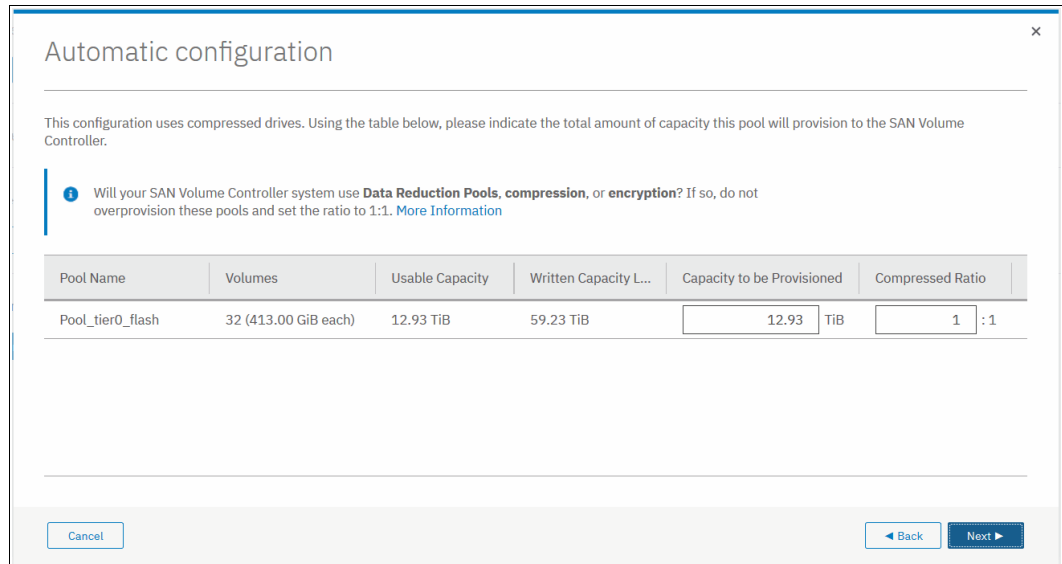


Figure 3-47 Automatic pool configuration

- Review the pool (or pools) configuration, as shown in Figure 3-48, and click **Proceed** to trigger commands that will apply it.

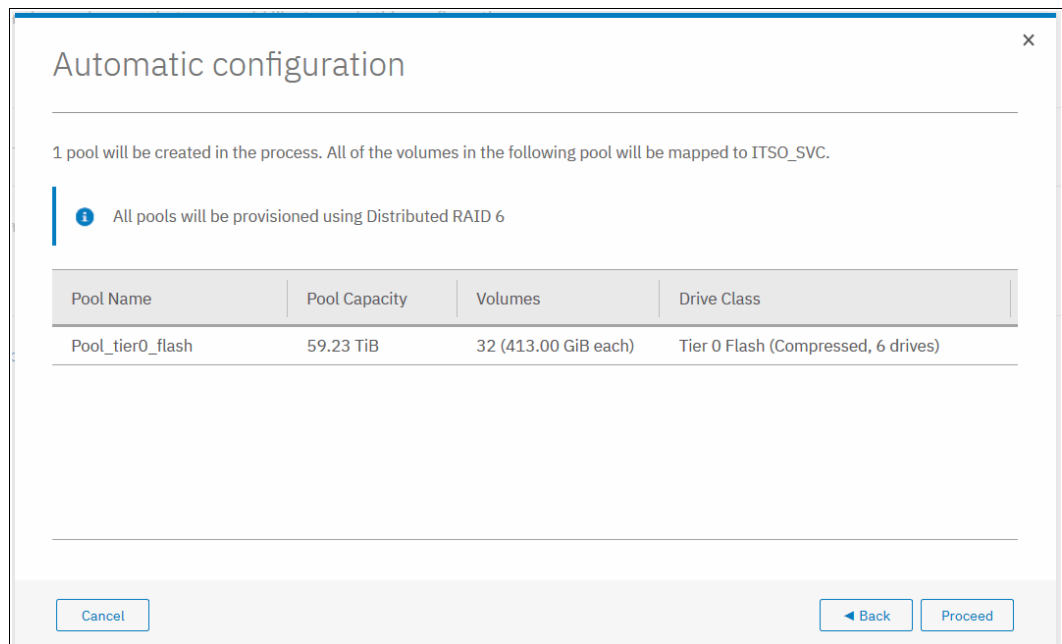


Figure 3-48 Pools configuration

- When the Automatic Configuration for Virtualization wizard completes, you see the window that is shown in Figure 3-49. After clicking **Close**, you may proceed to the SVC GUI and configure a new provisioned storage.

You can export the system volume configuration data in .csv format by using this window or anytime later by selecting **Settings** → **System** → **Automatic Configuration**.

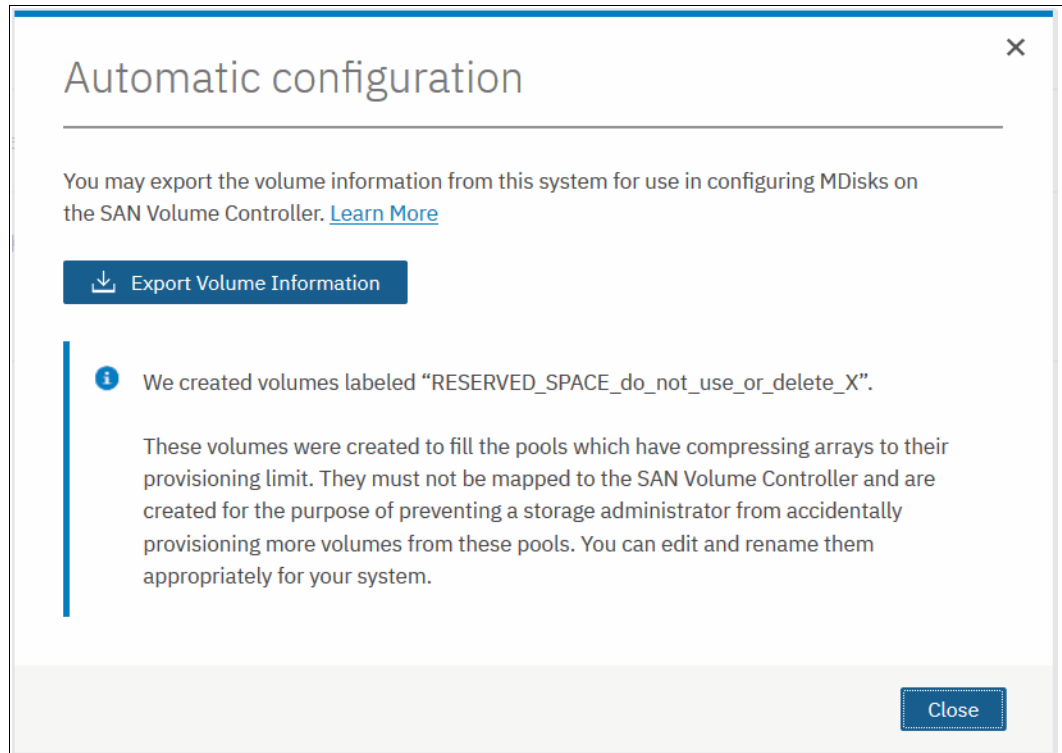


Figure 3-49 Automatic configuration complete

3.5 Configuring management access

The system can be managed by using the GUI and the CLI. Access to the system management interfaces require user authentication. User authentication and the secure communication implementation steps are described in this section.

3.5.1 Configuring secure communications

During system initialization, a *self-signed* Secure Sockets Layer (SSL) certificate is automatically generated by the system to encrypt communications between the browser and the system. Self-signed certificates generate web browser security warnings and might not comply with organizational security guidelines.

Signed SSL certificates are issued by a trusted CA. A browser maintains a list of trusted CAs that are identified by their *root* certificate. The root certificate must be included in this list in order for the signed certificate to be trusted.

To see the details of your system certificate, select **Settings** → **Security** and click **Secure Communications**, as shown in Figure 3-50, or run the `lssystemcert` command.

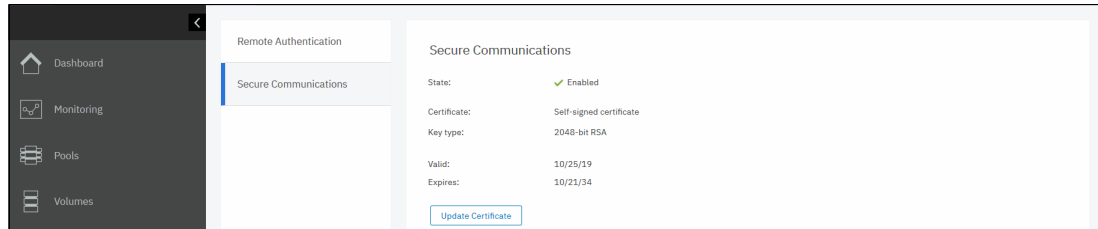


Figure 3-50 Accessing the Secure Communications window

Based on the security requirements for your system, you can create either a new self-signed certificate or install a signed certificate that is created by a third-party CA.

Generating a self-signed certificate

If a self-signed certificate is expired or its key type does not comply with your company's security policy, you can regenerate it. To renew a self-signed certificate, complete the following steps:

1. Select **Update Certificate** on the Secure Communications window, as shown in Figure 3-50.
2. Select **Self-signed certificate** and enter the details for the new certificate. The “Key type” and “Validity days” are the only mandatory fields.

Note: Before re-creating a self-signed certificate, ensure that your browser supports the type of keys that you are going to use for a certificate. See your organization's security policy to ensure what key type is required.

3. Click **Update**.

You are prompted to confirm the action. Click **Yes** to proceed. Close the browser, wait approximately 2 minutes, and reconnect to the management GUI.

To regenerate an SSL certificate by using a CLI, run the `chsystemcert` command, as shown in Example 3-10. Valid values for `-keytype` are `rsa2048`, `ecdsa384`, or `ecdsa521`.

Example 3-10 Regenerating a self-signed certificate

```
IBM_IBM FlashSystem:ITS0-FS9100:superuser>chsystemcert -mkselfsigned -keytype  
ecdsa521 -validity 365
```

Configuring a signed certificate

If your company's security policy requests certificates to be signed by a trusted authority, complete the following steps to configure a signed certificate:

1. Select **Update Certificate** in the Secure Communications window.
2. Select **Signed certificate** and enter the details for the new certificate signing request, as shown in Figure 3-51. All fields are mandatory except for the Subject Alternative Name. For the "Country" field, use a two-letter country code. Click **Generate Request**.

The screenshot shows a dialog box titled "Update Certificate" with a close button (X) in the top right corner. The dialog is divided into two main sections: "Certificate type" and "Certificate Signing Request".

Certificate type: There are two radio buttons: "Self-signed certificate" (unselected) and "Signed certificate" (selected).

Certificate Signing Request: This section contains several input fields and a button:

- Key type:** A dropdown menu showing "2048-bit RSA".
- Country:** A text input field containing "US".
- State:** A text input field containing "CA".
- City:** A text input field containing "San Jose".
- Organization:** A text input field containing "IBM".
- Organization unit:** A text input field containing "ITSO".
- Common name:** A text input field containing "9.155.123.198".
- Subject Alternative Name:** A text area containing the text "Use the suggested format below:" followed by "IP:123.45.67.91", "URI:http://www.mydomain.com", "DNS:cluster.mydomain.com", and "email:support@mydomain.com".
- Email address:** A text input field containing "redbooks@us.ibm.com".
- Generate Request:** A blue button.

Signed Certificate: This section is partially visible at the bottom, showing a "Signed certificate:" label and a dropdown menu with "Upload from Certificate Authority" selected.

At the bottom of the dialog, there are two buttons: "Cancel" and "Update".

Figure 3-51 Generating a certificate request

3. When prompted, save the `certificate.csr` file that contains the certificate signing request.

Until the signed certificate is installed, the Secure Communications window shows that an outstanding certificate request exists.

Attention: If you must update a field in the certificate request, generate a new request and submit it for signing by the proper CA. However, this process invalidates the previous certificate request and prevents the installation of the signed certificate that is associated with the original request.

4. Submit the request to the CA to receive a signed certificate. Notify the CA that you need a certificate (or certificate chain) in base64-encoded Privacy Enhanced Mail (PEM) format.
5. When you receive the signed certificate, select **Update Certificate** in the Secure Communications window again.
6. Select **Signed Certificate** and click the folder icon next to the **Signed Certificate** input field of the Update Certificate window, as shown in Figure 3-51 on page 141. Click **Update**.
7. You are prompted to confirm the action. Click **Yes** to proceed. After your certificate is installed, the GUI session disconnects. Close the browser window and wait approximately 2 minutes before reconnecting to the management GUI.
8. Reconnect to the GUI and select **Settings** → **Security** → **Secure Communications**. The window that opens should show that you are using a signed certificate, as shown in Figure 3-52.

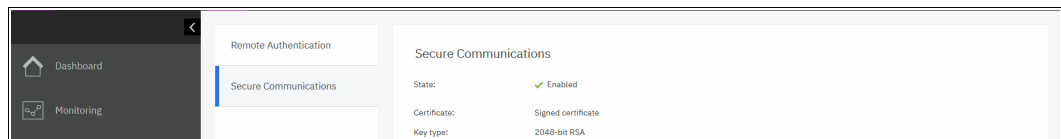


Figure 3-52 Signed certificate installed

3.5.2 Configuring password policies

There are set of options that were added to IBM Spectrum Virtualize V8.4.0 or later that allow a security admin to create policies for passwords, account lockout, and session timeout. A single system-wide policy applies to all local accounts (session timeouts also apply to remote accounts).

Password creation

The password creation options can be customized to employ the following policies:

- ▶ Minimum password length (6 – 64 characters)
- ▶ Minimum number of uppercase characters (1 – 3)
- ▶ Minimum number of lowercase characters (1 – 3)
- ▶ Minimum number of special characters (1 – 3)
- ▶ Minimum number of digits (1 – 3)

Note: A new policy does not apply retroactively to existing passwords. However, any new passwords must meet the current policy setting.

Password creation rules ensure that passwords that were used before do not match the new password:

- ▶ Password History checking can be enabled. Zero – 10 previous passwords can be checked.
- ▶ Stores the previous password hashes only (no plain text).
 - 0 = compare the current password only.
 - 10 = check that the new password does not match the current password or the 10 passwords that were used before the current password.
- ▶ The minimum required password age can be set (0 – 365 days).

A minimum age of 1 means that a user can change a password only once per day, which prevents a user from cycling through previous passwords.

Note: The password history is not checked when a security admin changes another user’s password because this function is not supported on IBM FlashSystem 5010.

From the GUI, set the password creation options and password creation rules policies by selecting **Settings** → **Security** and clicking **Password Policies**, as shown in Figure 3-53.

Remote Authentication

Encryption

Password Policies

Secure Communications

Inactivity Logout

Password creation

Define policies for when users are creating or changing a password.

Passwords must contain a certain number of characters.

Minimum required characters:

8

Require passwords to contain specific characters.

Minimum required lowercase letters: 0

Minimum required uppercase letters: 0

Minimum required special characters: 0

Minimum required numbers: 0

Prevent users from reusing previous passwords.

Number of previous passwords that users cannot reuse: 6

Number of days a user must wait before they can change their password again: 1

Reset Save

Figure 3-53 Creating password policies for password creation

Password expiration and account lockout

The following options can be used to apply password expiry to passwords:

- ▶ Passwords can be set to expire after 0 – 365 days.
- ▶ All existing passwords are set to expire in X days when the setting is first enabled.
- ▶ A user with an expired password can log in to the system, but cannot run any **svctask** commands until they change their password.
- ▶ An expiry warning can be enabled (0 – 30 days) which warns the user on login that their password expires in X days. (Only on the CLI on IBM Spectrum Virtualize V8.4)

The security admin can force a user to change their password at any time. The password expires immediately. If you use the CLI, you can expire individual users. If you use the GUI, you can reset all user passwords.

- ▶ Can be used when creating a user to require a password change on first login.
- ▶ Can be used after changing password policy settings.

There are two ways to force account locking:

- ▶ Account locking (manual):

The security admin can manually lock and unlock user accounts by using the CLI, as shown in Example 3-11.

Example 3-11 Manually lock and unlock a user account

```
IBM FlashSystem 7200:admin>svctask chuser -lock bill
IBM FlashSystem 7200:admin>svctask chuser -unlock ted
```

Note: A locked account is not allowed to log in to the system.

- ▶ Account locking (automatic):

Accounts can also be locked automatically as follows:

- By setting the maximum number of failed login attempts (0 – 10).

Note: The counter is reset on a successful login.

- By setting the length of time a user is locked out of the system (0 – 10080 minutes (which is 7 days). 0 = indefinite).

Disabling the superuser account and session timeouts is available only on platforms with a dedicated techport.

Note: This feature is not available on IBM FlashSystem 5010(E) and 5030(E).

Disabling the superuser account can be done either from the GUI or CLI by completing the following steps:

1. Use an explicit option to enable superuser locking, as shown in Example 3-12.

Example 3-12 Manually enable superuser account locking option

```
IBM FlashSystem 7200:admin>svctask chsecurity -superuserlocking enable
Changing the system security settings could result in a loss of access to the
system via SSH or the management GUI. Refer to the Command Line Interface help
```



```
for more information about the risks associated with each parameter. Are you
sure you wish to continue? (y/yes to confirm) yes
IBM FlashSystem 7200:admin>
```

2. The superuser can then be locked as shown in Example 3-13.

Example 3-13 Manually lock the superuser account

```
IBM FlashSystem 7200:admin>svctask chuser -lock superuser
IBM FlashSystem 7200:admin>
```

A good use case is assuming that some enterprises have policies that all systems should use remote authentication. So, configure remote authentication, create a remote security admin, and disable the superuser. Now, no local accounts can log in to the system.

Note: The superuser account is still required for **satask** actions and recovery actions, for example, T3/T4 recovery. It is automatically unlocked for recovery and must be manually relocked afterward.

Session timeouts can be configured for both CLI and GUI sessions as follows:

- ▶ A configurable CLI timeout of 5 – 240 minutes
- ▶ A separate configurable GUI timeout of 5 – 240 minutes

All the above options governing password expiry, requiring a password change or account lockouts, and disabling the superuser account and session timeouts can be done from the GUI by selecting **Settings** → **Security** and clicking **Password Policies**, as shown in Figure 3-54.

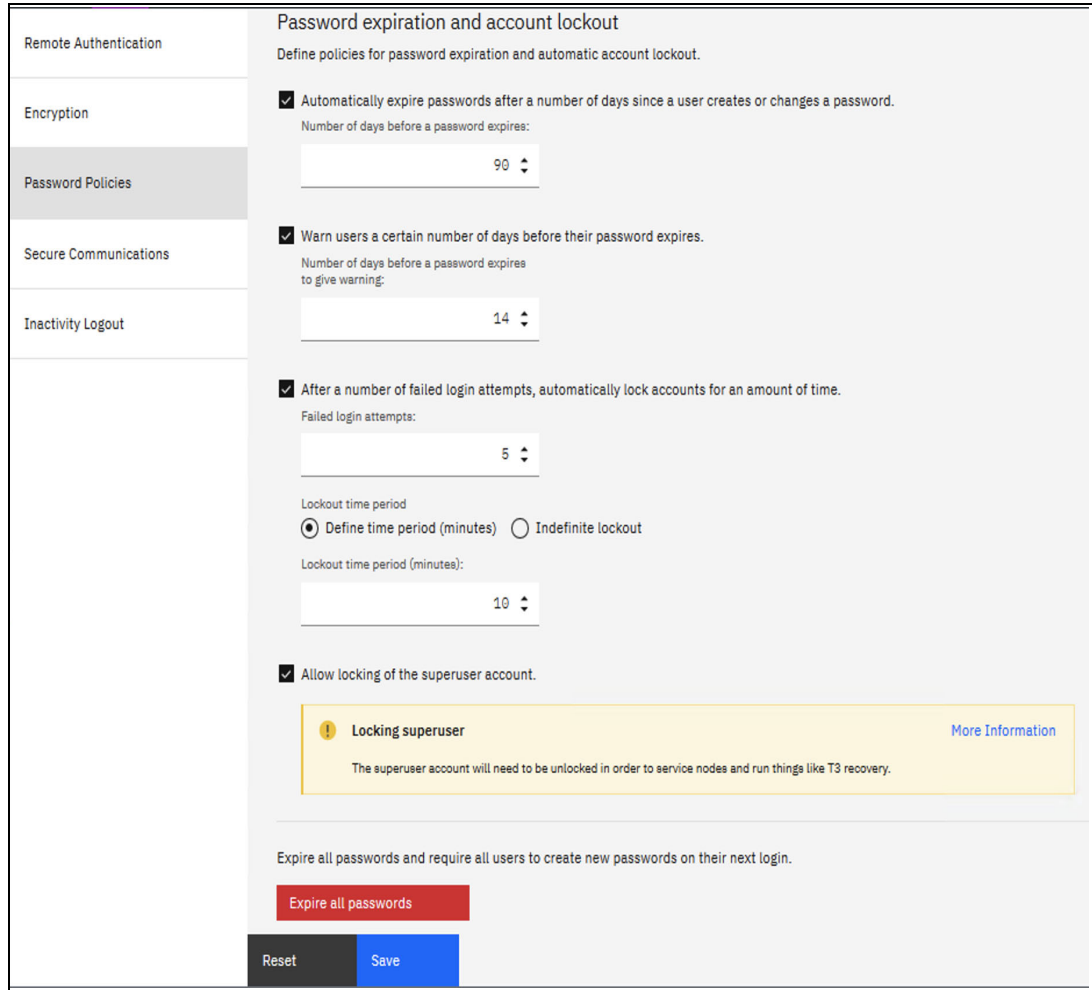


Figure 3-54 Creating password policies for password expiration and account lockout

3.5.3 Configuring user authentication

There are two methods of user authentication to control access to the GUI and to the CLI:

- ▶ *Local user authentication* is performed within the system. GUI users authenticate with user name and password. CLI users must provide a user name and a Secure Shell (SSH) public key or a password.
- ▶ *Remote user authentication* allows users to authenticate to the system by using credentials that are stored on an external authentication service. With this feature, you use user credentials and user groups that are defined on the remote service to simplify user management and access, enforce password policies more efficiently, and separate user management from storage management.

Locally administered users can coexist with remote authentication.

User roles and groups

User groups are used to determine what tasks the user is authorized to perform. Each user group is associated with a single role. Roles apply to both local and remote users on the system and are based on the user group to which the user belongs. A local user can belong only to a single group, so the role of a local user is defined by the single group to which that user belongs.

For a list of user roles and their tasks, and a description of a pre-configured user group, see [IBM Documentation](#) and expand **Product overview** → **Technical overview** → **User roles**.

Superuser account

Every system has a default user that is called the *superuser*. It cannot be deleted or modified, except for changing the password and SSH key. The superuser is a *local* user and cannot be authenticated remotely. The superuser has a *SecurityAdmin* user role, which has the most privileges within the system.

Note: The superuser is the only user that may log in to the Service Assistant interface. It is also the only user that may run **sa info** and **sa task** commands through the CLI.

The password for superuser is set during the system setup. The superuser password can be reset to its default value of `passwd` by using a procedure that is described in [IBM Documentation](#) by expanding **Troubleshooting** → **Resolving a problem** → **Procedure: Resetting the superuser password by using the management GUI or CLI**.

Note: The superuser password reset procedure uses system internal USB ports. The system may be configured to disable those ports. If the USB ports are disabled and there are no users with the *SecurityAdmin* role and a known password, the superuser password cannot be reset without replacing the system hardware and deleting the system configuration.

Local authentication

A *local user* is a user whose account is managed entirely on the system. A local user belongs to one user group only, and it must have a password, an SSH public key, or both. Each user has a username, which must be unique across all users in one system.

Username can contain up to 256 printable American Standard Code for Information Interchange (ASCII) characters. Forbidden characters are the single quotation mark ('), colon (:), percent symbol (%), asterisk (*), comma (,), and double quotation marks ("). A username cannot begin or end with a blank space.

Passwords for local users can be up to 64 printable ASCII characters, but cannot begin or end with a space.

When connecting to the CLI, encryption key authentication is attempted first with the username and password combination available as a fallback. The SSH key authentication method is available for CLI and file transfer access only. For GUI access, only the password is used.

To add a user that is authenticated without a password by using only an SSH key, select **Access** → **Users by Group**, click **Add user**, and then click **Browse** to select the SSH public key for that user, as shown in Figure 3-55. The Password field may be left blank. The system accepts public keys that are generated by PuTTY (SSH2), OpenSSH, and Request for Comments (RFC) 4716-compliant keys that are generated by other clients.

The screenshot shows a 'Create User' dialog box with the following fields and options:

- Name:** SCuser
- Authentication Mode:** Local (selected), Remote
- User Group:** Monitor
- Local Credentials:**
 - Local users must have a password, an SSH public key, or both.
 - Password: [Empty field]
 - Verify password: [Empty field]
 - SSH Public Key: Browse... publickey.pub
- Buttons:** Cancel, Create

Figure 3-55 Creating a user that is authenticated by an SSH key

If local authentication is used, user accounts must be created for each system. If you want access for a user on multiple systems, you must define the user in each system.

Remote authentication

A *remote user* is authenticated by using identity information that is accessible by using the Lightweight Directory Access Protocol (LDAP). The LDAP server must be available for the users to log in to the system. Remote users have their groups defined by the remote authentication service.

Users that are authenticated by an LDAP server can log in to the management GUI and the CLI. These users do not need to be configured locally for CLI access, and they do not need an SSH key that is configured to log in by using the CLI.

If multiple LDAP servers are available, you can configure more than one LDAP server to improve resiliency. Authentication requests are processed by those LDAP servers that are marked as preferred unless the connection fails or a user is not found. Requests are distributed across all preferred servers for load balancing in a round-robin fashion.

Note: All LDAP servers that are configured within the same system must be of the same type.

If users that are part of a group on the LDAP server are to be authenticated remotely, a user group with an identical name must exist on the system. The user group name is *case-sensitive*. The user group must also be enabled for remote authentication on the system.

A user who is authenticated remotely is granted permissions according to the role that is assigned to the user group of which the user is a member.

To configure remote authentication by using LDAP, start by enabling remote authentication by completing the following steps:

1. Select **Settings** → **Security**, click **Remote Authentication**, and then click **Configure Remote Authentication**, as shown in Figure 3-56.

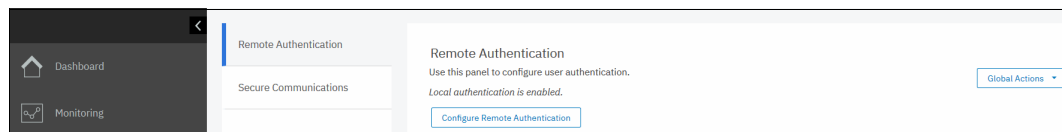


Figure 3-56 Configuring remote authentication

2. Enter the LDAP settings. These settings are not server-specific. They are applied to all LDAP servers that are configured in the system. Extra optional settings are available by clicking **Advanced Settings**, as shown in Figure 3-57.

Figure 3-57 Configure Remote Authentication settings

The following settings are available:

- LDAP type:
 - **IBM Security Directory Server** (for IBM Security Directory Server).
 - **Microsoft Active Directory** (AD).
 - **Other** (other LDAP v3-capable directory servers, for example, OpenLDAP).
 - Security:
 - **LDAP with StartTLS**: Select this option to use the StartTLS extension (RFC 2830). It works by establishing a non-encrypted connection with an LDAP server on a standard LDAP port (389), and then performing a TLS handshake over an existing connection.
 - **LDAPS**: Select to use LDAP over SSL and establish secure connections by using port 636.
 - **None**: Select to transport data in clear text format without encryption.
 - Service Credentials: Sets a username and password for administrative binding (the credentials of a user that has the authority to query the LDAP directory). Leave it empty if your LDAP server is configured to support anonymous bind.
- For AD, a username must be in User Principal Name (UPN) format.

- Advanced settings:

Speak to the administrator of the LDAP server to ensure that these fields are completed correctly:

- **User Attribute**

This LDAP attribute is used to determine the username of remote users. The attribute must exist in your LDAP schema and must be unique for each of your users.

This advanced setting defaults to `sAMAccountName` for AD and to `uid` for **IBM Security Directory Server** and **Other**.

- **Group Attribute**

This LDAP attribute is used to determine the user group memberships of remote users. The attribute must contain either the distinguished name of a group or a colon-separated list of group names.

This advanced setting defaults to `memberOf` for AD and **Other**, and to `ibm-allGroups` for **IBM Security Directory Server**. For **Other** LDAP type implementations, you might need to configure the `memberOf` overlay if it is not in place.

- **Audit Log Attribute**

This LDAP is an attribute that is used to determine the identity of remote users. When an LDAP user performs an audited action, this identity is recorded in the audit log. This advanced setting defaults to `userPrincipalName` for AD and to `uid` for **IBM Security Directory Server** and the **Other** type.

3. Enter the server settings for one or more LDAP servers, as shown in Figure 3-58.

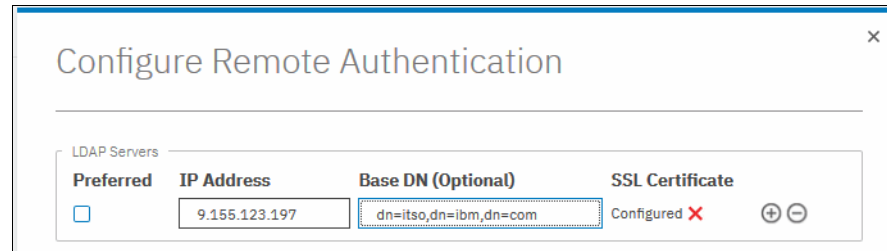


Figure 3-58 Configure Remote Authentication: Creating an LDAP server

The following settings are available:

- **Preferred**

One or more configured LDAP servers may be marked as **Preferred**. Requests are distributed among these servers, and use only non-preferred servers if all the preferred servers failed.

- **IP Address**

The IP address of the server.

- **Base DN**

The distinguished name to use as a starting point for searching for users on the server (for example, `dc=itso, dc=ibm, dc=com`).

- **SSL Certificate**

The SSL certificate that is used to securely connect to the LDAP server. This certificate is required only if you chose to use SSL or Transport Layer Security as a security method earlier.

Click **Browse** to select a server certificate. The system accepts certificates in base-64 encoded PEM format. To get a certificate in PEM format from your AD server, select **Base-64 Encoded X.509 (.CER)** in the MS Windows Certificate Export wizard, as shown in Figure 3-59.

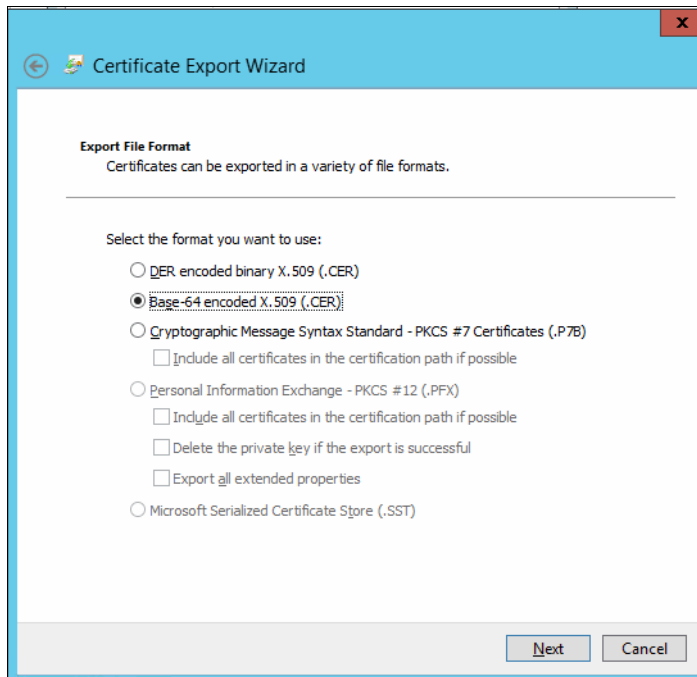


Figure 3-59 Exporting the AD certificate

Note: If your organization is using a tiered CA hierarchy, a server certificate that is exported for use on a system must include all the certificates in a chain. To accomplish this task, export the certificate in MS Windows in .P7B format and use third-party tools (OpenSSL) to convert it to PEM format. Otherwise, the exported certificate will not contain all certificates in the certification path.

If you set a certificate and you want to remove it, click the red cross next to **Configured**.

- Click the plus and minus signs to add or remove LDAP server records. You may define up to six servers.

Click **Finish** to save the settings.

4. To verify that LDAP is enabled, select **Settings** → **Security** → **Remote Authentication**, as shown in Figure 3-60. You may also test the server connection by selecting **Global Actions** → **Test LDAP connections** and verifying that all servers return “CMMVC7075! The LDAP task completed successfully”.

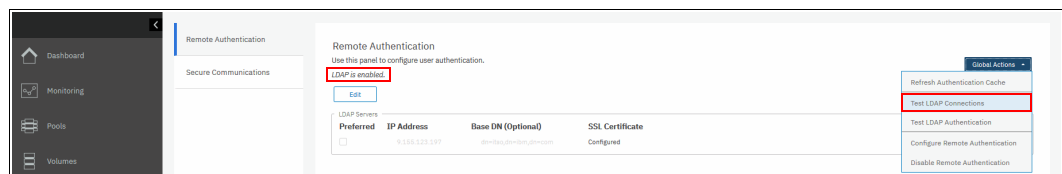


Figure 3-60 Verifying that LDAP is enabled

You can use the **Global Actions** menu to disable remote authentication and switch to local authentication only.

After remote authentication is enabled, the remote user groups must be configured. You can use the default built-in user groups for remote authentication. However, the name of the default user groups cannot be changed. If the LDAP server contains a group that you want to use and you do not want to create this group on the storage system, the name of the group must be changed on the server side to match the default name. Any user group, whether default or self-defined, must be enabled for remote authentication before LDAP authentication can be used for that group.

To create a user group with remote authentication enabled, complete the following steps:

1. Select **Access** → **Users by Group** and click **Create User Group**. Enter the name for the new group, select the **LDAP** checkbox, and choose a role for the users in the group, as shown in Figure 3-61.

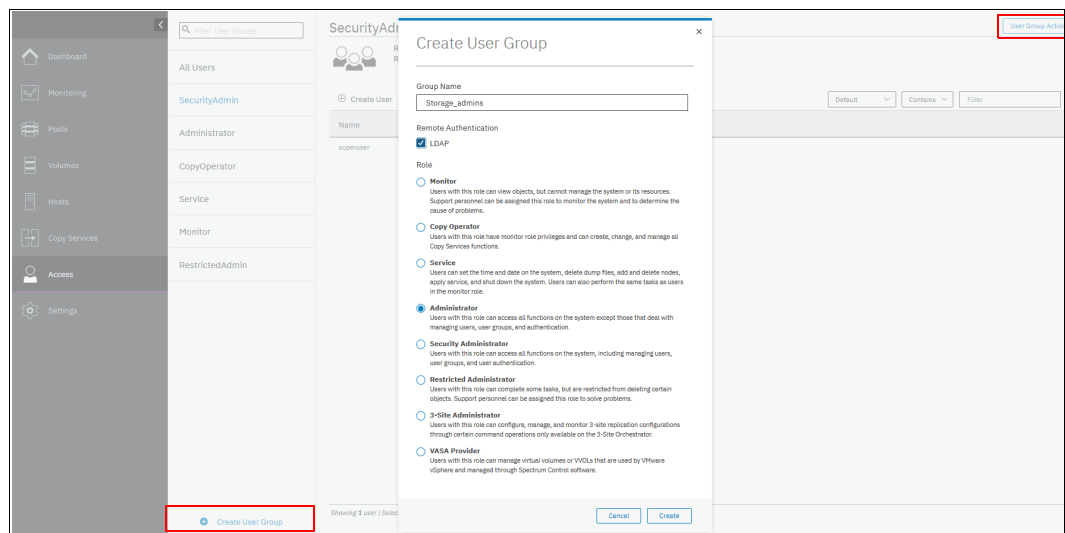


Figure 3-61 Creating a user group with remote authentication enabled

2. Enable LDAP for one of the existing groups, select it in the list, select **User Group Actions** → **Properties** in the upper right, and select the LDAP checkbox.
3. When you have at least one user group that is enabled for remote authentication, verify that you set up your user group on the LDAP server correctly by checking whether the following conditions are true:
 - The name of the user group on the LDAP server matches the one that you modified or created on the storage system.

Note: The user group name is case-sensitive.

- Each user that you want to authenticate remotely is a member of the LDAP user group that is configured for the system role.

4. To test the user authentication, select **Settings** → **Security** → **Remote Authentication**, and then select **Global Actions** → **Test LDAP Authentication** (for an example, see Figure 3-60 on page 152). Enter the user credentials of a user that is defined on the LDAP server and click **Test**. A successful test returns the message “CMMVC70751 The LDAP task completed successfully”.

A user can log in with their short name (that is, without the domain component) or with the fully qualified username in the UPN format (user@domain).



IBM Spectrum Virtualize GUI

This chapter describes an overview of the IBM Spectrum Virtualize GUI. The management GUI is a tool that is enabled and provided by IBM Spectrum Virtualize that helps you to monitor, manage, and configure your system.

This chapter explains the basic view and the configuration procedures that are required to get your system environment running as quickly as possible by using the GUI. This chapter does not describe advanced troubleshooting or problem determination and some of the complex operations (compression and encryption). For more information, see Chapter 13, “Reliability, availability, and serviceability, monitoring and logging, and troubleshooting” on page 793.

Throughout this chapter, all GUI menu items are introduced in a systematic, logical order as they appear in the GUI. However, topics that are described more in detail in other chapters of the book are only referred to here. For example, Storage pools (Chapter 5, “Storage pools” on page 237), Volumes (Chapter 6, “Volumes” on page 299), Hosts (Chapter 7, “Hosts” on page 405), and Copy Services (Chapter 10, “Advanced Copy Services” on page 553) are described in separate chapters.

Demonstration: The IBM Client Demonstration Center includes a demonstration of the Version 8.4.0 GUI. For more information, see the [IBM Client Demonstration Center](#) (log in required).

This chapter includes the following topics:

- ▶ 4.1, “Performing operations by using the GUI” on page 156
- ▶ 4.2, “GUI introduction” on page 160
- ▶ 4.3, “System Hardware - Overview window” on page 166
- ▶ 4.4, “Monitoring menu” on page 170
- ▶ 4.5, “Pools” on page 180
- ▶ 4.6, “Volumes” on page 180
- ▶ 4.7, “Hosts” on page 181
- ▶ 4.8, “Copy Services” on page 182
- ▶ 4.9, “Access” on page 182
- ▶ 4.10, “Settings” on page 196
- ▶ 4.11, “Additional frequent tasks in the GUI” on page 230

4.1 Performing operations by using the GUI

This section describes useful tasks that use the GUI that help administrators to manage, monitor, and configure the system as quickly as possible. For the example in this book, we configure the system in a standard topology.

The GUI is a built-in software component within the IBM Spectrum Virtualize Software. Multiple users can be logged in to the GUI. However, no locking mechanism exists, so be aware that if two users change the same object simultaneously, the last action that is entered from the GUI is the action that takes effect.

Important: Data entries that are made through the GUI are case-sensitive.

You must enable Java Script in your browser. For Mozilla Firefox, JavaScript is enabled by default and requires no other configuration steps.

4.1.1 Accessing the GUI

To access the IBM GUI, enter the IP address that was set during the initial setup process into your web browser. You can connect from any workstation that can communicate with the system. The login window opens (see Figure 4-1).

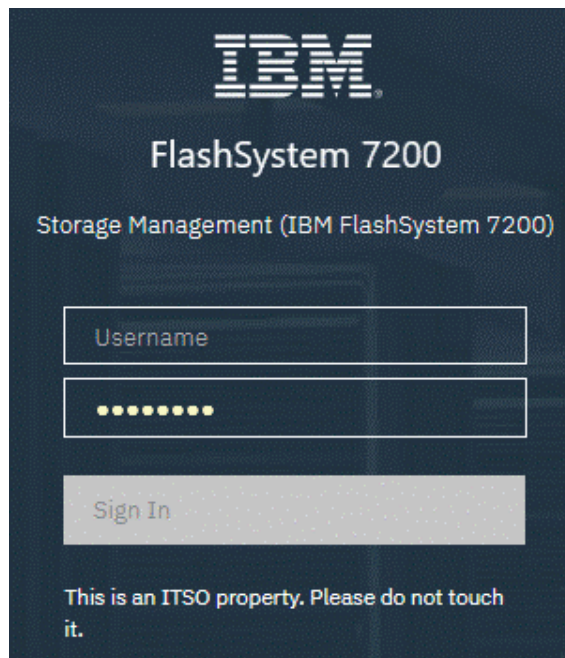


Figure 4-1 Login window of the GUI

Note: If you log in to the GUI by using the configuration node, you receive another option: Service Assistant Tool (SAT). Clicking this option takes you to the service assistant instead of the cluster GUI, as shown in Figure 4-2.

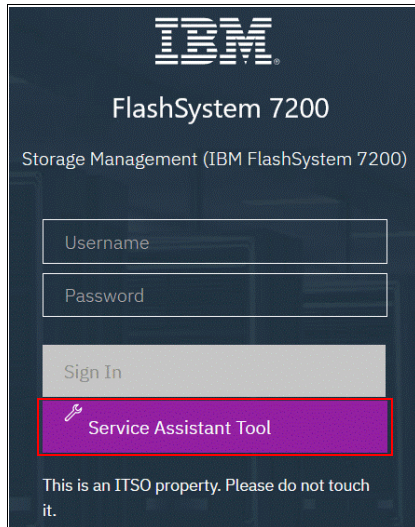


Figure 4-2 Login window of the storage system when it is connected to the configuration node

It is a best practice for each user to have their own unique account. The default user accounts should be disabled for use or their passwords changed and kept secured for emergency purposes only. This approach helps to identify personnel working on the systems and track all important changes that are done by them. The *superuser* account should be used for initial configuration only.

After a successful login, the Version 8.4 Welcome window opens and displays the system dashboard (see Figure 4-3).

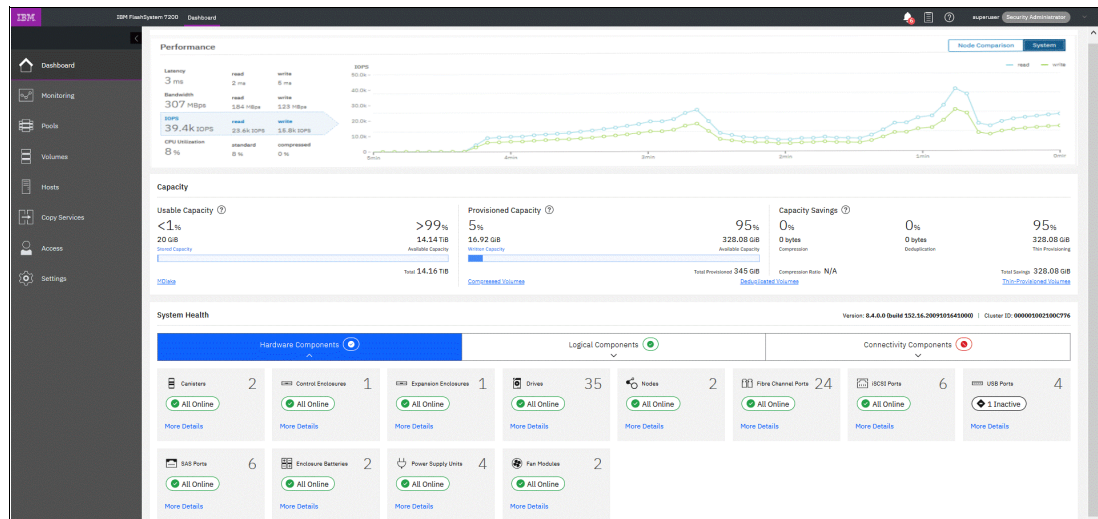


Figure 4-3 Welcome page with the system dashboard

The Dashboard is divided into three sections:

► Performance

This section provides important information about latency, bandwidth, input/output operations per second (IOPS), and CPU utilization. All this information can be viewed at the system or canister levels. A “Node comparison” view shows the differences in characteristics of each node (see Figure 4-4). The performance graph is updated with new data every 5 seconds.

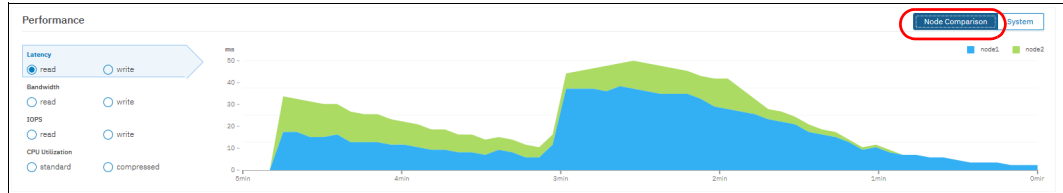


Figure 4-4 Performance statistics

► Capacity

This section (Figure 4-5) shows the current utilization of attached storage and its usage. Apart from the usable capacity, it also shows provisioned capacity and capacity savings. You can select the **Compressed Volumes**, **Deduplicated Volumes**, or **Thin Provisioned Volumes** options to display a complete list of the options in the Volumes tab.

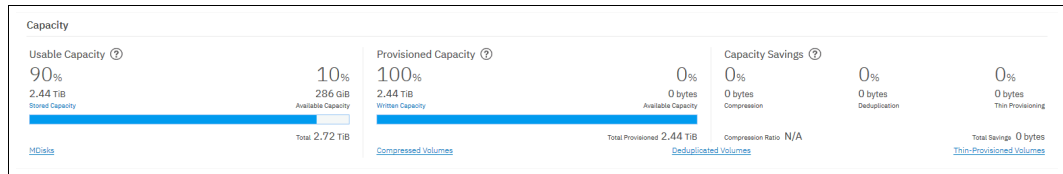


Figure 4-5 Capacity overview

If the ‘Overprovisioned External Systems’ section appears, you can then click it to see a list of related managed disks (MDisks) and pools, as shown in Figure 4-6.

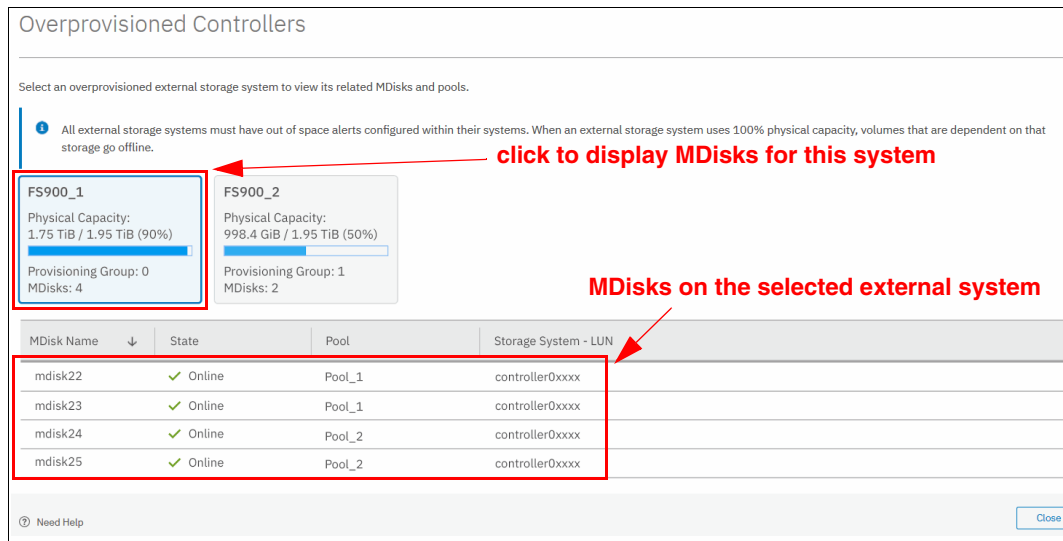


Figure 4-6 List that shows overprovisioned external storage

You also see a warning when assigning MDisks to pools if the MDisk is on an overprovisioned external storage controller.

► **System Health**

This section indicates the status of all critical system components, which are grouped in three categories: Hardware, logical, and connectivity components, as shown in Figure 4-7. When you click **Expand**, each component is listed as a subgroup. You can then go directly to the section of GUI where the component that you are interested in is managed.

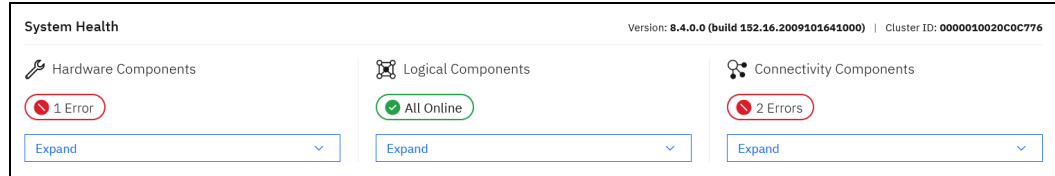


Figure 4-7 System Health overview window

Figure 4-8 shows the expanded view.

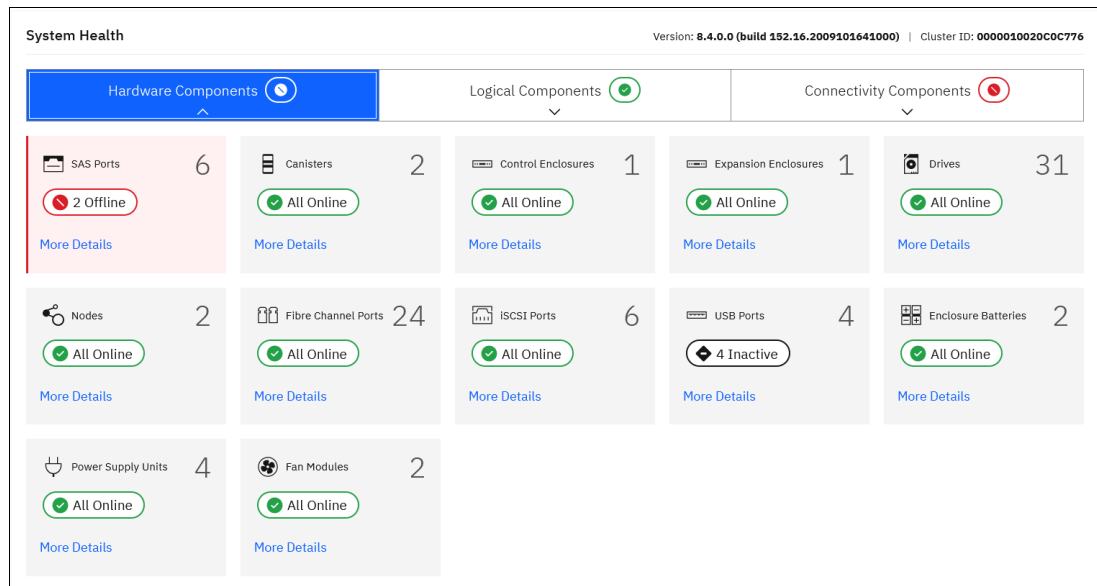


Figure 4-8 Expanded System health view

The dashboard in Version 8.4 displays as a welcome page instead of the system window as in previous versions. This system overview was moved to the **Monitoring** → **System Hardware** menu.

Although the Dashboard window provides key information about system behavior, the System menu is a preferred starting point to obtain the necessary details about your system components.

4.2 GUI introduction

As shown in Figure 4-9, the former GUI System window was moved to **Monitoring** → **System Hardware**.

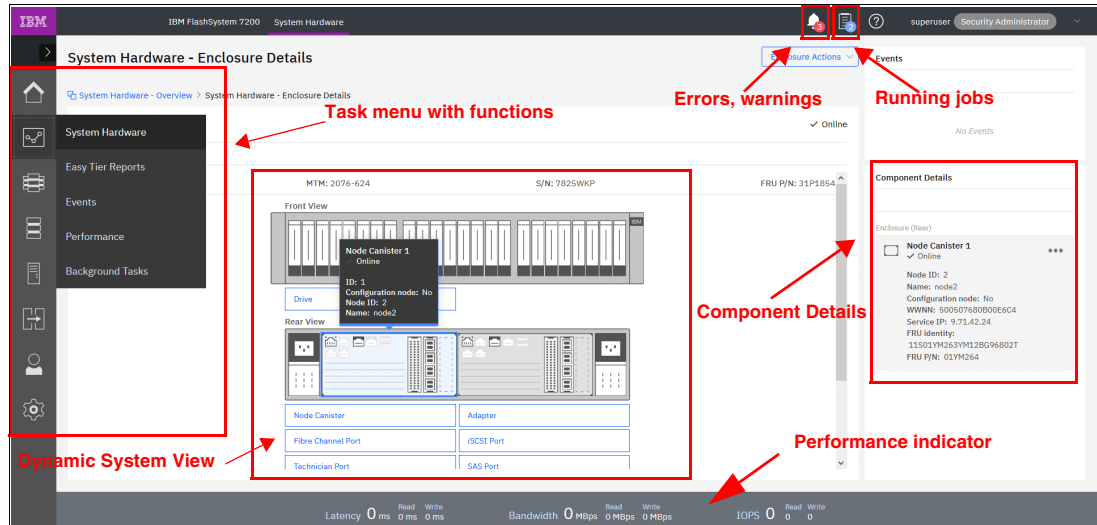


Figure 4-9 IBM Storage System Hardware window

4.2.1 Task menu

The IBM Spectrum Virtualize GUI task menu is always available on the left side of the GUI window. To browse by using this menu, click the action and choose a task that you want to display, as shown in Figure 4-10.

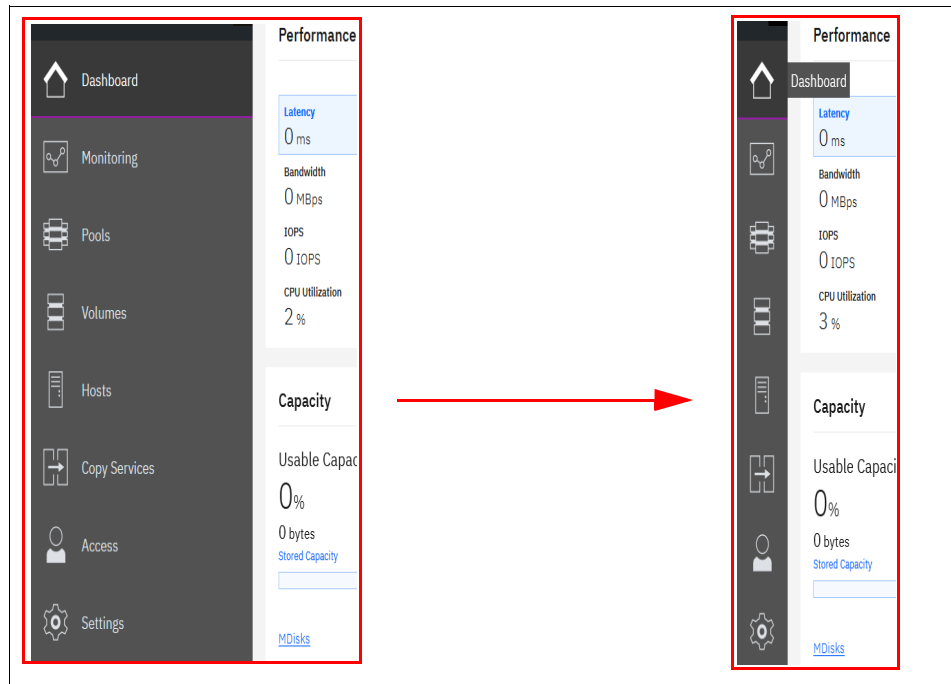


Figure 4-10 The task menu on the left side of the GUI

By reducing the horizontal size of your browser window, the wide task menu shrinks to the icons only.

4.2.2 Suggested tasks

After the initial configuration process is complete, IBM Spectrum Virtualize shows the information about suggested tasks to notify the administrator that several key functions are not yet configured. If necessary, this indicator can be closed and these tasks can be performed at any later time.

Figure 4-11 shows the suggested tasks in the System window.

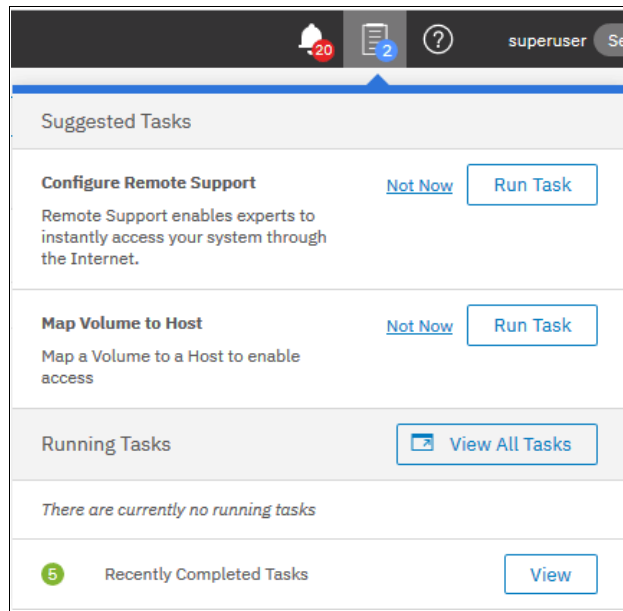


Figure 4-11 Suggested tasks

In this case, the GUI has two suggested tasks that help with the general administration of the system: You can directly perform the tasks from this window, or cancel them and run the procedure later. Other suggested tasks that typically appear after the initial system configuration are to create a volume and configure a storage pool.

The dynamic IBM Spectrum Virtualize menu contains the following windows:

- ▶ Dashboard
- ▶ Monitoring
- ▶ Pools
- ▶ Volumes
- ▶ Hosts
- ▶ Copy Services
- ▶ Access
- ▶ Settings

4.2.3 Notification icons and help

Three notification icons are in the upper navigation area of the GUI (see Figure 4-12). The left icon indicates warning and error alerts that were recorded in the event log. The middle icon shows running jobs and suggested tasks. The third rightmost icon offers a help menu with content that is associated with the current tasks and the currently opened GUI menu.



Figure 4-12 Notification area

Alerts indication

The left icon in the notification area informs administrators about important alerts in the systems. Click the icon to list warning messages in yellow and errors in red (see Figure 4-13).

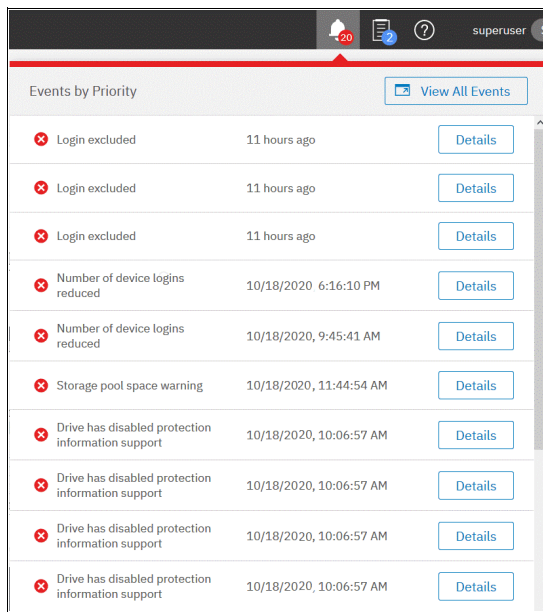


Figure 4-13 System alerts

You can go directly to the Events menu by clicking the **View All Events** option, as shown in Figure 4-14.

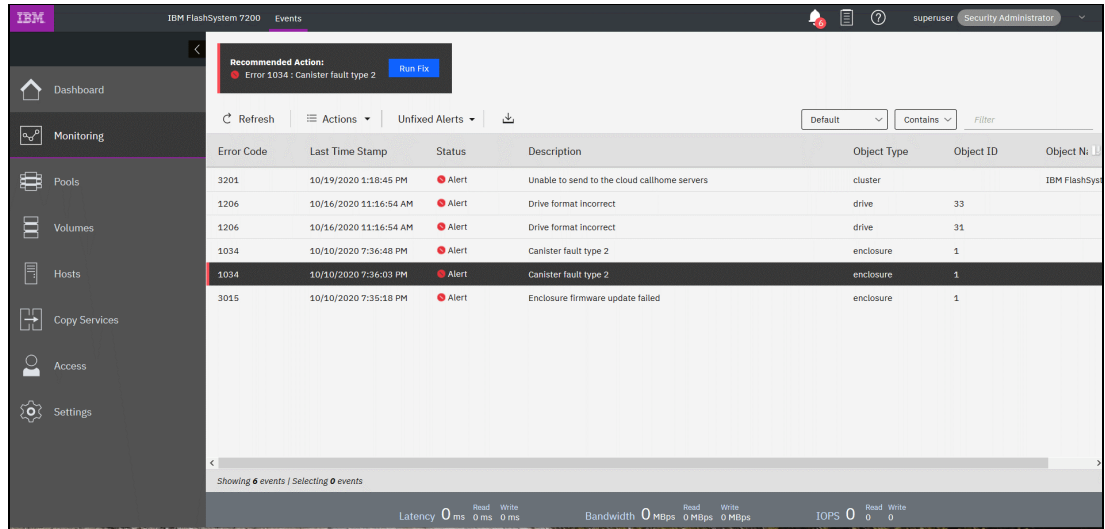


Figure 4-14 View all Events

You can see each event message separately by clicking the **Details** icon of the specific message. Then, you can analyze the content and eventually run the suggested fix procedure, as shown in Figure 4-15.

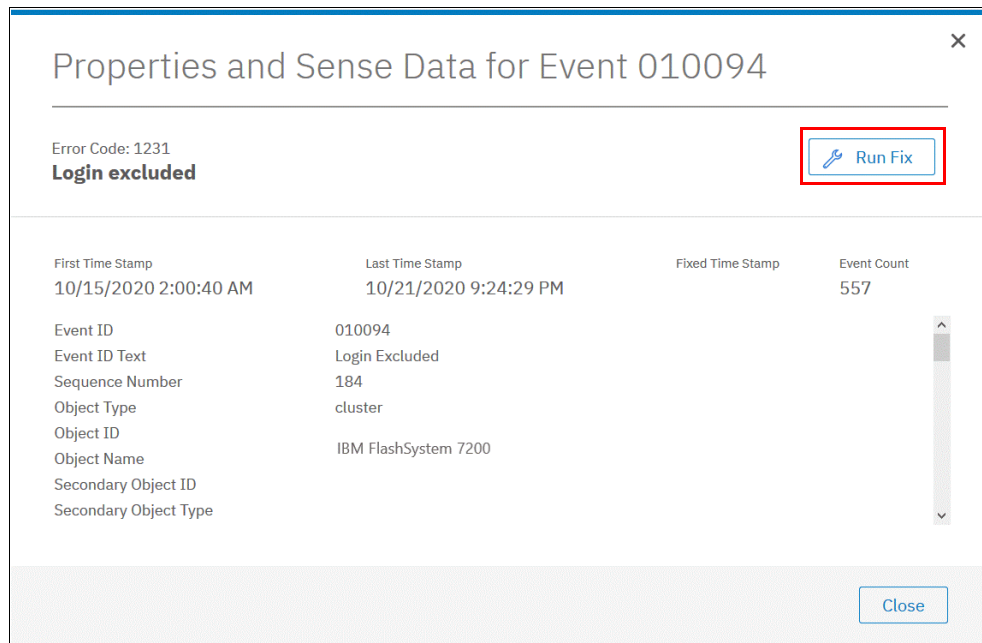


Figure 4-15 Login excluded

Running tasks and suggested tasks

The middle icon in the notification area provides an overview of currently running tasks that are triggered by administrator. It also includes the suggested tasks that recommend that users perform specific configuration actions.

In the example that is shown in Figure 4-16, we have not yet defined remote support in the system. Therefore, the system suggests that we do so and offers us direct access to the associated **Remote Support** menu. Click **Run Task** to define remote support according to the procedure that is explained in Chapter 13, “Reliability, availability, and serviceability, monitoring and logging, and troubleshooting” on page 793. If you do not want to define remote support now, click **Not Now** and the suggestion message disappears.

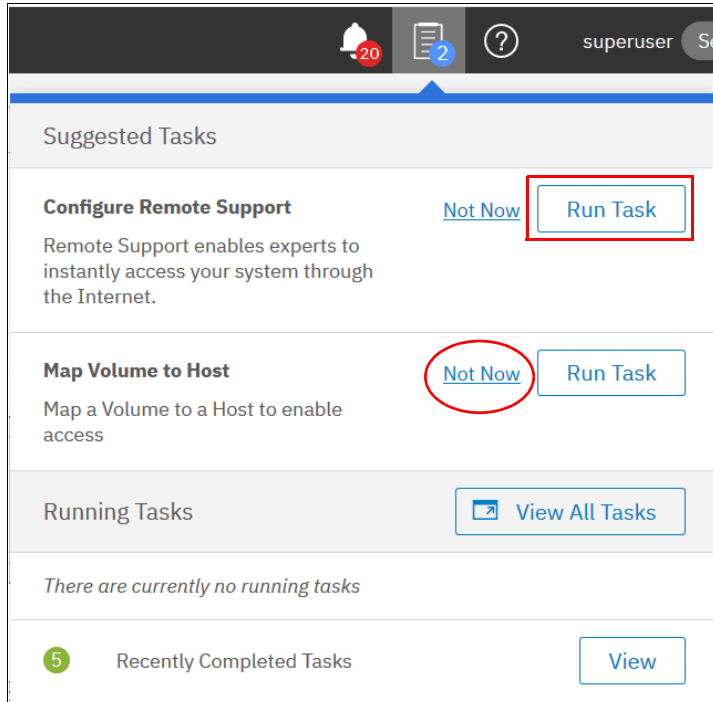


Figure 4-16 Configure Remote Support window

Similarly, you can analyze the details of running tasks (all of them together in one window or of a single task). Click **View** to open the volume format job, as shown in Figure 4-17.

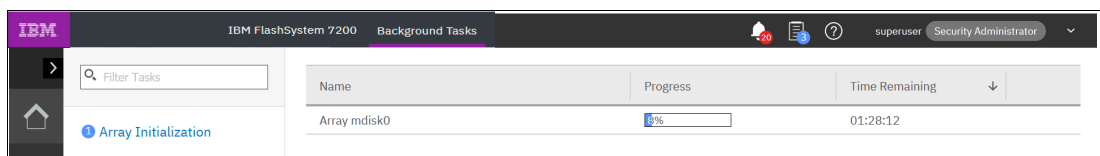


Figure 4-17 Details of a running task

The following information can be displayed as part of the running tasks:

- ▶ Volume migration
- ▶ MDisk removal
- ▶ Image mode migration
- ▶ Extent migration
- ▶ IBM FlashCopy
- ▶ Metro Mirror (MM) and Global Mirror (GM)
- ▶ Volume formatting
- ▶ Space-efficient copy repair
- ▶ Volume copy verification and synchronization
- ▶ Estimated time for the task completion

Making selections

Recent updates to the GUI brought improved selection making. You can now select multiple items more easily. Go to a wanted window, press and hold the Shift or Ctrl key, and make your selection.

Pressing and holding the Shift key, select the first item in your list that you want, and then select the last item. All items between the two that you choose are also selected, as shown in Figure 4-18.

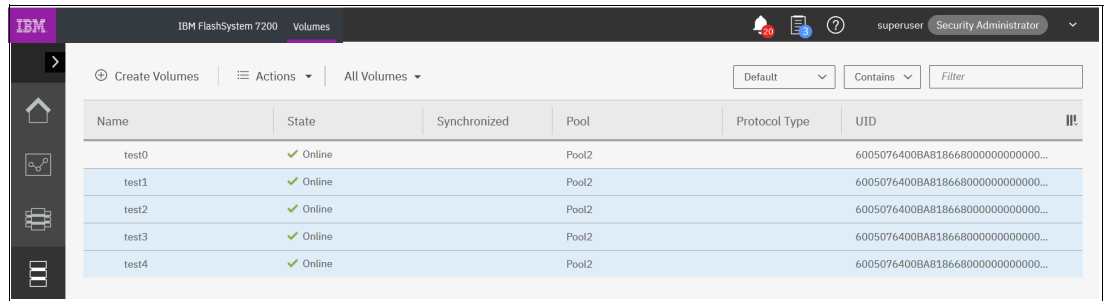


Figure 4-18 Selecting items by using the Shift key

Pressing and holding the Ctrl key, select any items from the entire list. You can select items that do not appear in sequential order, as shown in Figure 4-19.

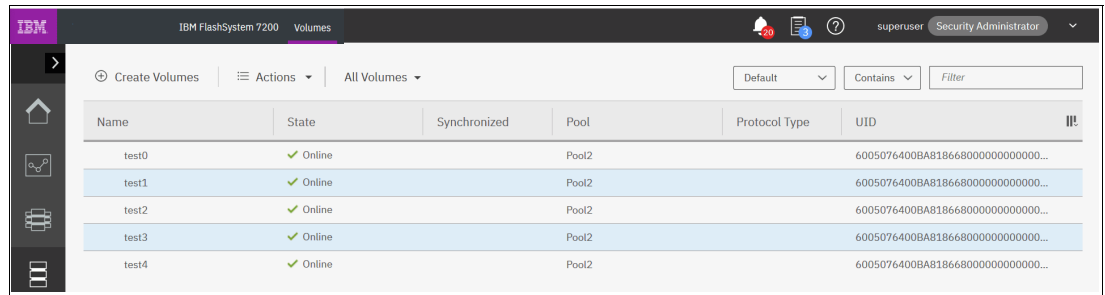


Figure 4-19 Selecting items by using the Ctrl key

You can also select items by using the built-in filtering function. For more information, see 4.3.1, “Content-based organization” on page 166.

Help

If you need help, you can select the (?) button, as shown in Figure 4-20.

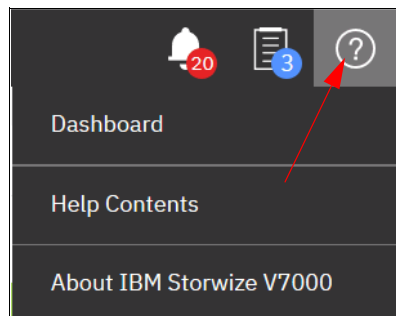


Figure 4-20 Access help menu

You see two options:

- ▶ The first option opens a new tab with plain text information about the window you are on and its contents.
- ▶ The second option shows the same information in IBM Documentation. This option requires an internet connection, but the first option does not because the information is stored locally on the system.

For example, in the Dashboard window, you can open help information that is related to the dashboard-provided information, as shown in Figure 4-21.

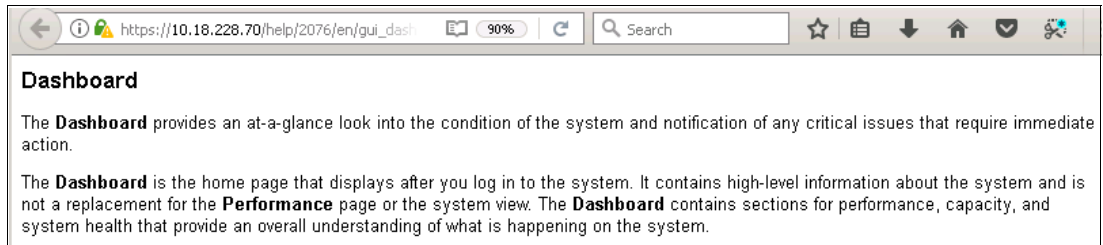


Figure 4-21 Example of Dashboard help content

4.3 System Hardware - Overview window

The System Hardware - Overview window is shown in Figure 4-22.

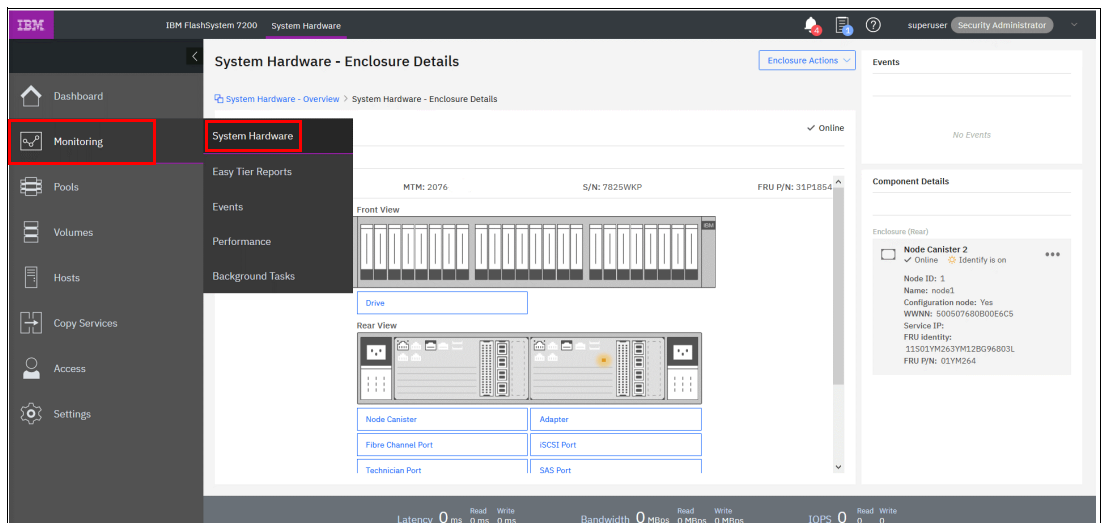


Figure 4-22 The System Hardware - Overview window

The next section describes the structure of the window and how to go to various system components to manage them more efficiently and quickly.

4.3.1 Content-based organization

The following sections describe several view options within the GUI in which you can filter (to minimize the amount of data that is shown on the window), sort, and reorganize the content of the window.

Table filtering

On most pages, a Filter box is available at the upper right of the window. Use this option if the list of object entries is too long and you want to search for something specific.

To use search filtering, complete the following steps:

1. In the **Filter** box that is shown in Figure 4-23, enter a search term by which you want to filter. You can also use the drop-down menus to modify what the system searches for. For example, if you want an exact match to your filter, select **=** instead of **Contains**. The first drop-down list limits your filter to search through a specific column only, for example, Name and State.

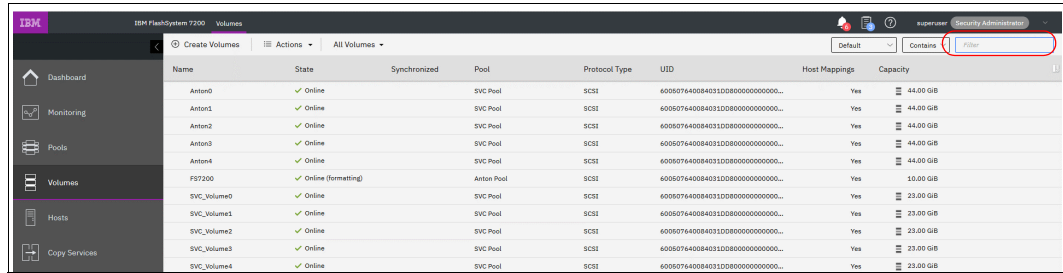


Figure 4-23 Filter search box

2. Enter the text string that you want to filter and press Enter.

By using this function, you can filter your table based on column names. In our example, a volume list is displayed that contains the names that include *Anton* somewhere in the name. *Anton* is highlighted in amber, as are any columns that contain this information, as shown in Figure 4-24. The search option is not case-sensitive.

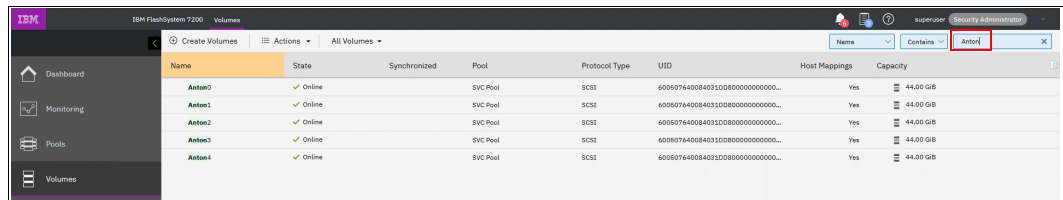


Figure 4-24 Showing filtered rows

3. Remove this filtered view by clicking the **X** icon that displays in the Filter box or by deleting what you searched for and pressing Enter, as shown in Figure 4-25.

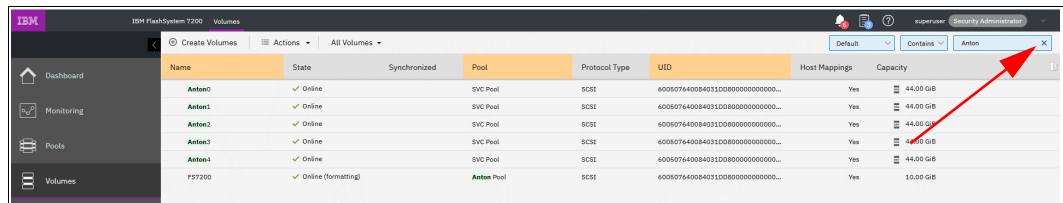


Figure 4-25 Removing the filtered view

Filtering: This filtering option is available in most menu options of the GUI.

Table information

In the table view, you can add or remove the information in the tables on most pages.

For example, on the Volumes window, complete the following steps to add a column to the table:

1. Right-click any column headers of the table or select the icon in the upper left of the table header. A list of all of the available columns displays, as shown in Figure 4-26.

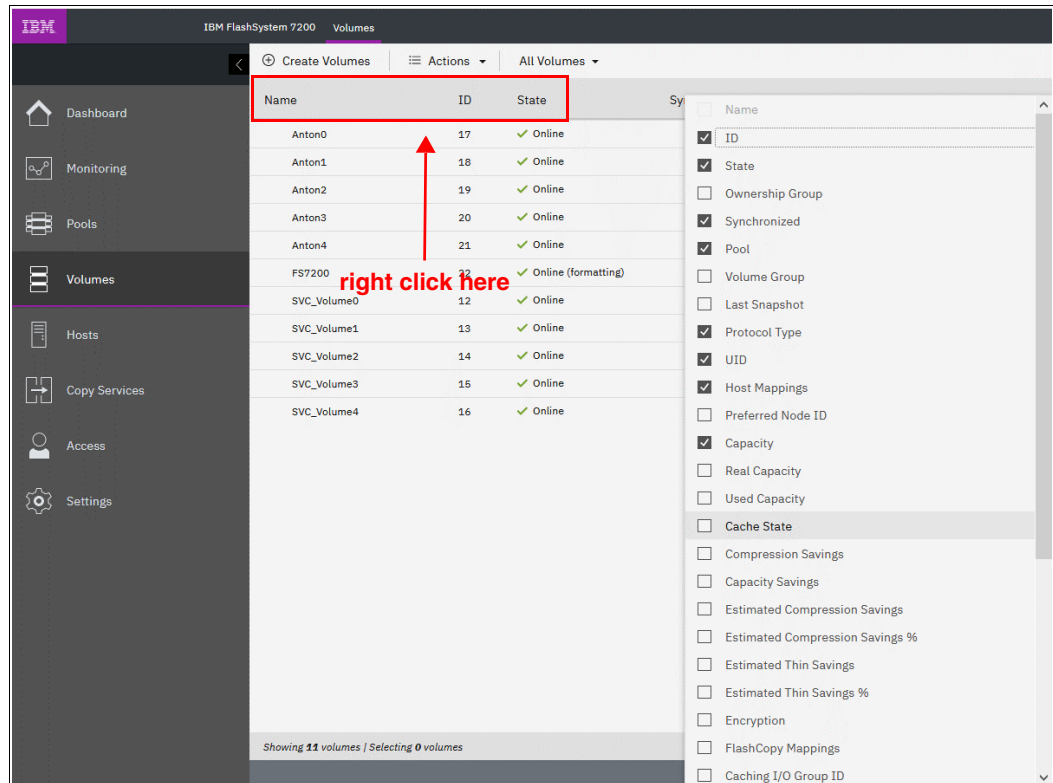


Figure 4-26 Adding or removing details in a table

- Select the column that you want to add or remove from this table. In our example, we added the volume ID column and sorted the content by ID, as shown on the left in Figure 4-27.

Name	State	ID ↑	Synchronized
SVC_Volume0	✓ Online		
SVC_Volume1	✓ Online		
SVC_Volume2	✓ Online		
SVC_Volume3	✓ Online	15	
SVC_Volume4	✓ Online	16	
Anton0	✓ Online	17	
Anton1	✓ Online	18	
Anton2	✓ Online	19	
Anton3	✓ Online	20	
Anton4	✓ Online	21	
FS7200	✓ Online (formatting)	22	

Figure 4-27 Table with an added ID column

- You can repeat this process several times to create custom tables to meet your requirements.
- Return to the default table view by selecting **Restore Default View** (the last entry) in the column selection menu.

Sorting: By clicking a column, you can sort a table based on that column in ascending or descending order.

Shifting columns in tables

You can move columns by clicking and moving the column right or left, as shown in Figure 4-28. In this example, we attempt to move the Capacity column before the Pool column.

Name	State	ID ↑	Synchronized	Pool	Protocol Type	UID	Host Mappings	Capacity
SVC_Volume0	✓ Online	12	Capacity	SVC Pool	SCSI	60807640084031D080000000000000...	Yes	23.00 GiB
SVC_Volume1	✓ Online	13		SVC Pool	SCSI	60807640084031D080000000000000...	Yes	23.00 GiB
SVC_Volume2	✓ Online	14		SVC Pool	SCSI	60807640084031D080000000000000...	Yes	23.00 GiB
SVC_Volume3	✓ Online	15		SVC Pool	SCSI	60807640084031D080000000000000...	Yes	23.00 GiB
SVC_Volume4	✓ Online	16		SVC Pool	SCSI	60807640084031D080000000000000...	Yes	23.00 GiB
Anton0	✓ Online	17		SVC Pool	SCSI	60807640084031D080000000000000...	Yes	44.00 GiB
Anton1	✓ Online	18		SVC Pool	SCSI	60807640084031D080000000000000...	Yes	44.00 GiB
Anton2	✓ Online	19		SVC Pool	SCSI	60807640084031D080000000000000...	Yes	44.00 GiB
Anton3	✓ Online	20		SVC Pool	SCSI	60807640084031D080000000000000...	Yes	44.00 GiB
Anton4	✓ Online	21		SVC Pool	SCSI	60807640084031D080000000000000...	Yes	44.00 GiB
FS7200	✓ Online (formatting)	22		Anton Pool	SCSI	60807640084031D080000000000000...	Yes	10.00 GiB

Figure 4-28 Reorganizing table columns

4.4 Monitoring menu

Click the **Monitoring** icon in left pane to open the **Monitoring** menu (see Figure 4-29). The **Monitoring** menu offers these navigation options:

- ▶ **System Hardware:** This option opens an overview of the system. It shows all control enclosures and groups them into I/O groups if more than one control enclosure is present. Useful information about each enclosure is displayed, including status, number of events against each enclosure, and key enclosure information, such as ID and serial number. For more information, see 4.4.1, “System Hardware overview” on page 171.
- ▶ **Easy Tier Reports:** This option gives you an overview of the Easy Tier Activities in your system. Easy Tier eliminates manual intervention when you assign highly active data on volumes to faster responding storage. In this dynamically tiered environment, data movement is seamless to the host application regardless of the storage tier in which the data belongs. However, you can manually change the default behavior. For example, you can turn off Easy Tier on pools that have any combination of the four types of MDisks. If a pool contains one type of MDisk, Easy Tier goes into balancing mode. When the pool contains multiple types of MDisks, Easy Tier is automatically turned on. For more information, see 4.4.2, “Easy Tier Reports” on page 175.
- ▶ **Events:** This option tracks all informational, warning, and error messages that occurred in the system. You can apply various filters to sort the messages according to your needs or export the messages to an external comma-separated value (CSV) file. For more information, see 4.4.3, “Events” on page 177.
- ▶ **Performance:** This option reports the general system statistics that relate to the processor (CPU) utilization, host and internal interfaces, volumes, and MDisks. With this option, you can switch between megabytes per second (MBps) or IOPS. For more information, see 4.4.4, “Performance” on page 178.
- ▶ **Background Tasks:** The option shows the progress of all tasks running in the background as listed in 4.4.5, “Background Tasks” on page 179.

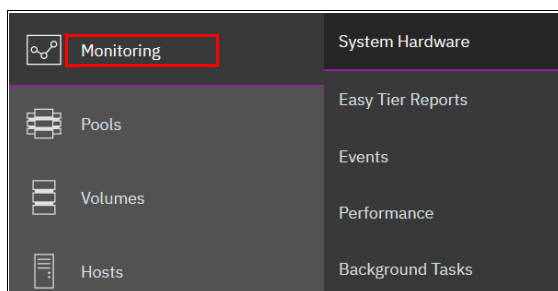


Figure 4-29 Monitoring menu

4.4.1 System Hardware overview

The **System Hardware** option on the **Monitoring** menu provides a general overview. If you have more than one control enclosure in a cluster, each enclosure has its own I/O group section (see Figure 4-30).

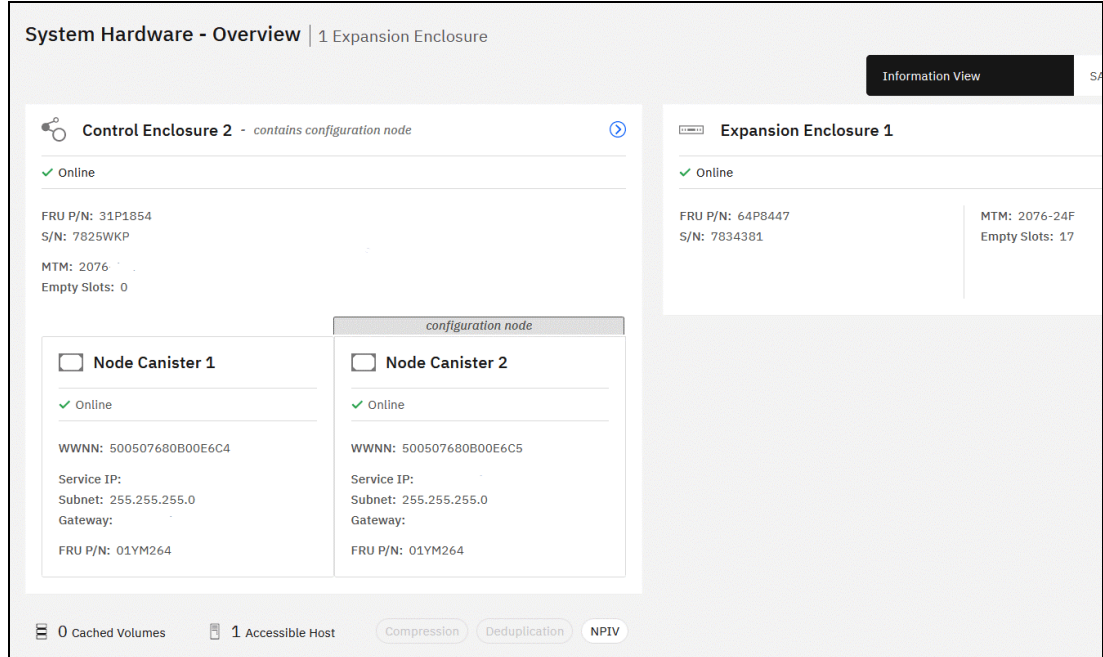


Figure 4-30 System Hardware overview

Figure 4-31 shows how to see more about the System Hardware Enclosure Details.

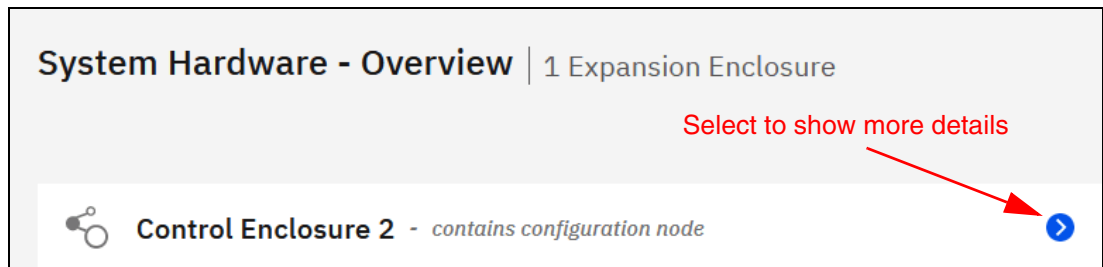


Figure 4-31 Selecting more System Hardware Enclosure Details

This view shows all external components in real time. In Figure 4-32, you can see that one of the canisters identifies light-emitting diodes (LEDs) is lit (see red box). You can click any component in the graphic view or on the list view at the bottom to view details.

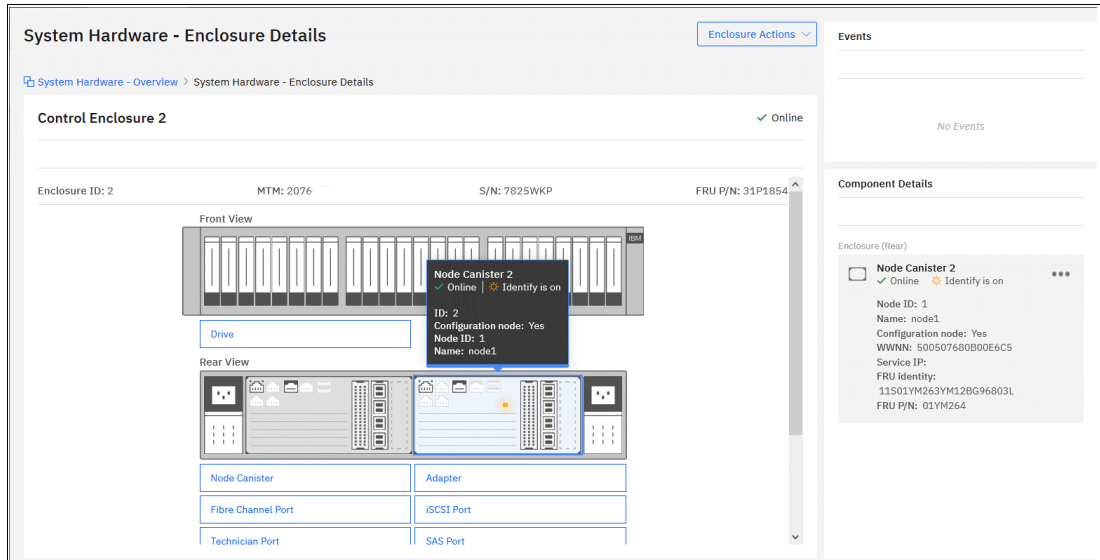


Figure 4-32 System Hardware Enclosure Details

For example, clicking a node brings up details, such as whether the node is online and which node is the configuration node, as shown in Figure 4-33. For more information about the component, see the right side under the Component Details section that shows when a component is selected.

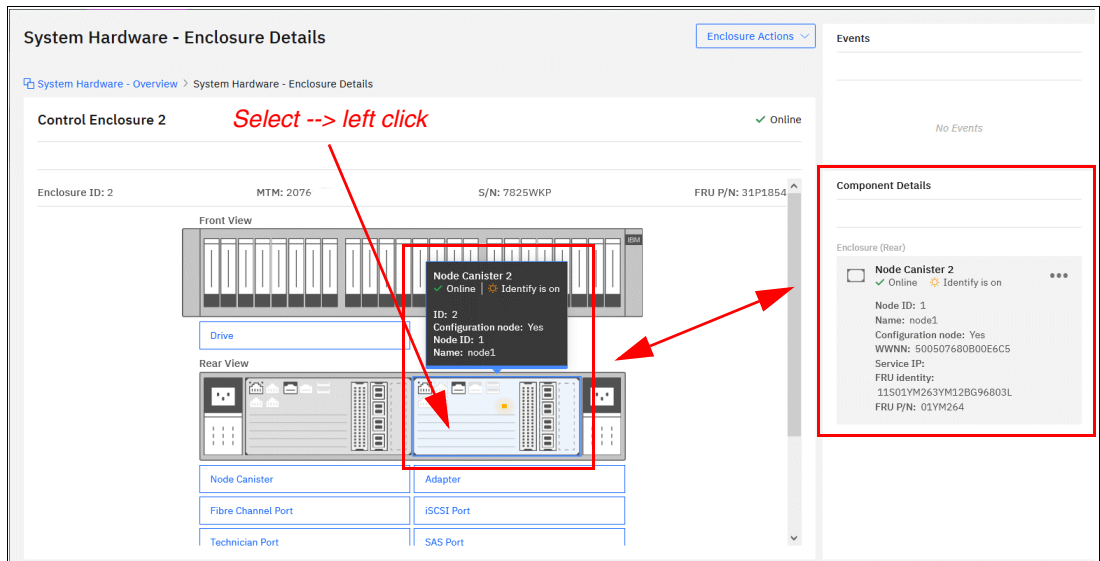


Figure 4-33 Showing the node canister details

By right-clicking and selecting **Properties**, you see detailed technical parameters, such as WWNN, Memory, CPU, and field-replaceable unit (FRU) number, as shown in Figure 4-34.

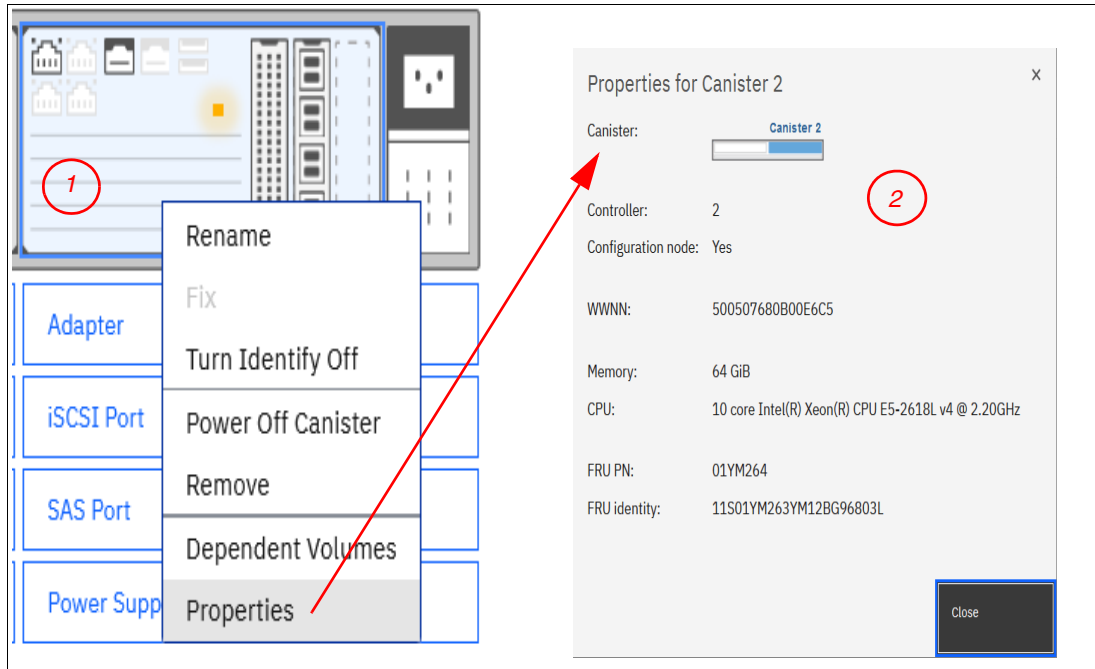


Figure 4-34 Canister information

In an environment with multiple IBM Storage System clusters, you can easily direct the onsite personnel or technician to the correct device by enabling the identification LED on the front panel by completing the following steps:

1. Select the appropriate drive and click **Turn Identify On**, as shown in Figure 4-35.

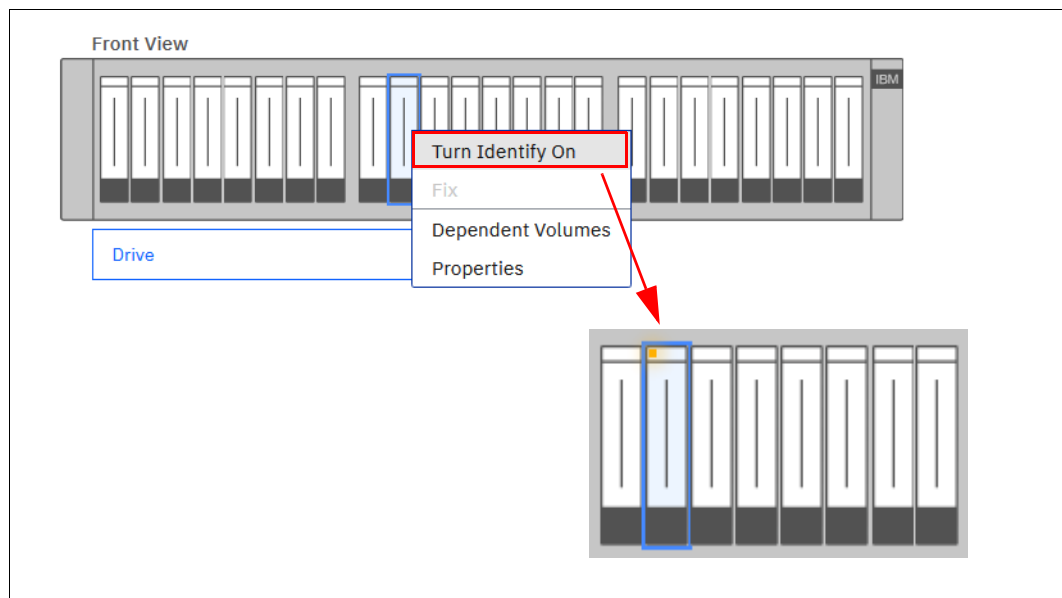


Figure 4-35 Turning on the Identify LED

2. Wait for confirmation from the technician that the device in the data center was correctly identified. In the GUI, you see a flashing light, which indicates that the Identify LED was turned on.
3. After the confirmation, click **Turn Identify Off** (see Figure 4-36).

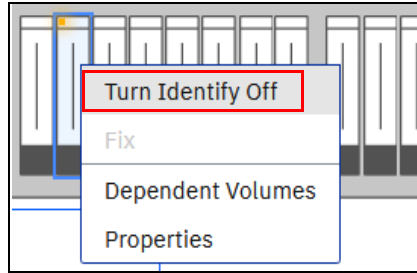


Figure 4-36 Turning off the Identify LED

Alternatively, you can use the command-line interface (CLI) to get the same results. Enter the following commands in this sequence:

1. Type `svctask chenclosure -identify yes 1` (or enter `chenclosure -identify yes 1`).
2. Type `svctask chenclosure -identify no 1` (or enter `chenclosure -identify no 1`).

You can use the same CLI to obtain results for a specific controller or drive.

To view internal components (components that cannot be seen from the outside), review the bottom of the GUI underneath where the list of external components is displayed. You can select any of these components and details display in the right pane, as with the external components. Figure 4-37 shows the backside of the enclosure.



Figure 4-37 Viewing the internal components

You can also choose **SAS Chain View** to view directly attached expansion enclosures, as shown in Figure 4-38. A useful view of the entire serial-attached Small Computer System Interface (SCSI) (SAS) chain is displayed, with selectable components that show port numbers and canister numbers, along with a cable diagram for easy cable tracking.

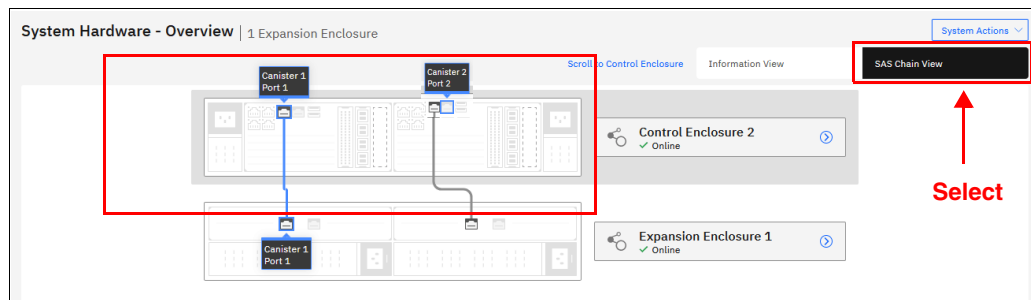


Figure 4-38 SAS Chain View

You can select any enclosure to get more information, including serial number and model type, as shown in Figure 4-39, where Expansion Enclosure 1 is selected. You can also see the Events and Component Details areas at the right side of the window, which shows information that relates to the enclosure or component that you select.

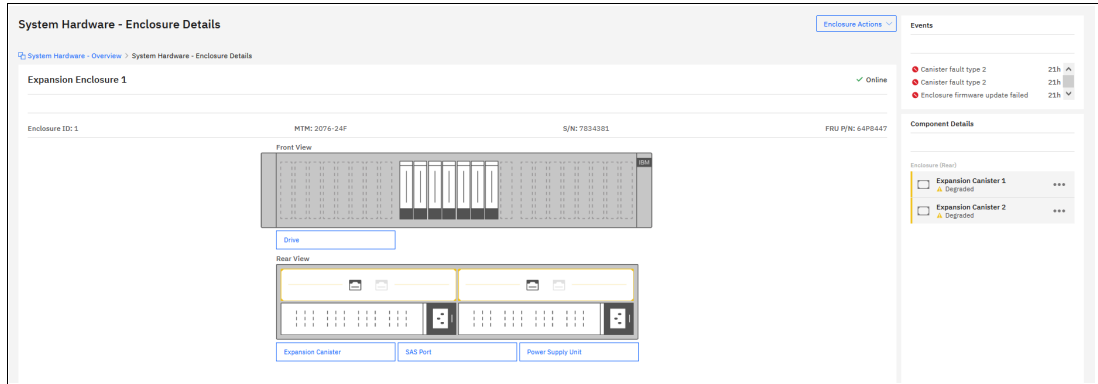


Figure 4-39 Enclosure Details window

With directly attached expansion enclosures, the view is condensed to show all expansion enclosures on the right side, as shown in Figure 4-40. The number of events against each enclosure and the enclosure status are displayed for quick reference. Each enclosure is selectable, which brings you to the Expansion Enclosure View window.

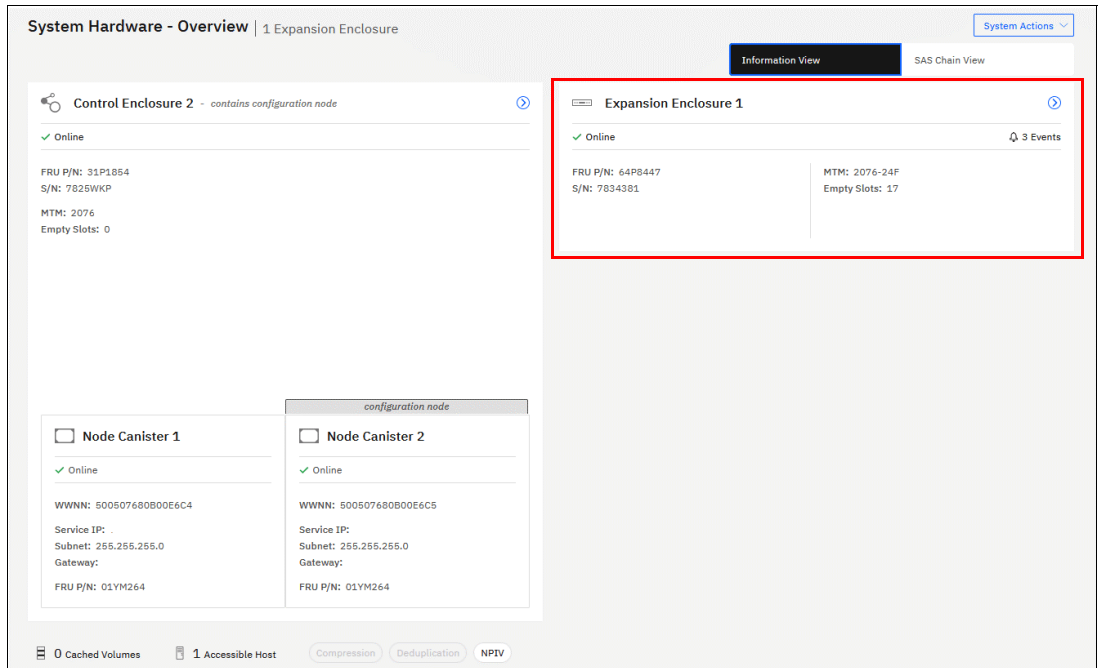


Figure 4-40 System Overview with attached enclosures

4.4.2 Easy Tier Reports

The management GUI supports monitoring Easy Tier data movement in graphical reports to help you understand what is happening with the performance of your storage device. Charts for data movement, tier composition, and workload skew comparison can be viewed as web-generated HTML files in a browser, or can be downloaded as CSV files.

Data is collected by the IBM Storage Tier Advisor Tool (IBM STAT) tool in 5-minute increments. When data that is displayed in increments that are larger than 5 minutes (for example, 1 hour), the data that is displayed for that 1 hour is the sum of all the data points that were received for that 1-hour time span.

To view Easy Tier data and reports in the management GUI, select one of the following paths:

- ▶ From the management GUI, select **Monitoring** → **Easy Tier Reports**.
- ▶ From the management GUI, select **Pools** → **View Easy Tier Reports**.

Data Movement statistics

The Data Movement chart displays the migration actions that are triggered by Easy Tier.

Tier composition statistics

The Tier Composition chart displays the distributed workload between the top tier, middle tier, and bottom tier. Each tier is composed of one or more tier types.

Workload Skew Comparison

The Workload Skew Comparison chart displays the percentage of I/O workload compared to the total capacity.

Figure 4-41 shows how to export the Easy Tier Reports.

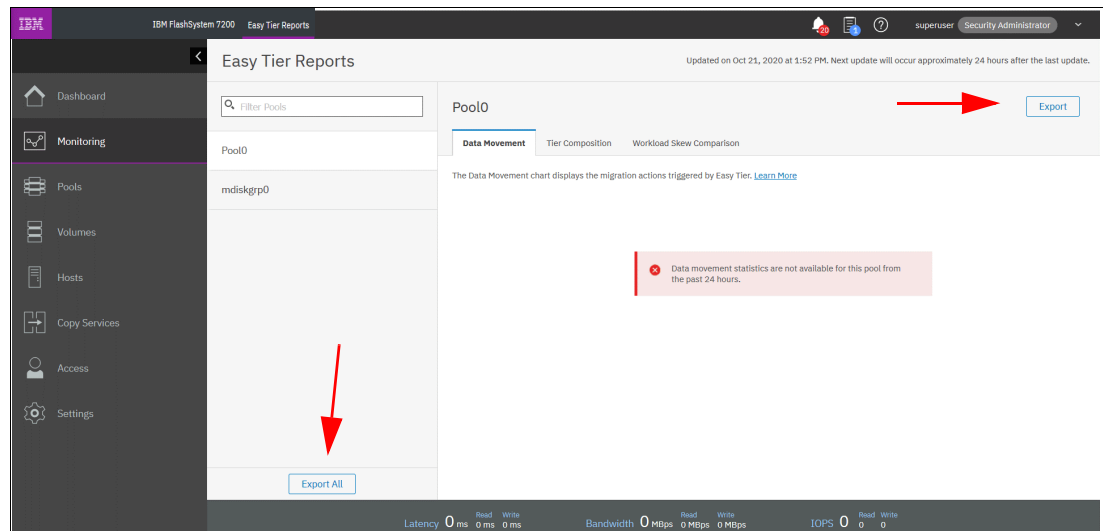


Figure 4-41 Easy Tier Reports

You can export your Easy Tier stats to a CSV file for further analysis. For more information about Easy Tier Reports, see Chapter 9, “Advanced features for storage efficiency” on page 509.

4.4.3 Events

The **Events** option, which is available in the **Monitoring** menu, tracks all informational, warning, and error messages that occur in the system. You can apply various filters to sort them, or export them to an external CSV file. A CSV file can be created from the information that is shown here. Figure 4-42 provides an example of records in the system Event log.

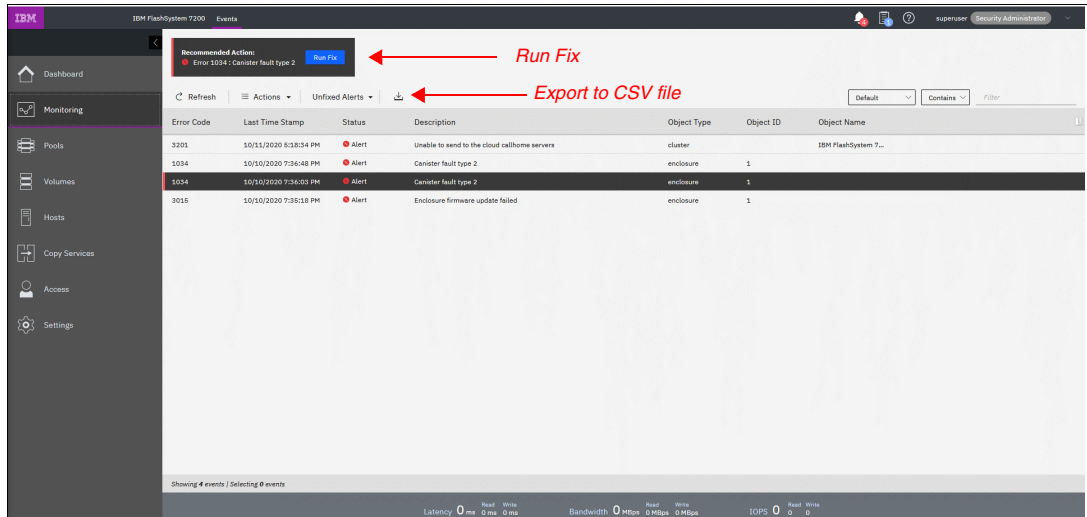


Figure 4-42 Event log list

For the error messages with the highest internal priority, perform corrective actions by running fix procedures. Click **Run Fix** (see Figure 4-42), and the fix procedure wizard opens, as shown in Figure 4-43.

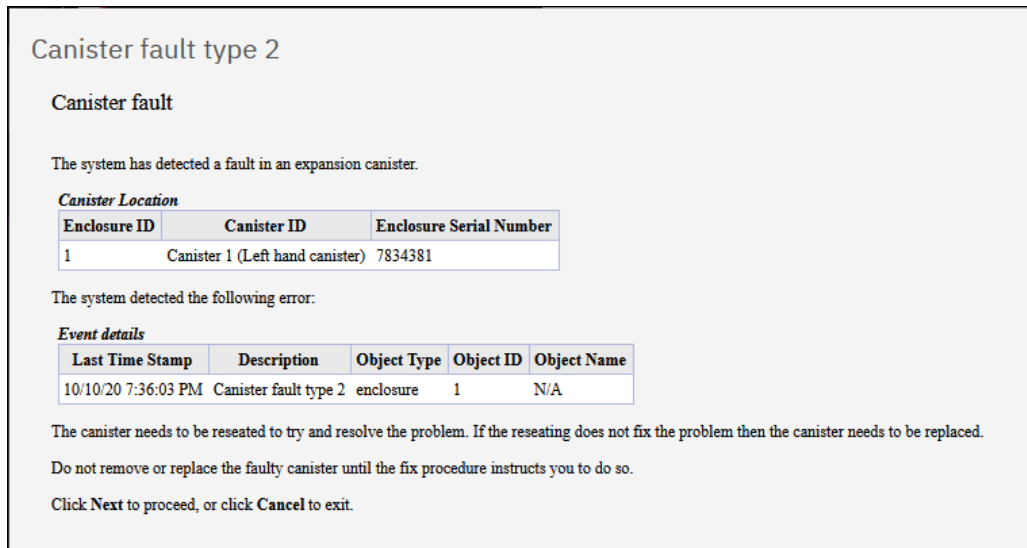


Figure 4-43 Performing a fix procedure

The wizard guides you through the troubleshooting and fixing process from a hardware or software perspective. If you determine that the problem cannot be fixed without a technician's intervention, you can cancel the procedure execution at any time.

For more information about fix procedures, see Chapter 13, "Reliability, availability, and serviceability, monitoring and logging, and troubleshooting" on page 793.

4.4.4 Performance

The Performance pane reports the general system statistics that relate to processor (CPU) utilization, host and internal interfaces, volumes, and MDisks. You can switch between MBps or IOPS, and drill down in the statistics to the node level. This capability might be useful when you compare the performance of each control canister in the system if problems exist after a node failover occurs (see Figure 4-44).

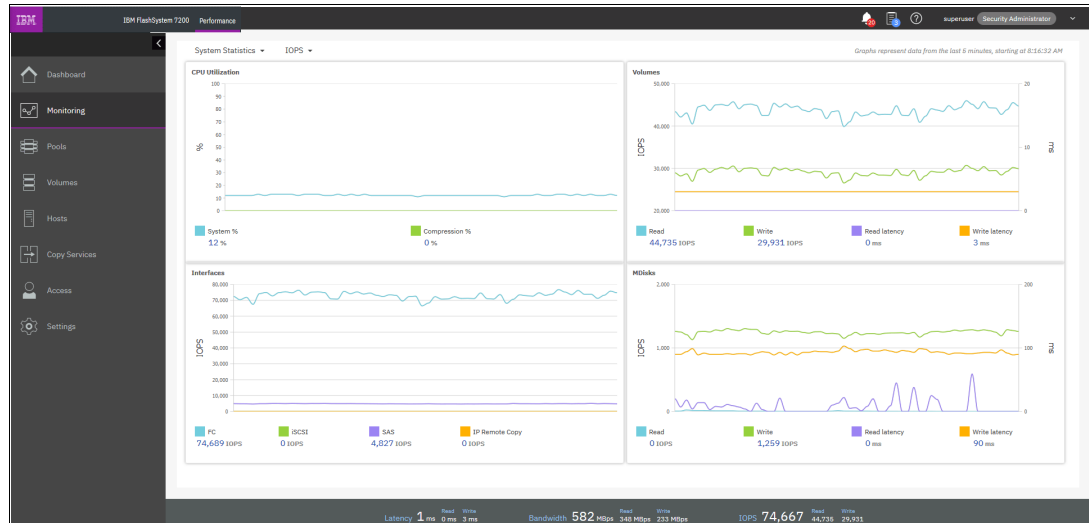


Figure 4-44 Performance statistics of the IBM Storage System

The performance statistics in the GUI show, by default, the latest 5 minutes of data. To see details of each sample, click the graph and select the timestamp, as shown in Figure 4-45.

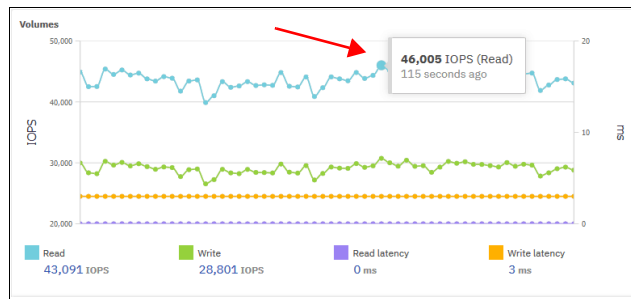


Figure 4-45 Sample details

The charts that are shown in Figure 4-45 represent 5 minutes of the data stream. For in-depth storage monitoring and performance statistics with historical data about your system, use IBM Spectrum Control or IBM Storage Insights.

You can also obtain a no-charge unsupported version of the Quick Performance Overview ([qperf](#)) from [this website](#).

4.4.5 Background Tasks

Use the Background Tasks window to view and manage current tasks that are running on the system (see Figure 4-46).

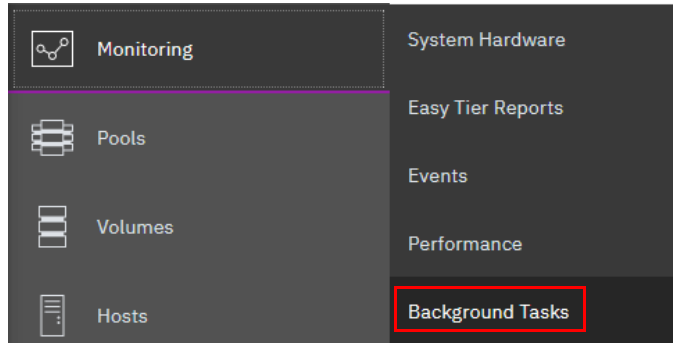


Figure 4-46 Selecting Background Tasks

This menu provides an overview of currently running tasks that are triggered by the administrator. In contrast to the Running jobs and Suggested tasks indication in the middle of top window, it does not list the suggested tasks that administrators should consider performing. The overview provides more details than the indicator, as shown in Figure 4-47.



Figure 4-47 List of running tasks

You can switch between each type (group) of operation, but you cannot show them all in one list (see Figure 4-48).



Figure 4-48 Switching between types of background tasks

4.5 Pools

The **Pools** menu option is used to configure and manage storage pools, internal, and external storage, MDisks, and to migrate old attached storage to the system.

The **Pools** menu contains the following items accessible from GUI (see Figure 4-49):

- ▶ Pools
- ▶ Volumes by Pool
- ▶ Internal Storage
- ▶ External Storage
- ▶ MDisks by Pool
- ▶ System Migration

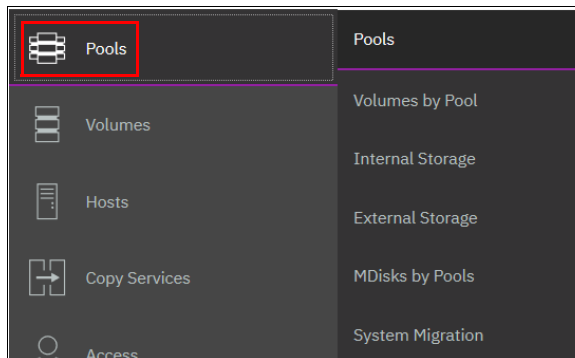


Figure 4-49 Pools menu

For more information about storage pool configuration and management, see Chapter 5, “Storage pools” on page 237.

4.6 Volumes

A **volume** is a logical disk that the system presents to attached hosts. By using GUI operations, you can create different types of volumes depending on the type of topology that is configured on your system.

The **Volumes** menu contains the following items, as shown in Figure 4-50 on page 181:

- ▶ Volumes
- ▶ Volumes by Pool
- ▶ Volumes by Host
- ▶ Volumes by Host Cluster
- ▶ Cloud Volumes

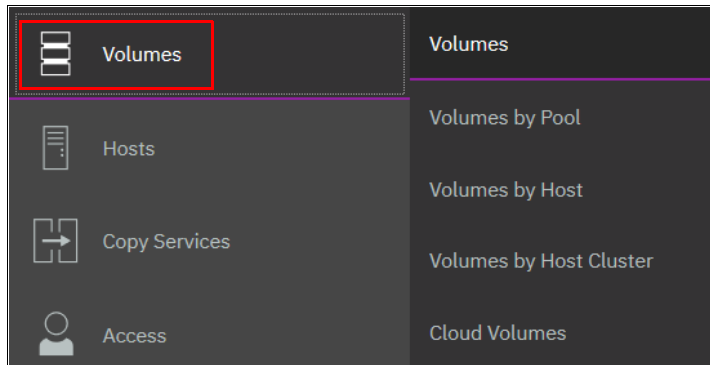


Figure 4-50 Volumes menu

For more information about these tasks and configuration and management process guidance, see Chapter 6, “Volumes” on page 299.

4.7 Hosts

A host system is a computer that is connected to the system through a Fibre Channel (FC) interface or an IP network. It is a logical object that represents a list of worldwide port names (WWPNs) that identify the interfaces that the host uses to communicate with your System. FC and SAS connections use WWPNs to identify the host interfaces to the systems.

The **Hosts** menu consists of the following choices, as shown in Figure 4-51:

- ▶ Hosts
- ▶ Host Clusters
- ▶ Ports by Host
- ▶ Mappings
- ▶ Volumes by Host
- ▶ Volumes by Host Cluster

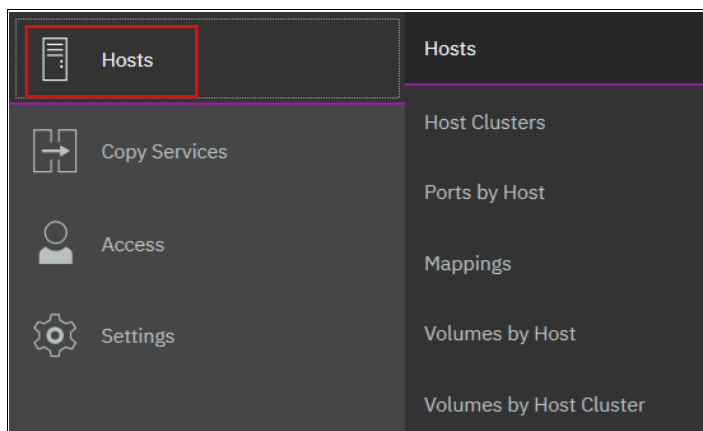


Figure 4-51 Hosts menu

For more information about configuration and management of hosts by using the GUI, see Chapter 7, “Hosts” on page 405.

4.8 Copy Services

The IBM Spectrum Virtualize Copy Services and Volumes Copy operations are based on the FlashCopy function. In its basic mode, the function creates copies of content on a source volume to a target volume. Any data on the target volume is lost and is replaced by the copied data.

More advanced functions allow FlashCopy operations to occur on multiple source and target volumes. Management operations are coordinated to provide a common, single point-in-time (PiT) for copying target volumes from their respective source volumes. This technique creates a consistent copy of data that spans multiple volumes.

The Copy Services menu offers the following operations in the GUI, as shown in Figure 4-52:

- ▶ FlashCopy
- ▶ Consistency groups
- ▶ FlashCopy Mappings
- ▶ Remote Copy (RC)
- ▶ Partnerships

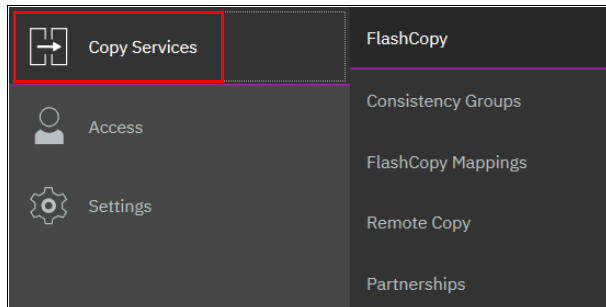


Figure 4-52 Copy Services in GUI

Because Copy Services is one of the most important features for resiliency solutions, see Chapter 10, “Advanced Copy Services” on page 553.

4.9 Access

The **Access** menu in the GUI maintains who can log in to the system, defines the access rights to the user, and tracks what was done by each privileged user to the system. It is logically split into three categories:

- ▶ Ownership groups
- ▶ Users by group
- ▶ Audit log

In this section, we explain how to create, modify, or remove a user, and how to see records in the audit log.

The **Access** menu is available from the left pane, as shown in Figure 4-53.

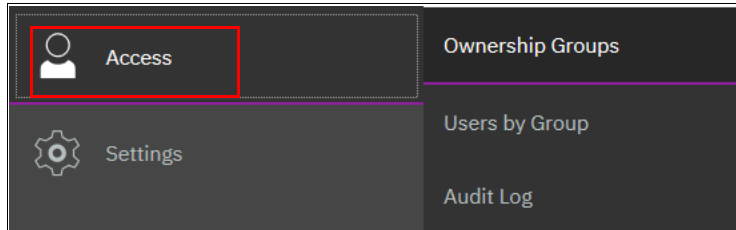


Figure 4-53 Access menu

4.9.1 Ownership groups

An *ownership group* defines a subset of users and objects within the system. You can create ownership groups to further restrict access to specific resources that are defined in the ownership group. Only users with Administrator or Security Administrator roles can configure and manage ownership groups. Ownership groups restrict access to only those objects that are defined within that ownership group. An *owner* is a user with an ownership group that can view and manipulate objects within that group.

The first time that you start the Ownership Group task, you see the window that is shown in Figure 4-54.

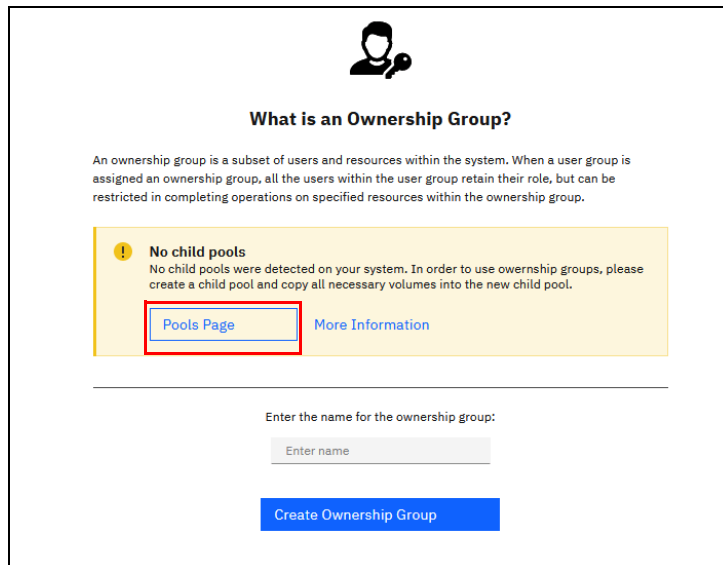


Figure 4-54 Start of an Ownership Group

In our example, no child pool exists, so the GUI guides you to the Pools page to create child pools.

The system supports several resources that you assign to ownership groups:

- ▶ Child pools
- ▶ Volumes
- ▶ Volume groups
- ▶ Hosts
- ▶ Host clusters
- ▶ Host mappings

- ▶ FlashCopy mappings
- ▶ FlashCopy consistency groups
- ▶ User groups

Two basic use cases can be applied to using ownership groups on the system:

- ▶ New objects are created within the ownership group. There also can be other existing objects on the system that are not in the ownership group.
- ▶ On a system where these supported objects are already configured, and you want to migrate these objects to use ownership groups.

When a user group is assigned to an ownership group, the users in that user group retain their role but are restricted to only those resources within the same ownership group. User groups can define the access to operations on the system, and the ownership group can further limit access to individual resources. For example, you can configure a user group with the Copy Operator role, which limits access of the user to Copy Services functions, such as FlashCopy and RC operations. Access to individual resources, such as a specific FlashCopy consistency group, can be further restricted by adding it to an ownership group. When the user logs on to the management GUI, only resources that they have access to through the ownership group are displayed. Additionally, only events and commands that are related to the ownership group in which a user belongs are viewable by those users.

Inheriting ownership

Depending on the type of resource, ownership can be defined explicitly or ownership can be inherited from the user, user group, or from other parent resources. Objects inherit their ownership group from other objects whenever possible:

- ▶ Volumes inherit the ownership group from the child pool that provides capacity for the volumes.
- ▶ FlashCopy mappings inherit the ownership group from the volumes that are configured in the mapping.
- ▶ Hosts inherit the ownership group from the host cluster they belong to, if applicable.
- ▶ Host mappings inherit the ownership group from both the host and the volume to which the host is mapped.

These objects cannot be explicitly moved to a different ownership group without creating inconsistent ownership.

Ownership groups are also inherited from the user. Objects that are created by an owner inherit the ownership group of the owner. If the owner is in more than one ownership group (only possible for remote users), then the owner must choose the group when the object is created.

Figure 4-55 on page 185 shows how different objects inherit ownership from ownership groups.

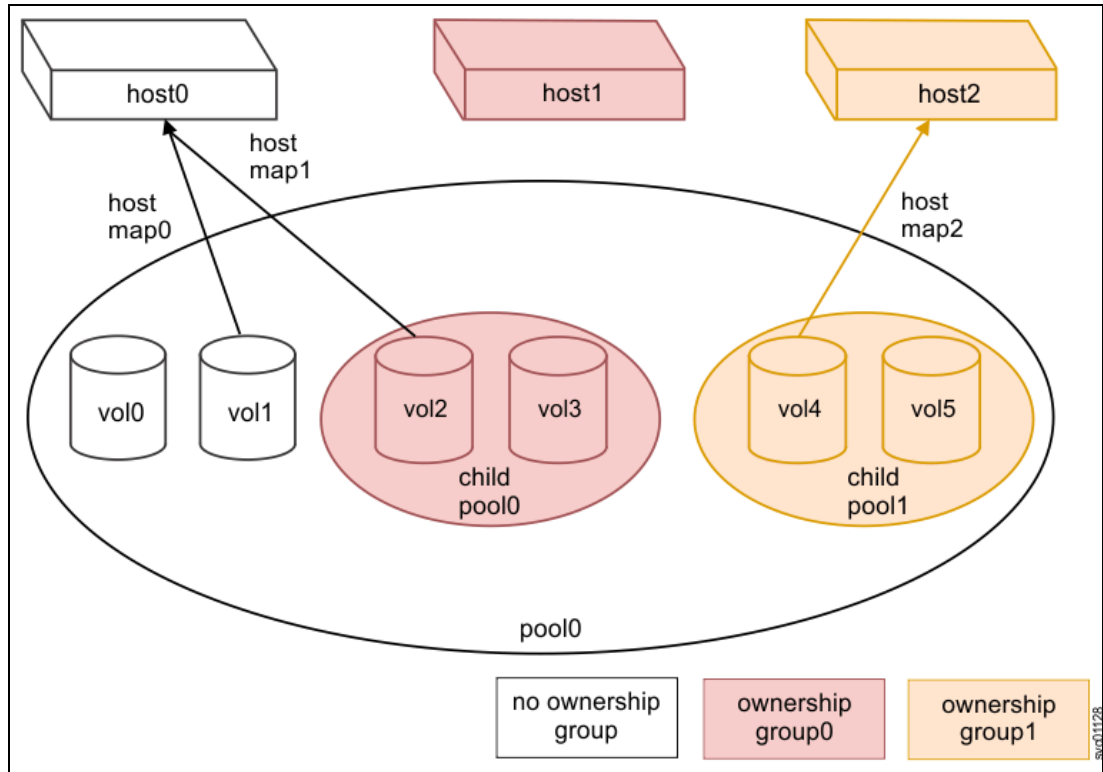


Figure 4-55 Ownership group inheritance

The following objects have ownership that is assigned explicitly and do not inherit ownership from other parent resources:

- ▶ Child pools
- ▶ Host clusters
- ▶ Hosts that are not part of a host cluster
- ▶ Volume groups
- ▶ FlashCopy consistency groups
- ▶ User groups
- ▶ Hosts that are a part of a host cluster
- ▶ Volumes
- ▶ Users
- ▶ Volume-to-host mappings
- ▶ FlashCopy mappings
- ▶ Configuring ownership groups
- ▶ Migrating to ownership groups

Child pools

The following rules apply to child pools that are defined in ownership groups:

- ▶ Child pools can be assigned to an ownership group when you create a pool or change a pool.
- ▶ Users who assign the child pool to the ownership group cannot be defined within that ownership group.
- ▶ Resources that are within the child pool inherit the ownership group that is assigned for the child pool.

Host clusters

The following rules apply to host clusters that are defined in ownership groups:

- ▶ If the user who is creating the host cluster is defined in only one ownership group, the host cluster inherits the ownership group of that user.
- ▶ If the user is defined in an ownership group but is also defined in multiple user groups, the host cluster inherits the ownership group. The system uses the lowest role that the user has from the user group. For example, if a user is defined in two user groups with the roles of Monitor and Copy Operator, the host cluster inherits the Monitor role.
- ▶ Only users not within an ownership group can assign ownership groups when a host cluster is created or changed.

Hosts that are not part of a host cluster

The following rules apply to a host that are not part of a host cluster that is defined in ownership groups:

- ▶ If the user who is creating the host is in only one ownership group, the host cluster inherits the ownership group of that user.
- ▶ If the user is defined in an ownership group but is also defined in multiple user groups, the host inherits the ownership group. The system uses the lowest role that the user has from the user group. For example, if a user is defined in two user groups with the roles of Monitor and Copy Operator, the host inherits the Monitor role.
- ▶ Only users not within an ownership group can assign ownership groups when you create a new host or change an existing host.

Volume groups

Volume groups can be created to manage multiple volumes that are used with Transparent Cloud Tiering (TCT) support. The following rules apply to volume groups that are defined in ownership groups:

- ▶ If the user that is creating the volume group is defined in only one ownership group, the volume group inherits the ownership group of that user.
- ▶ If the user is defined in an ownership group but is also defined in multiple user groups, the volume group inherits the ownership group. The system uses the lowest role that the user has from the user group. For example, if a user is defined in two user groups with the roles of Monitor and Copy Operator, the host inherits the Monitor role.
- ▶ Only users not within an ownership group can assign ownership groups when you create a new volume group or change an existing volume group.
- ▶ Volumes can be added to a volume group if both the volume and the volume group are within the same ownership group or if both are not in an ownership group. There are situations where a volume group and its volumes can belong to different ownership groups. Volume ownership can be inherited from the ownership group or from one or more child pools.
- ▶ The ownership of a volume group does not affect the ownership of the volumes it contains. If a volume group and its volumes are owned by different ownership groups, then the owner of the child pool that contains the volumes can change the volume directly. For example, the owner of the child pool can change the name of a volume within it. The owner of the volume group can change the volume group itself and indirectly change the volume, such as deleting a volume from the volume group. Neither the ownership group of the child pools or the owner of the volume group can directly manipulate the resources that are not defined in their ownership group.

FlashCopy consistency groups

FlashCopy consistency groups can be created to manage multiple FlashCopy mappings. The following rules apply to FlashCopy consistency groups that are defined in ownership groups:

- ▶ If the user that is creating the FlashCopy consistency group is in only one ownership group, the FlashCopy consistency group inherits the ownership group of that user.
- ▶ If the user is defined in an ownership group but is also defined in multiple user groups, the FlashCopy consistency group inherits the ownership group. The system uses the lowest role that the user has from the user group.
- ▶ Only users not within an ownership group can assign ownership groups when a FlashCopy consistency is created or changed.
- ▶ FlashCopy mappings can be added to a consistency group if the volumes in the mapping and the consistency group are within the same ownership group. You can also add a FlashCopy mapping to a consistency group if it and all of its dependent resources are not in an ownership group.
- ▶ There are situations where a FlashCopy consistency group and its resources can belong to different ownership groups.
- ▶ As with volume groups and volumes, the ownership of the consistency group has no impact on the ownership of the mappings it contains.

User groups

The following rules apply to user groups that are defined in ownership groups:

- ▶ If the user that is creating the user group is in only one ownership group, the user group inherits the ownership group of that user.
- ▶ If the user is with multiple user groups, the user group inherits the ownership group of the user group with the lowest role.
- ▶ Only users not within an ownership group can assign an ownership group when a user group is created or changed.

These resources inherit ownership from the parent resource. A user cannot change the ownership group of the resource, but can change the ownership group of the parent object.

Hosts that are a part of a host cluster

The following rules apply to hosts that are defined in ownership groups:

- ▶ The host inherits the ownership group of the host cluster to which it belongs.
- ▶ If a host is removed from a host cluster within an ownership group, the host inherits the ownership group of the host cluster to which it used to belong.
- ▶ If a host is removed from a host cluster that is not within an ownership group, the host inherits no ownership groups.
- ▶ Hosts can be added to a host cluster if the host and host cluster have the same ownership group.
- ▶ Changing the ownership group of a host cluster automatically changes the ownership group of all the hosts inside the host cluster.

Volumes

The following rules apply to volumes that are defined in ownership groups:

- ▶ The volume inherits the ownership group of the child pools that provide capacity for the volume and its copies.
- ▶ If the child pool that provides capacity for the volume or its copies is defined in different ownership groups, then the volume cannot be created in an ownership group.
- ▶ When creating a volume copy or migrating a volume in the CLI, use the **-inconsistentownershipgroup** flag to allow for inconsistent ownership groups. However, you should not leave volumes or volume copies in different ownership groups. After the migration, the user with the Security Administrator role must ensure that all volumes or copies are within the same ownership group as the users who need access.
- ▶ With volume groups, the volume group and its volumes can belong to different ownership groups. However, the ownership of a volume group does not impact the ownership of the volumes that it contains.

Users

The following rules apply to users that are defined in ownership groups:

- ▶ A user inherits the ownership group of the user group to which it belongs.
- ▶ Users that use Lightweight Directory Access Protocol (LDAP) for remote authentication can belong to multiple user groups and multiple ownership groups.

Volume-to-host mappings

The following rules apply to volume-to-host mappings that are defined in ownership groups:

- ▶ Volume-to-host mappings inherit the ownership group of the host or host cluster and volume in the mapping.
- ▶ If host or host cluster and volume are within different ownership groups, then the mapping cannot be assigned an ownership group.

FlashCopy mappings

The following rules apply to FlashCopy mappings that are defined in ownership groups:

- ▶ FlashCopy mappings inherit the ownership group of both volumes that are defined in the mapping.
- ▶ If the volumes are within different ownership groups, then the mapping cannot be assigned to an ownership group.
- ▶ Like with FlashCopy consistency groups, it is possible for a consistency group and its mappings to belong to different ownership groups. However, the ownership of the consistency group has no impact on the ownership of the mappings that it contains.

Configuring ownership groups

You can configure ownership groups to manage access to resources on the system. An ownership group defines a subset of users and objects within the system. You can create ownership groups to further restrict access to specific resources that are defined in the ownership group. Only users with Administrator or Security Administrator roles can configure and manage ownership groups.

Migrating to ownership groups

If you updated your system to a software level that supports ownership groups, you must reconfigure certain resources if you want to configure ownership groups. An ownership group defines a subset of users and objects within the system. You can create ownership groups to further restrict access to specific resources that are defined in the ownership group. Only users with the Administrator or Security Administrator roles can configure and manage ownership groups.

Figure 4-56 shows an example of an ownership group.

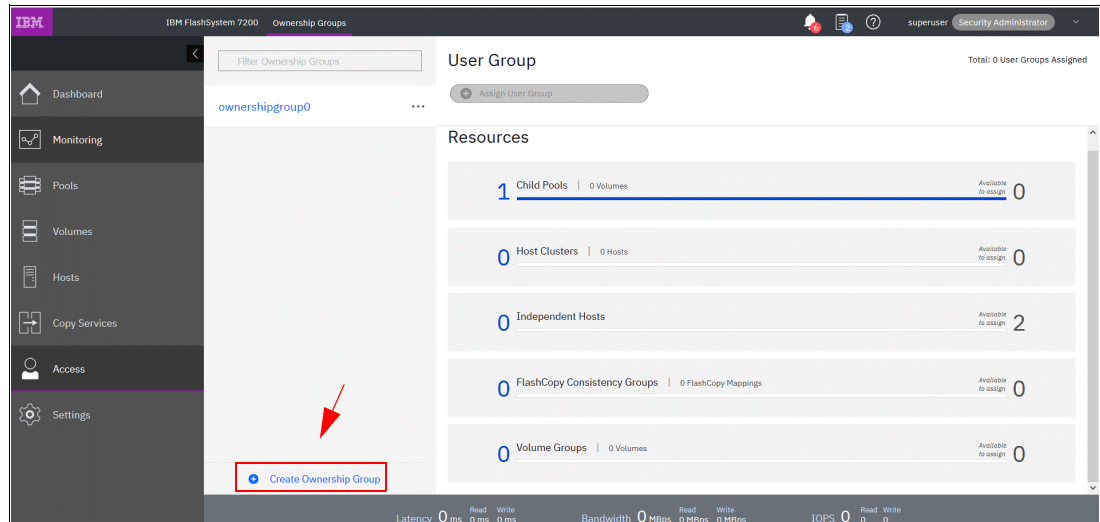


Figure 4-56 Ownership by Groups

To create an ownership group, select **Create Ownership Group**, as shown in Figure 4-56.

4.9.2 Users by groups

You can create local users who can access the system. These user types are defined based on the administrative privileges that they have on the system.

Local users must provide a password, Secure Shell (SSH) key, or both. Local users are authenticated through the authentication methods that are configured on the system. If the local user needs access to the management GUI, a password is needed for the user. If the user requires access to the CLI through SSH, a password or a valid SSH key file is necessary.

Local users must be part of a user group that is defined on the system. User groups define roles that authorize the users within that group to a specific set of operations on the system.

To define your user group in your system, select **Access** → **Users by Group**, as shown in Figure 4-57.

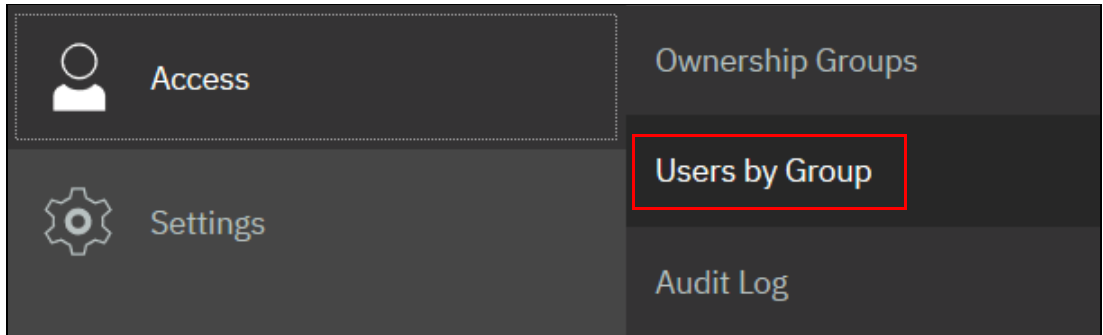


Figure 4-57 Accessing Users by Group

Select **Create User Group**, as shown in Figure 4-58.

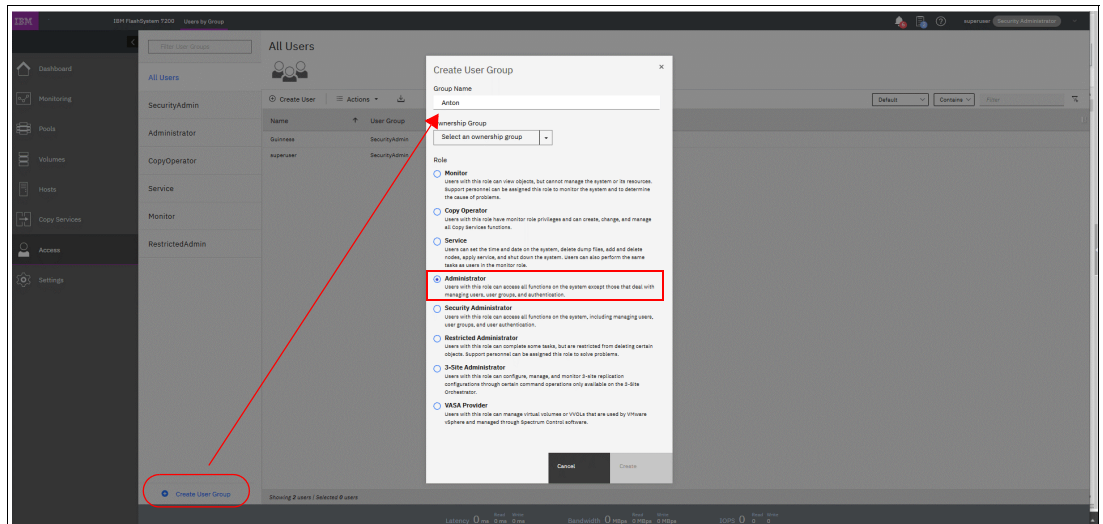


Figure 4-58 Defining a User Group

Figure 4-59 on page 191 shows the newly created User Group.

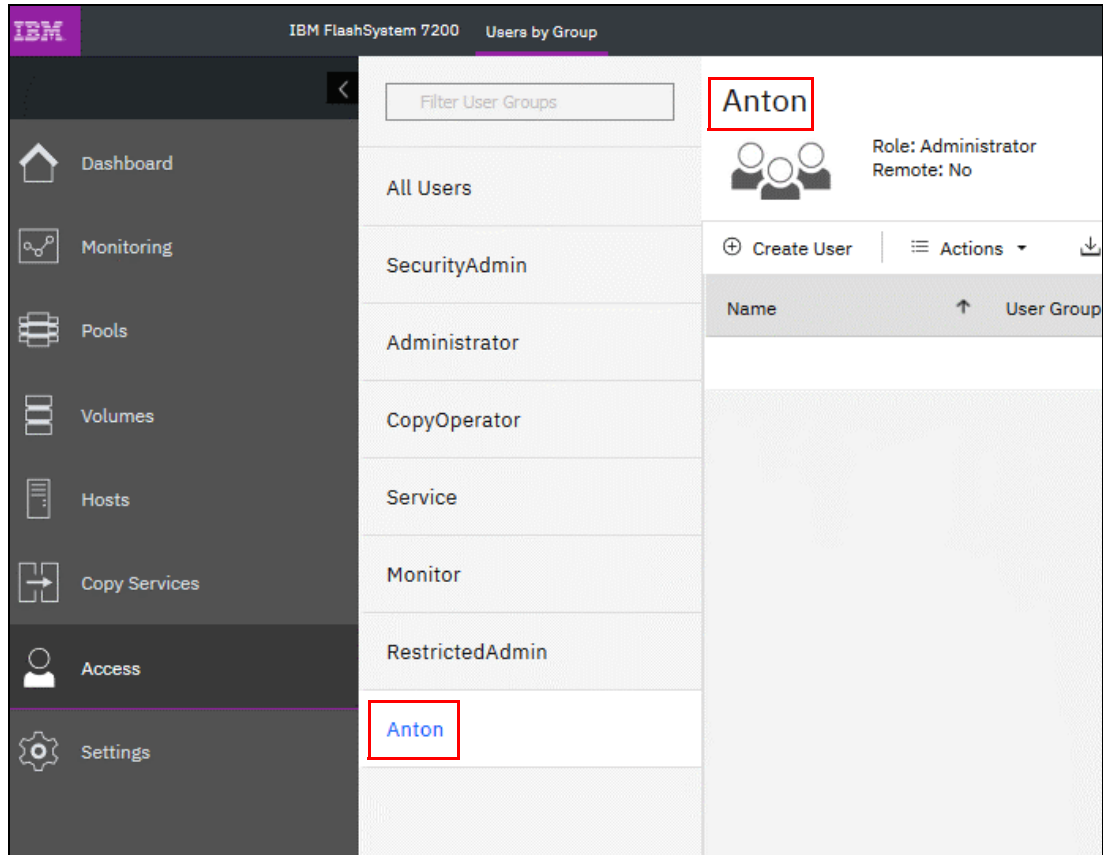


Figure 4-59 User Group

The following privileged user group roles exist in IBM Spectrum Virtualize:

► Monitor

These users can access all system viewing actions. Monitor role users cannot change the state of the system or the resources that the system manages. Monitor role users can access all information-related GUI functions and commands, back up configuration data, and change their own passwords.

► Copy Operator

These users can start and stop all existing FlashCopy, MM, and GM relationships. Copy Operator role users can run the system commands that Administrator role users can run that deal with FlashCopy, MM, and GM relationships.

► Service

These users can set the time and date on the system, delete dump files, add and delete nodes, apply service, and shut down the system. Users can also complete the same tasks as users in the monitor role.

► Administrator

These users can manage all functions of the system except for those functions that manage users, user groups, and authentication. Administrator role users can run the system commands that the Security Administrator role users can run from the CLI, except for commands that deal with users, user groups, and authentication.

► Security Administrator

These users can manage all functions of the system, including managing users, user groups, user authentication, and configuring encryption. Security Administrator role users can run any system commands from the CLI. However, they cannot run the **sa info** and **sa task** commands from the CLI. Only the superuser ID can run those commands.

► Restricted Administrator

These users can perform the same tasks and run most of the same commands as Administrator role users. However, users with the Restricted Administrator role are not authorized to run the **rmvdisk**, **rmvdiskhostmap**, **rmhost**, or **rmmdiskgrp** commands. Support personnel can be assigned this role to help resolve errors and fix problems.

► 3-Site Administrator

These users can configure, manage, and monitor 3-site replication configurations through certain command operations that are available only on the 3-Site Orchestrator. Before you can work with 3-Site Orchestrator, a user profile must be created.

► vSphere application programming interfaces (APIs) for Storage Awareness (VASA) Provider

Users with this role can manage virtual volumes or VMware vSphere Virtual Volume (VVOLs) that are used by VMware vSphere and managed through IBM Spectrum Control software.

Registering a user

After you define your group, you can register a user within this group by clicking **Create User** (see Figure 4-60).

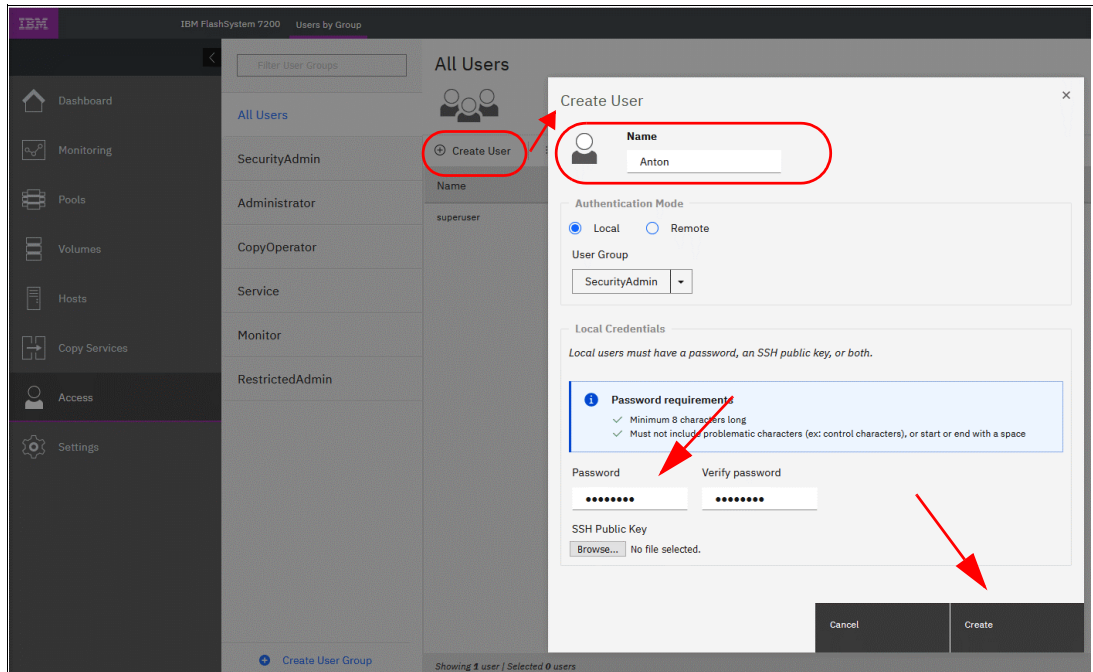


Figure 4-60 Registering a user account

Deleting a user

To remove a user account, right-click the user in the **All Users** list and select **Delete**, as shown in Figure 4-61.

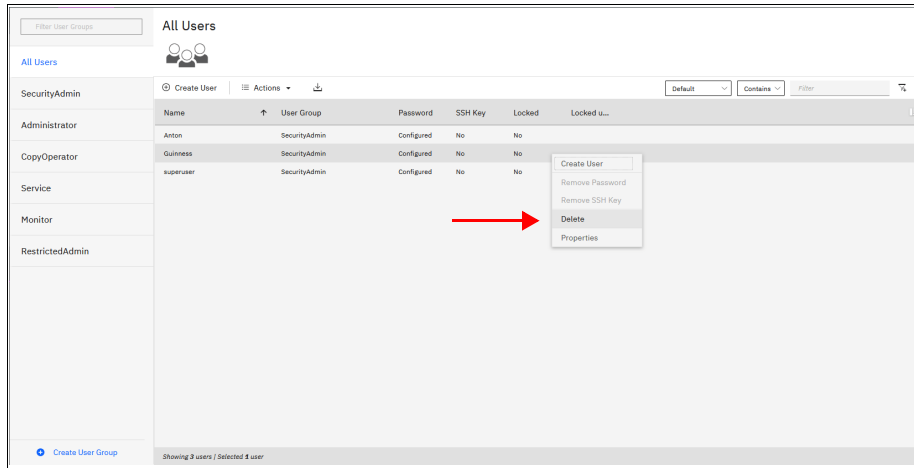


Figure 4-61 Deleting a user account

Attention: When you click **Delete**, the user account is directly deleted. No other confirmation request is presented.

Setting a new password

To set a new password for the user, right-click the user (or click **Actions**) and select **Properties**. In this window, you can either assign the user to a different group or reset their password (see Figure 4-62).

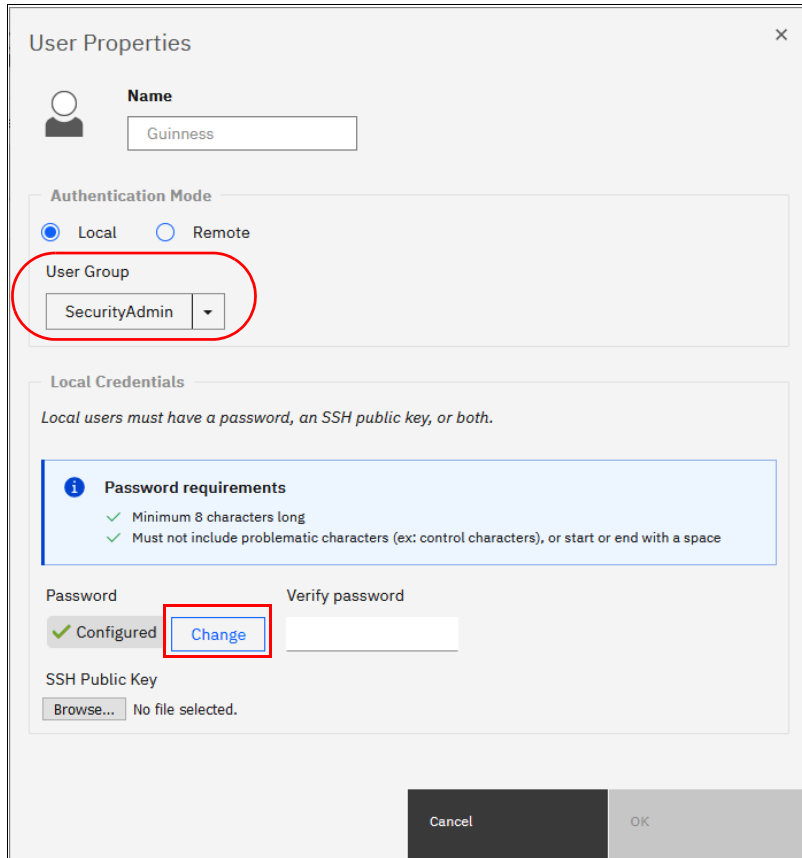


Figure 4-62 Setting a new password

4.9.3 Audit log

An *audit log* documents actions that are submitted through the management GUI or the CLI. You can use the audit log to monitor user activity on your system.

The audit log entries provide the following information:

- ▶ Time and date when the action or command was submitted.
- ▶ Name of the user who completed the action or command.
- ▶ IP address of the system where the action or command was submitted.
- ▶ Name of source and target node on which the command was submitted.
- ▶ Parameters that were submitted with the command, excluding confidential information.
- ▶ Results of the command or action that completed successfully.
- ▶ Sequence number and the object identifier that is associated with the command or action.

An example of the audit log is shown in Figure 4-63 on page 195.

Date and Time	User Name	Command	Object ID
10/11/2020 7:38:17 PM	superuser	svctask mkuser -gui -name Guinness -password #### -usergr...	2
10/11/2020 7:36:38 PM	superuser	svctask mkuser -gui 2	
10/11/2020 7:32:23 PM	superuser	svctask mkuser -gui -name Guinness -password #### -usergr...	2
10/11/2020 7:31:44 PM	superuser	svctask mkuser -gui -name Anton -password #### -usergrp 0	1
10/11/2020 6:48:00 PM	superuser	svctask mkownershipgroup -gui	
10/11/2020 6:38:10 PM	superuser	svctask chlogrp -gui -maintenance no 3	
10/11/2020 6:38:09 PM	superuser	svctask chlogrp -gui -maintenance no 1	
10/11/2020 6:38:09 PM	superuser	svctask chlogrp -gui -maintenance no 2	
10/11/2020 6:37:03 PM	superuser	svctask chlogrp -gui -maintenance no 0	
10/11/2020 6:35:58 PM	superuser	svctask chlogrp -gui -maintenance yes 0	
10/11/2020 5:00:02 PM	superuser	svctask chenclosureslot -gui -identify yes -slot 10 2	
10/11/2020 4:51:57 PM	superuser	svctask chenclosureslot -gui -identify no -slot 10 2	
10/11/2020 4:49:54 PM	superuser	svctask chenclosureslot -gui -identify yes -slot 10 2	
10/11/2020 12:28:49 PM	superuser	svctask chenclosurecanister -canister 2 -gui -identify yes 2	
10/11/2020 12:28:05 PM	superuser	svctask chenclosurecanister -canister 2 -gui -identify yes 2	
10/11/2020 1:00:16 AM	superuser	sataisk cpfiles -prefix /dumps/svc.config.cron*_7825WKP-2 -s...	
10/11/2020 1:00:03 AM	superuser	svctask detectmdisk	
10/10/2020 7:32:24 PM	superuser	svctask mkhost -fowppn 5009507680C11CD66-5009507680C12...	0
10/10/2020 7:32:24 PM	superuser	svctask chhost -gui -statuspolicy redundant 0	
10/10/2020 7:31:10 PM	superuser	svctask sendcloudcallhome -connectiontest -gui	
10/10/2020 7:29:47 PM	superuser	svctask mkdistributedarray -driveclass 0 -drivecount 10 -gui -l...	0

Figure 4-63 Audit log

The following commands are not documented in the audit log:

- ▶ **dumpconfig**
- ▶ **cpdumps**
- ▶ **finderr**
- ▶ **dumperrlog**

The following items are also not documented in the audit log:

- ▶ Commands that fail are not logged.
- ▶ A result code of 0 (success) or 1 (success in progress) is not logged.
- ▶ Result object ID of node type (for the **addnode** command) is not logged.
- ▶ Views are not logged.

Important: Failed commands are not recorded in the audit log. Commands that are triggered by IBM Support personnel are recorded with the flag **Challenge** because they use challenge-response authentication.

4.10 Settings

Use the Settings window to configure system options for notifications, security, IP addresses, and preferences that are related to display options in the management GUI (see Figure 4-64).

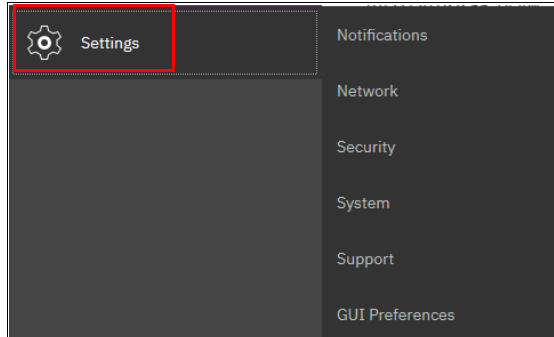


Figure 4-64 Settings menu

The following options are available for configuration from the **Settings** menu:

- ▶ **Notifications:** The system can use Simple Network Management Protocol (SNMP) traps, syslog messages, and Call Home emails to notify you and IBM Support Center when significant events are detected. Any combination of these notification methods can be used simultaneously.
- ▶ **Network:** Use the Network window to manage the management IP addresses for the system, service IP addresses for the nodes, and internet Small Computer Systems Interface (iSCSI) and FC configurations. The system must support FC or Fibre Channel over Ethernet (FCoE) connections to your storage area network (SAN).
- ▶ **Security:** Use the Security window to configure and manage remote authentication services.
- ▶ **System:** Use the **System** menu to manage overall system configuration options, such as licenses, updates, and date and time settings.
- ▶ **Support:** Use this option to configure and manage connections, and upload support packages to the support center.
- ▶ **GUI Preferences:** Configure the welcome message that appears after you log in, and refresh internals and GUI logout timeouts.

These options are described in more detail in the following sections.

4.10.1 Notifications

Your IBM Storage System can use SNMP traps, syslog messages, and Call Home email to notify you and the IBM Support Center when significant events are detected. Any combination of these notification methods can be used simultaneously.

Notifications are normally sent immediately after an event is raised. However, events can occur because of service actions that are performed. If a recommended service action is active, notifications about these events are sent only if the events are still unfixed when the service action completes.

SNMP notifications

SNMP is a standard protocol for managing networks and exchanging messages. The system can send SNMP messages that notify personnel about an event. You can use an SNMP manager to view the SNMP messages that are sent by your storage system.

To view the SNMP configuration, click the **Settings** icon and select **Notification** → **SNMP** (see Figure 4-65).

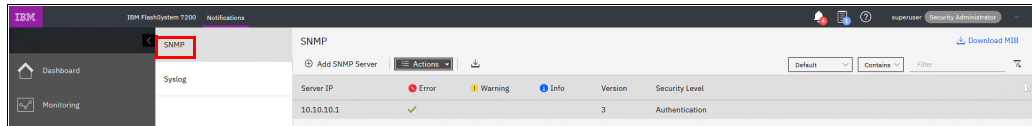


Figure 4-65 Setting the SNMP server and traps

In Figure 4-65, you can view and configure an SNMP server to receive various informational, error, or warning notifications by setting the following information:

- ▶ IP Address

The address for the SNMP server.

- ▶ Community

SNMP Community strings are used only by devices that support the SNMPv1 and SNMPv2c protocols. SNMPv3 uses username and password authentication, along with an encryption key. By convention, most SNMPv1 to v2c equipment ships from the factory with a read-only community string set to “public”.

- ▶ Server Port

The remote port (RPORT) number for the SNMP server. The RPORT number must be a value of 1 - 65535.

- ▶ Event Notifications

Consider the following points about event notifications:

- Select **Error** if you want the user to receive messages about problems, such as hardware failures, that must be resolved immediately.

Important: Browse to **Recommended Actions** to run the fix procedures on these notifications.

- Select **Warning** if you want the user to receive messages about problems and unexpected conditions. Investigate the cause immediately to determine any corrective action.

Important: Browse to **Recommended Actions** to run the fix procedures on these notifications.

- Select **Info** if you want the user to receive messages about expected events. No action is required for these events.

► Engine ID

The Engine ID is only used by SNMPv3 entities to uniquely identify them. An SNMP agent is considered an authoritative SNMP engine. Each SNMP agent maintains local information that is used in SNMPv3 message exchanges. The default SNMP Engine ID is composed of the enterprise number and the default Media Access Control (MAC) address.

To remove an SNMP server, click the minus sign (-). To add another SNMP server, click the plus sign (+).

► Security Name

The username must not exceed 31 characters.

► Authentication Protocol

- Uses an MD5 message digest algorithm in HMAC:
 - Directly provides data integrity checks.
 - Indirectly provides data origin authentication.
 - Uses a private key that is known by the sender and receiver.
 - 16-byte key.
 - 128-bit digest (truncates to 96 bits).
- SHA, an optional alternative algorithm.
- Loosely synchronized monotonically, which increases time indicator values to defend against certain message stream modification attacks.

► Privacy Protocol

A user-based privacy mechanism that is based on the following items:

- Data Encryption Standard (DES) Cipher Block Chaining (CBC) mode:
 - Provides data confidentiality.
 - Uses encryption.
 - Subject to export and use restrictions in many jurisdictions.
- Uses a 16-byte key (56-bit DES key, 8-byte DES initialization vector) that is known by the sender and receiver.
- Has multiple levels of compliance regarding DES due to problems that are associated with international use.
- Triple Data Encryption Standard (Triple DES).
- Advanced Encryption Standard (AES) (128, 192, and 256, bit keys).

Syslog notifications

The syslog protocol is a standard protocol for forwarding log messages from a sender to a receiver on an IP network. The IP network can be IPv4 or IPv6. The system can send syslog messages that notify personnel about an event. You can use the Syslog window to view the syslog messages that are sent by the system. To view the Syslog configuration, go to the System pane and click **Settings**, and select **Notification** → **Syslog** (see Figure 4-66 on page 199). A domain name server (DNS) server is required to use domain names in syslog.

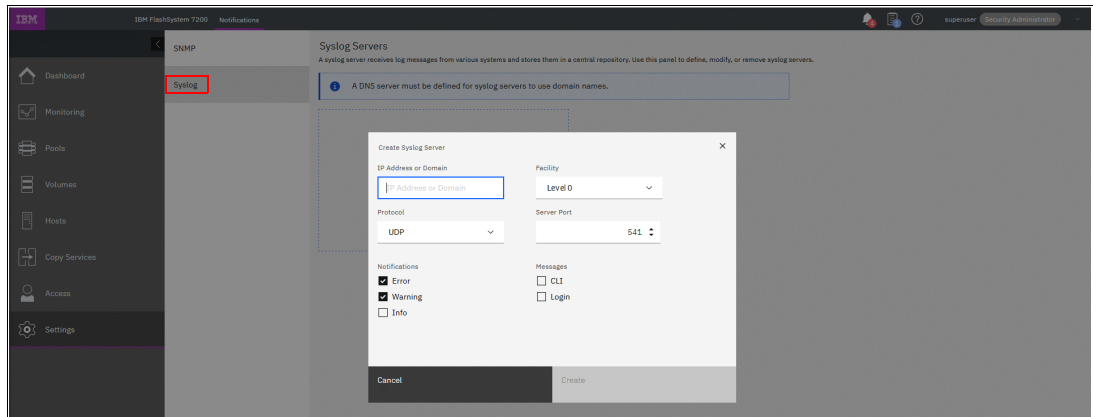


Figure 4-66 Setting the syslog messages

From this window, you can view and configure a syslog server to receive log messages from various systems and store them in a central repository by entering the following information:

- ▶ **IP Address**
The IP address for the syslog server.
- ▶ **Facility**
The facility determines the format for the syslog messages. The facility can be used to determine the source of the message.
- ▶ **Protocol of the transmission protocol**
Select **UDP** or **TCP**.
- ▶ **Port**
Port number of the syslog server.
- ▶ **Event Notifications**
 - Select **Error** if you want the user to receive messages about problems, such as hardware failures, that must be resolved immediately.

Important: Browse to **Recommended Actions** to run the fix procedures on these notifications.

- Select **Warning** if you want the user to receive messages about problems and unexpected conditions. Investigate the cause immediately to determine whether any corrective action is necessary.
- Select **Info** if you want the user to receive messages about expected events. No action is required for these events.
- ▶ **Message Format**
The message format depends on the facility. The system can transmit syslog messages in the following formats:
 - The concise message format provides standard details about the event.
 - The expanded format provides more details about the event.

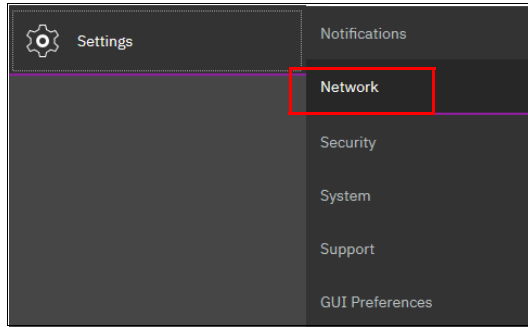


Figure 4-68 Accessing network information

Configuring the network

The procedure to set up and configure an IBM Storage System network interface is described in Chapter 3, “Initial configuration” on page 107.

Management IP addresses

To view the management IP addresses of IBM Spectrum Virtualize, select **Settings** → **Network**, and click **Management IP Addresses**. The GUI shows the management IP address by pointing to the network ports, as shown in Figure 4-69.

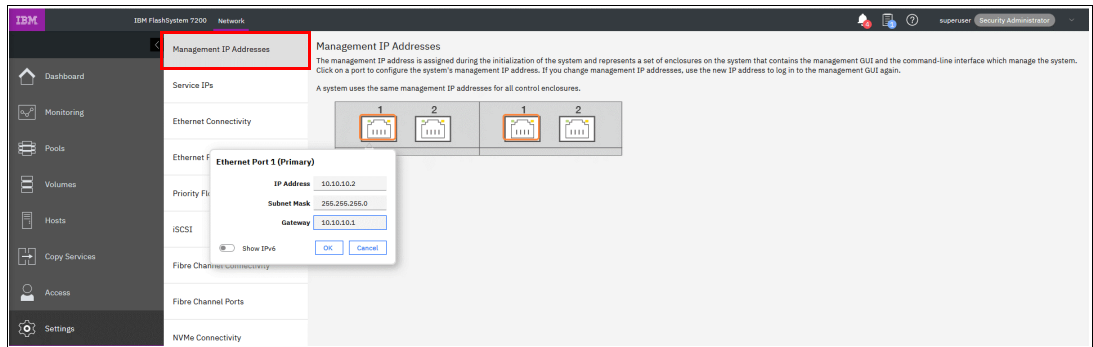


Figure 4-69 Viewing the management IP addresses

Service IP information

To view the Service IP information of your IBM Spectrum Virtualize installation, select **Settings** → **Network**, as shown in Figure 4-68. Click the **Service IPs** option to view the properties, as shown in Figure 4-70.

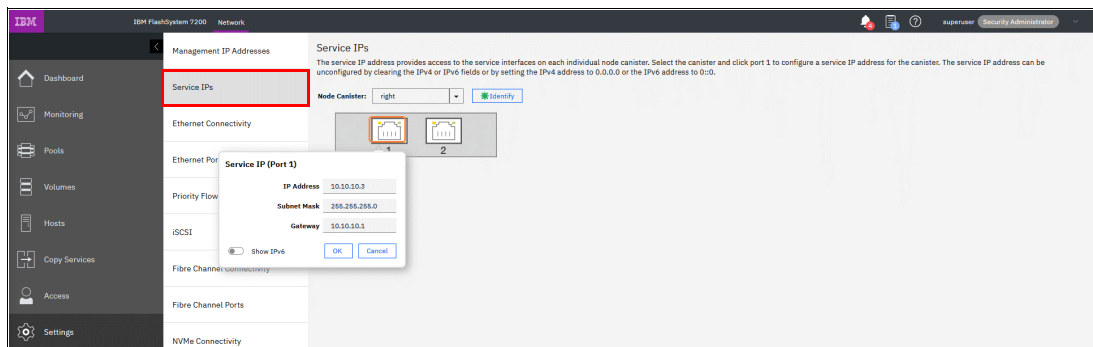


Figure 4-70 Viewing service IP addresses

The service IP address is commonly used to provide access to the network interfaces on each individual node of the control enclosure.

Instead of reaching the management IP address, the service IP address directly connects to each individual node canister for service operations. You can select a node canister of the control enclosure from the drop-down list and then click any of the ports that are shown in the GUI. The service IP address can be configured to support IPv4 or IPv6.

Ethernet Connectivity

Ethernet Connectivity displays the connectivity between nodes that are attached through the Ethernet network, as shown in Figure 4-71.

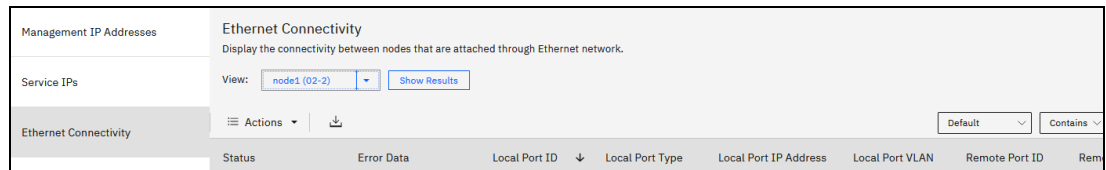


Figure 4-71 Ethernet Connectivity

Ethernet ports

Ethernet ports for each node are at the rear of the system and used to connect the system to hosts, external storage systems, and to other systems that are part of RC partnerships. Depending on the model of your system, supported connection types include FC, when the ports are FCoE-capable, iSCSI, and iSCSI Extensions for Remote Direct Memory Access (RDMA) (iSER). iSER connections use either the RDMA over Converged Ethernet (RoCE) protocol or the internet Wide Area RDMA Protocol (iWARP). The panel indicates whether a specific port is being used for a specific purpose and traffic.

You can modify how the port is used by selecting **Actions**. Select either **Modify VLAN**, **Modify IP Settings**, **Modify Remote Copy**, **Modify iSCSI Hosts**, **Modify Storage Ports**, or **Modify Maximum Transmission Unit** to change the use of the port. You can also display the login information for each host that is logged in to a selected node.

To display this information, select **Settings** → **Network** → **Ethernet Ports** and right-click the node and select **IP Login Information**. This information can be used to detect connectivity issues between the system and hosts and to improve the configuration of iSCSI host to optimize performance. Select **Ethernet Ports** for an overview from the menu, as shown in Figure 4-72. For planning, see Chapter 2, “Planning” on page 71.

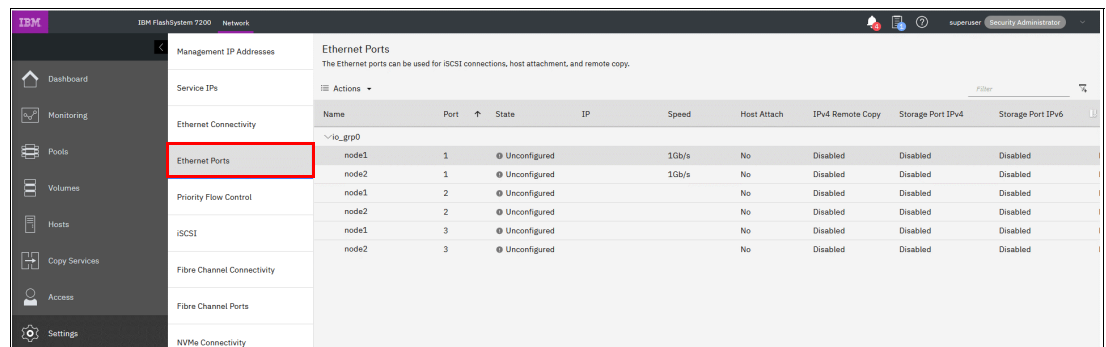


Figure 4-72 Ethernet Ports

Priority flow control

Priority flow control (PFC) is an Ethernet protocol that you can use to select the priority of different types of traffic within the network. With PFC, administrators can reduce network congestion by slowing or pausing certain classes of traffic on ports, thus providing better bandwidth for more important traffic. The system supports PFC on various supported Ethernet-based protocols on three types of traffic classes: system, host attachment, and storage traffic. You can configure a priority tag for each of these traffic classes. The priority tag can be any value 0 - 7. You can set identical or different priority tag values to all these traffic classes. You can also set bandwidth limits to ensure quality of service (QoS) for these traffic classes by using the Enhanced Transmission Selection (ETS) setting on the network. When you plan to configure PFC, follow these guidelines and examples.

To use PFC and ETS, ensure that the following tasks are completed:

- ▶ Ensure that ports support 10 Gb or higher bandwidth to use PFC settings.
- ▶ Configure a virtual local area network (VLAN) on the system to use PFC capabilities for the configured IP version.
- ▶ Ensure that the same VLAN settings are configured on the all entities, including all switches between the communicating end points.
- ▶ Configure the QoS values (priority tag values) for host attachment, storage, or system traffic by running the `chsystemethernet` command.
- ▶ To enable priority flow for host attachment traffic on a port, make sure that the host flag is set to yes on the configured IP on that port.
- ▶ To enable priority flow for storage traffic on a port, make sure that storage flag is set to yes on the configured IP on that port.
- ▶ On the switch, enable the Data Center Bridging Exchange (DCBx). DCBx enables switch and adapter ports to exchange parameters that describe traffic classes and PFC capabilities. For these steps, check your switch documentation for details.
- ▶ For each supported traffic class, configure the same priority tag on the switch. For example, if you plan to have a priority tag setting of 3 for storage traffic, ensure that the priority is also set to 3 on the switch for that traffic type.
- ▶ If you are planning on using the same port for different types of traffic, ensure that the ETS settings are configured on the network.

4.10.3 Using the management GUI

To set PFC on the system, complete the following steps:

1. In the management GUI, select **Settings** → **Network** → **Priority Flow Control**, as shown in Figure 4-73.

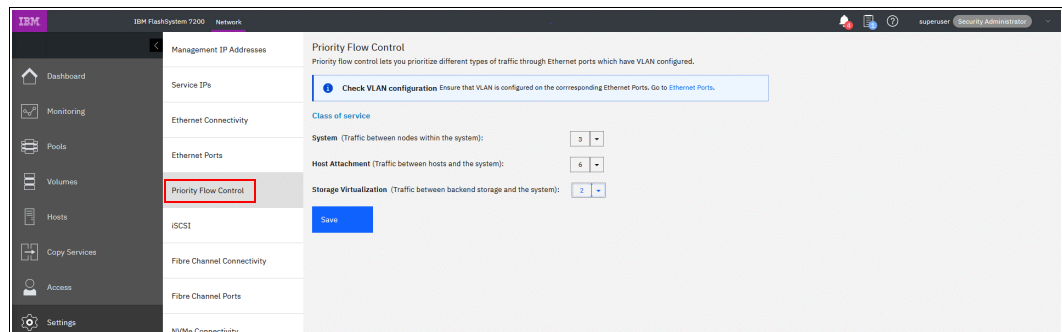


Figure 4-73 Priority flow control

2. For each of following classes of service, select the priority setting for that traffic type:
 - System

Set a value 0 - 7 for the system traffic, which includes communication between nodes within the system. The system priority tag is supported on iSCSI connections and on systems that support RDMA over Ethernet connections between nodes. Ensure that you set the same priority tag on the switch to use PFC capabilities.
 - Host attachment

Set the priority tag 0 - 7 for system to host traffic. The host attachment priority tag is supported on iSCSI connections and on systems that support RDMA over Ethernet connections. Ensure that you set the same priority tag on the switch to use PFC capabilities.
 - Storage virtualization

Set the priority tag 0 - 7 for system to external storage traffic. The storage virtualization priority tag is supported on storage traffic over iSCSI connections. Ensure that you set the same priority tag on the switch to use PFC capabilities.

Make sure that IP is configured with VLAN.

iSCSI information

From the iSCSI pane in the **Settings** menu, you can display and configure parameters for the system to connect to iSCSI-attached hosts, as shown in Figure 4-74.

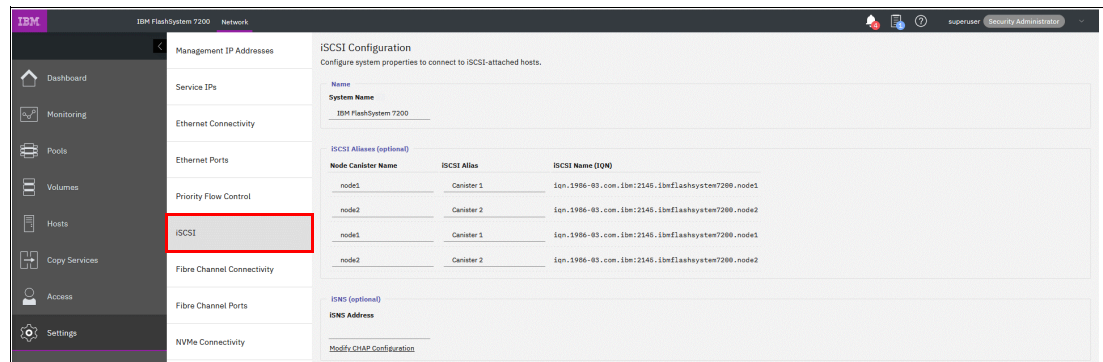


Figure 4-74 iSCSI Configuration pane

The following parameters can be updated:

► **System Name**

It is important to set the system name correctly because it is part of the iSCSI Qualified Name (IQN) for the node.

Important: If you change the name of the system after iSCSI is configured, you might need to reconfigure the iSCSI hosts.

To change the system name, click the system name and specify the new name.

System name: You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The name can be 1 - 63 characters.

- ▶ iSCSI aliases (optional)

An *iSCSI alias* is a user-defined name that identifies the node to the host. Complete the following steps to change an iSCSI alias:

- Click an iSCSI alias.
- Specify a name for it.

Each node has a unique iSCSI name that is associated with two IP addresses. After the host starts the iSCSI connection to a target node, this IQN from the target node is visible in the iSCSI configuration tool on the host.

- ▶ Internet Storage Name Service (iSNS) and Challenge Handshake Authentication Protocol (CHAP)

You can specify the IP address for the iSNS. Host systems use the iSNS server to manage iSCSI targets and for iSCSI discovery.

You can also enable CHAP to authenticate the system and iSCSI-attached hosts with the specified shared secret.

The CHAP secret is the authentication method that is used to restrict access for other iSCSI hosts that use the same connection. You can set the CHAP for the whole system under the system properties or for each host definition. The CHAP must be identical on the server and the system and host definition. You can create an iSCSI host definition without using CHAP.

Fibre Channel information

As shown in Figure 4-75, you can use the Fibre Channel Connectivity window to display the Fibre Channel connection (IBM FICON®) between nodes and other storage systems and hosts that attach through the FC network. You can filter by selecting one of the following fields:

- ▶ All nodes, storage systems, and hosts
- ▶ Systems
- ▶ Nodes
- ▶ Storage systems
- ▶ Hosts

You can view Fibre Channel Connectivity as shown in Figure 4-75.

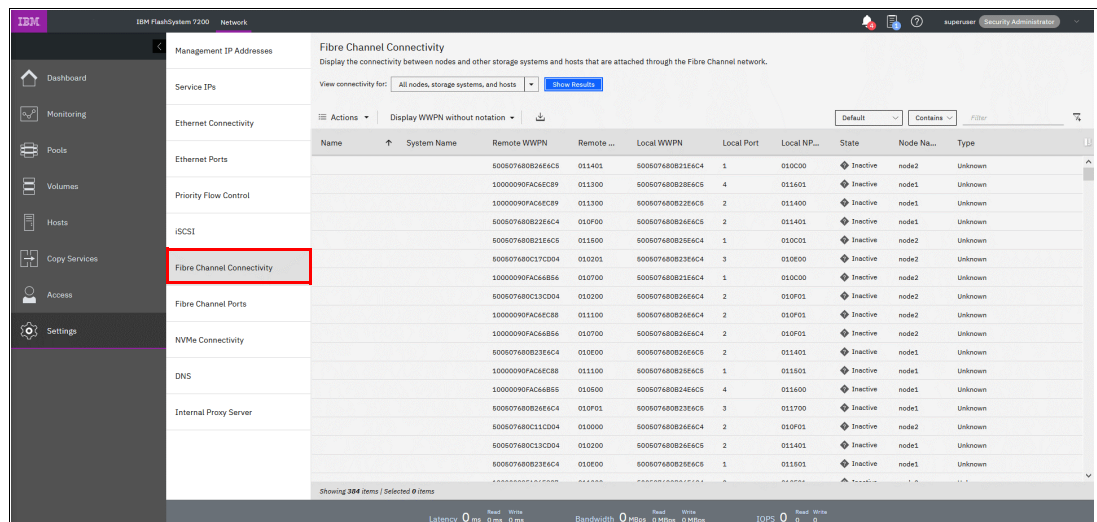


Figure 4-75 Fibre Channel Connectivity

In the Fibre Channel Ports window, you can use this view to display how the FC port is configured across all control node canisters in the system. This view helps, for example, to determine which other clusters and hosts the port may communicate with, and which ports are virtualized. “No” indicates that this port cannot be online on any node other than the owning node (see Figure 4-76).

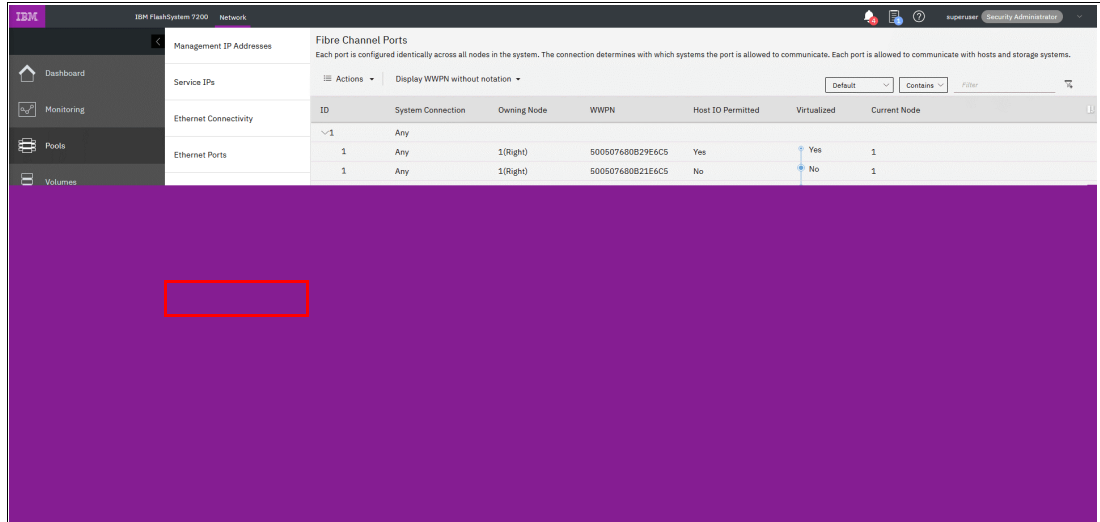


Figure 4-76 Viewing Fibre Channel Port properties

Non-Volatile Memory Express connectivity

A Non-Volatile Memory Express (NVMe) over FC host can be attached to the system. For other specific information about Non-Volatile Memory Express over Fibre Channel (FC-NVMe), such as interoperability requirements, see [V8.4.0.x Configuration Limits and Restrictions](#).

If your system supports an FC-NVMe connection between nodes and hosts, you can display details about each side of the connection. To display node details, select the node from the drop-down menu and select **Show Results**. You can also display the host details for the connection or for all hosts and nodes. Use this window to troubleshoot issues between nodes and hosts that use FC-NVMe connections.

For these connections, the Status column displays the current state of the connection. The following states for the connection are possible:

- ▶ **Active**
Indicates that the connection between the node and host is being used.
- ▶ **Inactive**
Indicates that the connection between the node and host is configured, but no FC-NVMe operations occurred in the last 5 minutes. Since the system sends periodic heartbeat message to keep the connection open between the node and the host, it is unusual to see an inactive state for the connection. However, it can take up to 5 minutes for the state to change from inactive to active. If the inactive state remains beyond the 5-minute refresh interval, it can indicate a connection problem between the host and the node. If a connection problem persists between the host and the node, a reduced node login count or the status of the host indicates it is degraded, which you can view by selecting **Hosts** → **Ports by Host** in the management GUI. Verify these values in the management GUI, and view the messages by selecting **Monitoring** → **Events**.

Figure 4-77 shows the NVMe Connectivity menu.

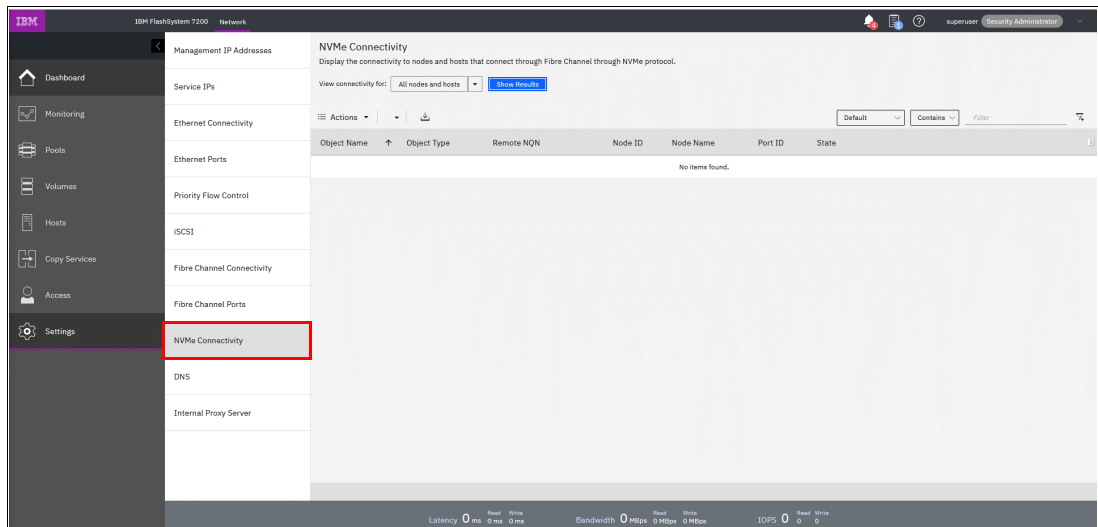


Figure 4-77 NVMe Connectivity window

Consider FC-NVMe target limits when you plan and configure the hosts. Include the following points in your plan:

- ▶ An NVMe host can connect to four NVMe controllers on each system node. The maximum per node is four with an extra four in failover.
- ▶ Zone up to four ports in a single host to detect up to four ports on a node. To allow failover and avoid outages, zone the same or extra host ports to detect an extra four ports on the second node in the I/O group.
- ▶ A single I/O group can contain up to 256 FC-NVMe I/O controllers. The maximum number of I/O controllers per node is 128 plus an extra 128 in failover. Zone a total maximum of 16 hosts to detect a single I/O group. Also, consider that a single system target port may have up to 16 NVMe I/O controllers.

When you install and configure attachments between the system and a host that runs the Linux operating system (OS), follow specific guidelines. For more information about these guidelines, see [Linux specific guidelines](#).

Domain name server

IBM Spectrum Virtualize allows DNS entries to be manually set up in the system. The information about the DNS helps the system to access the DNS and resolve the names of the computer resources that are in the external network.

To view and configure DNS information in IBM Spectrum Virtualize, complete the following steps:

1. In the left pane, click the **DNS** icon and enter the **IP address** and **Name** of each DNS. IBM Spectrum Virtualize supports up two DNSs for IPv4 or IPv6 (see Figure 4-78).

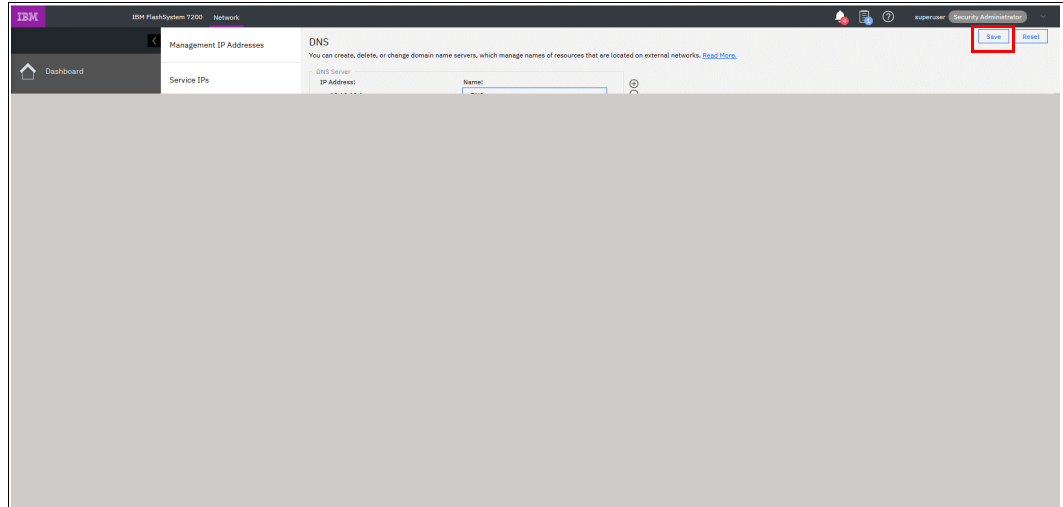


Figure 4-78 DNS information

2. Click **Save** after you finish entering the DNS information.

Internal Proxy Server

You can configure an internal proxy server to manage incoming and outgoing connections to the system, as shown in Figure 4-79.

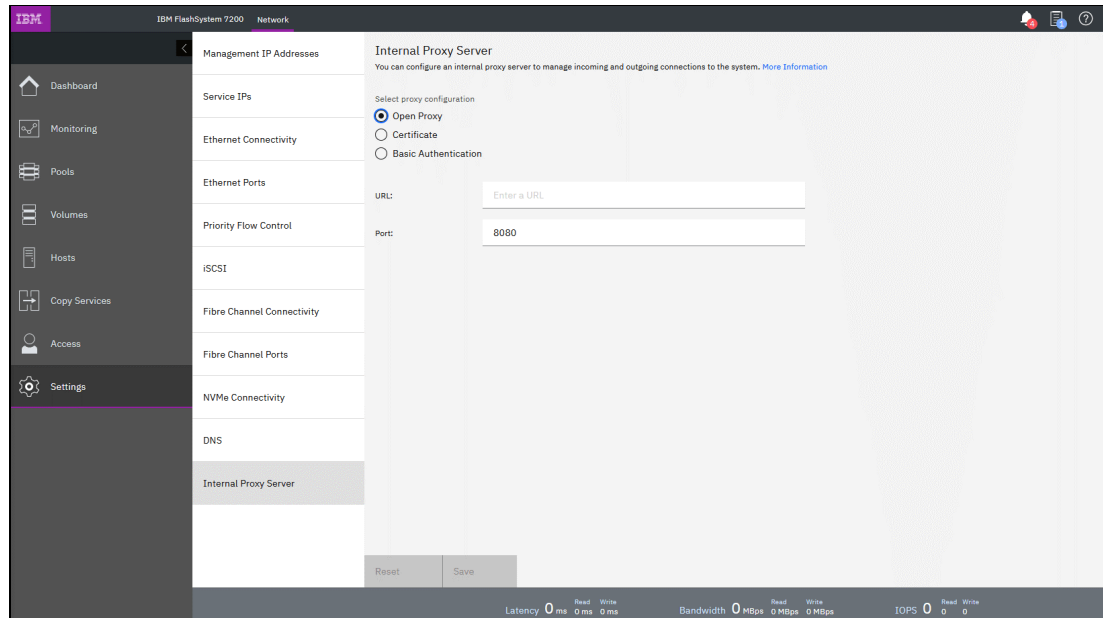


Figure 4-79 Internal Proxy Server

4.10.4 Security

Use the Security option from the **Settings** menu (as shown in Figure 4-80) to view and change security settings, authenticate users, and manage secure connections.

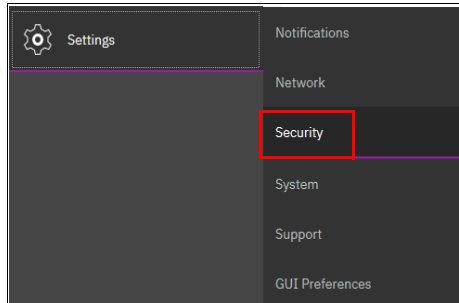


Figure 4-80 Security menu

Remote Authentication

In the Remote Authentication pane, you can configure remote authentication with LDAP, as shown in Figure 4-81. By default, the system has local authentication enabled. When you configure remote authentication, you do not need to configure users on the system or assign more passwords. Instead, you can use your passwords and user groups that are defined on the remote service to simplify user management and access, enforce password policies more efficiently, and separate user management from storage management.

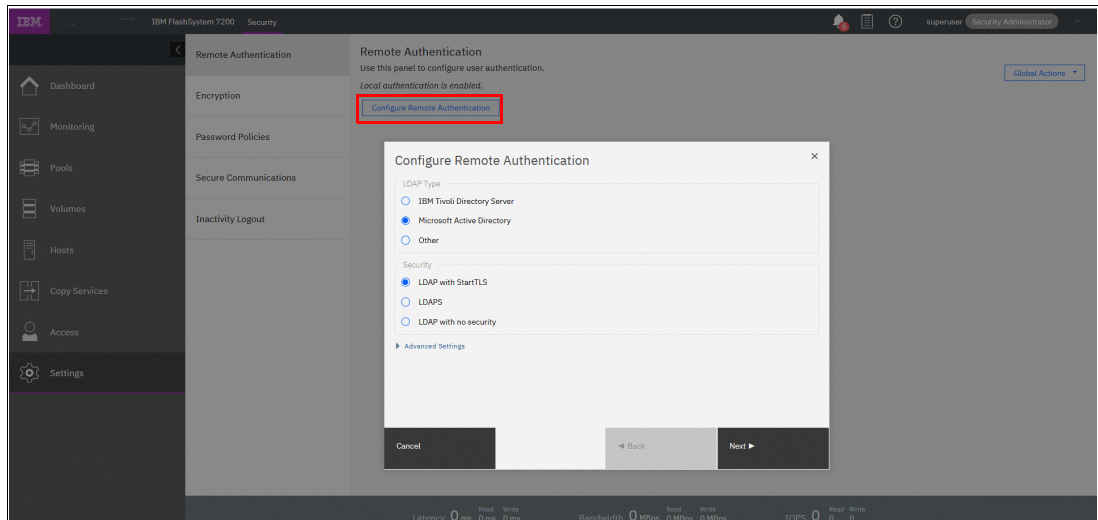


Figure 4-81 Configuring Remote Authentication

For more information about how to configure remote login, see [IBM Documentation](#).

Encryption

As shown in Figure 4-82, you can enable or disable the encryption function on an IBM Storage System. In our example, encryption is already enabled. For more information, see Chapter 12, “Encryption” on page 735.

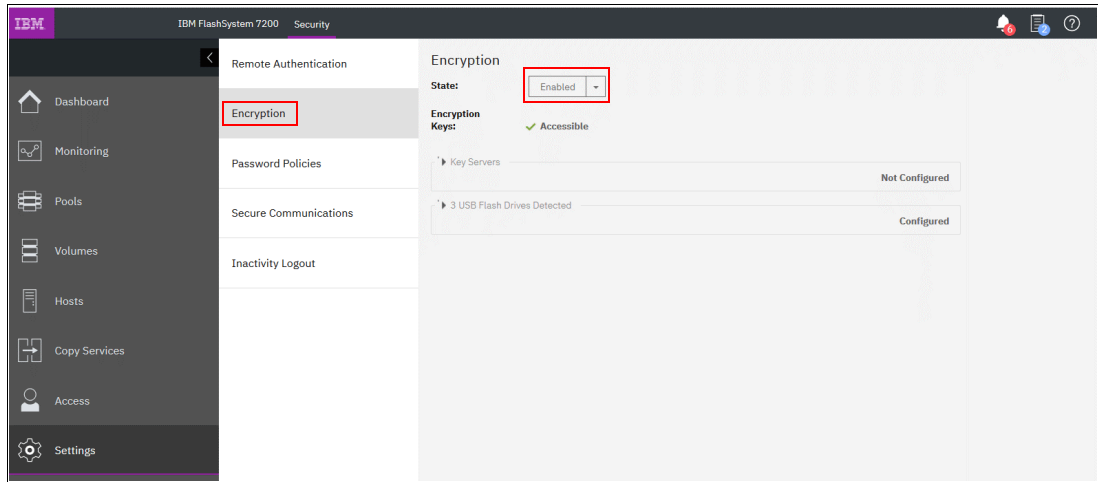


Figure 4-82 Enable Encryption

Password Policies

In this window, you can define policies for password management and expiration, as shown in Figure 4-83.

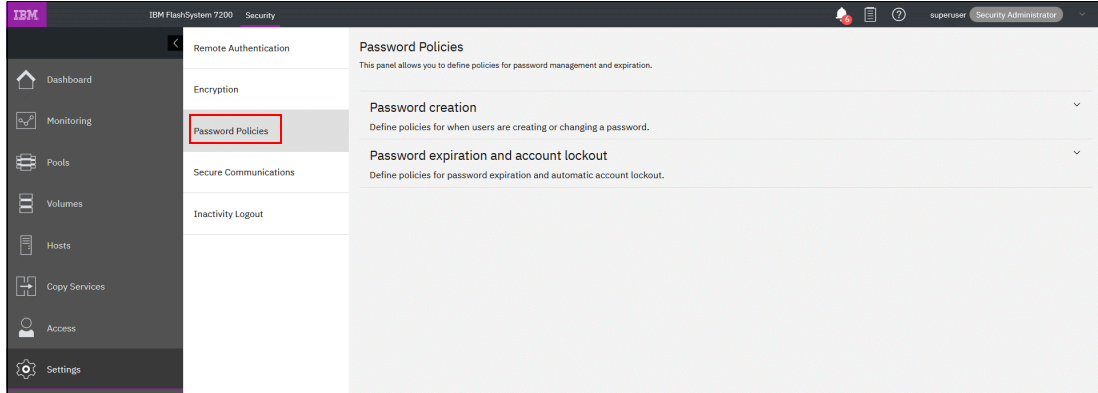


Figure 4-83 Password Policies

With password policy support, system administrators can set security requirements that are related to password creation and expiration, timeout for inactivity, and actions after failed logon attempts. Password policy support allows administrators to set security rules that are based on their organization's security guidelines and restrictions.

The system supports the following password and security-related rules:

► Password creation rules:

An administrator can set and manage the following rules for all passwords that are created on the system:

- Specify password length requirements for all users.
- Require passwords to use uppercase and lowercase characters.

- Require passwords to contain special characters.
 - Prevent users from reusing recent passwords.
 - Require users to change their password on next login under any of these conditions:
 - Their password expired.
 - An administrator created new accounts with temporary passwords.
- Password expiration and account locking rules:
- The administrator can create the following rules for password expiration:
- Set the password expiration limit.
 - Set a password to expire immediately.
 - Set the number of failed login attempts before the account is locked.
 - Set the time for locked accounts.
 - Automatic logout for inactivity.
 - Locking superuser account access.

Note: Systems that support a dedicated technician port can lock the superuser account. The superuser account is the default user that can complete installation, initial configuration, and other service-related actions on the system. If the superuser account is locked, service tasks cannot be completed.

Secure Communications

To enable or manage secure communications, select the **Secure Communications** window, as shown in Figure 4-84. Before you create a request for either type of certificate, ensure that your current browser does not have restrictions about the type of keys that are used for certificates.

Some browsers limit the use of specific key-types for security and compatibility issues. Select **Update Certificate** to add new certificate details, including certificates that were created and signed by a third-party certificate authority (CA).

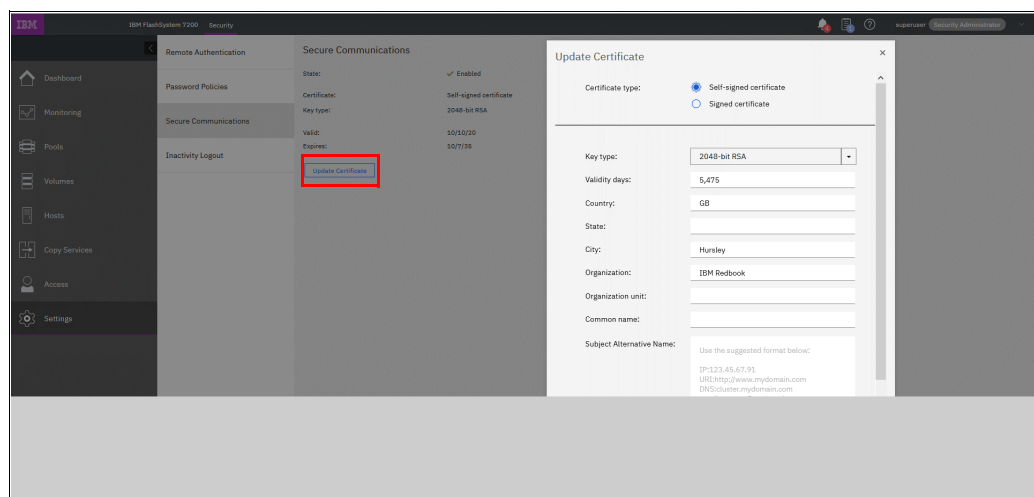


Figure 4-84 Configuring Secure Communications

Inactivity Logout

In this window, you set the inactivity time that is allowed before the system logs out users, as shown in Figure 4-85.

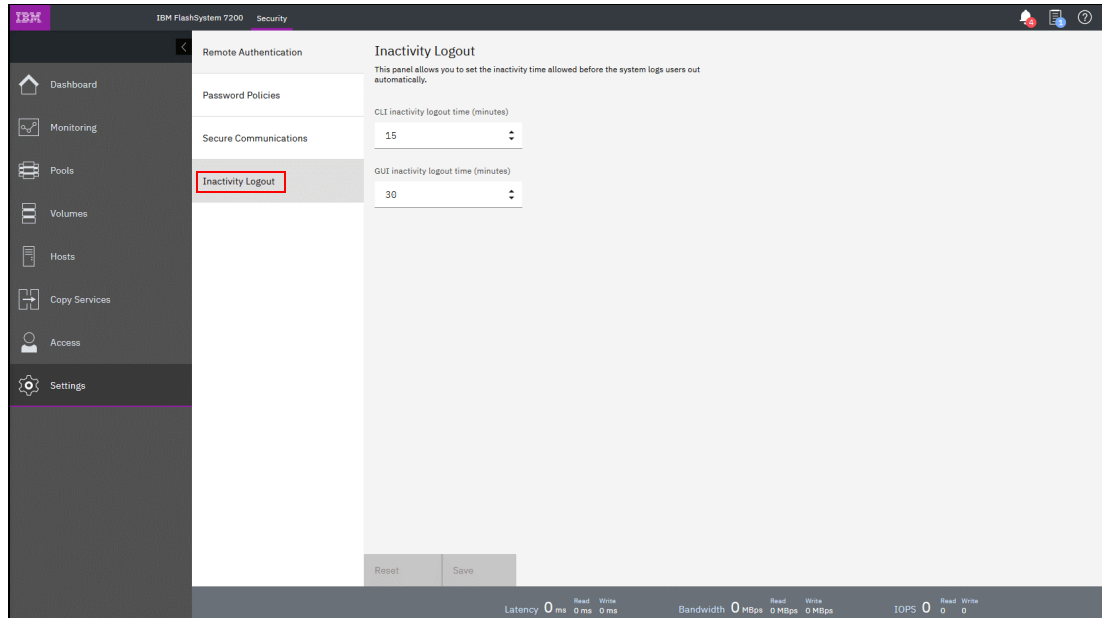


Figure 4-85 Inactivity Logout

4.10.5 System menus

Click the **System** option from the **Settings** menu (see Figure 4-86) to view and change the date and time settings, work with licensing options, download configuration settings, work with VVOLs and IP Quorum, or download software upgrade packages.

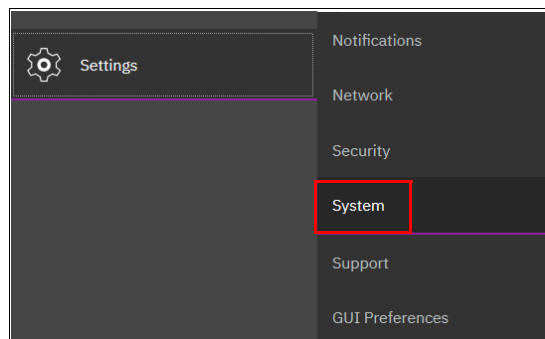


Figure 4-86 System option

Date and time

To view or configure the date and time settings, complete the following steps:

1. From the main System window, click **Settings** and click **System**.
2. In the left column, select **Date and Time**, as shown in Figure 4-87 on page 213.

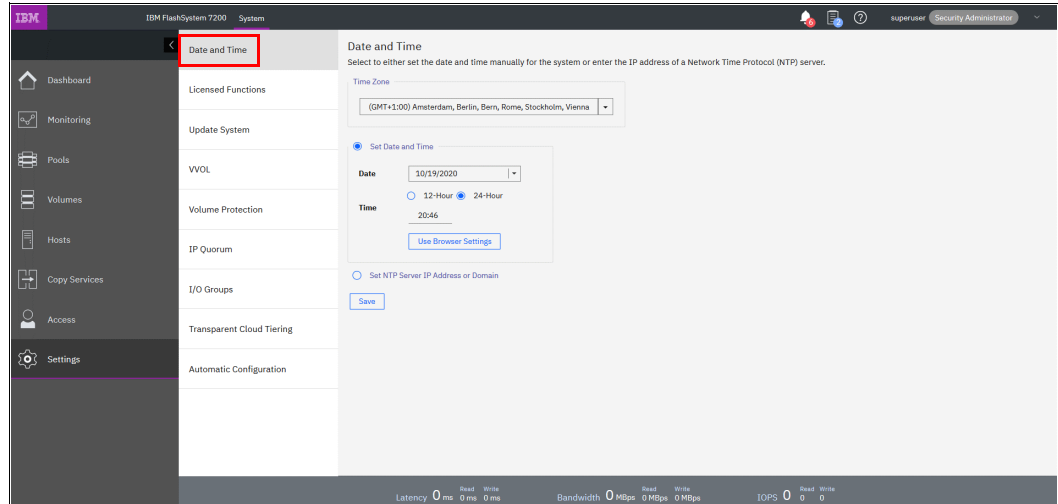


Figure 4-87 Date and Time window

3. From this window, you can modify the following information:

- Time zone

Select a time zone for your system by using the drop-down list.

- Date and time

The following options are available:

- If you are not using a Network Time Protocol (NTP) server, select **Set Date and Time**, and then manually enter the date and time for your system, as shown in Figure 4-88. You can click **Use Browser Settings** to automatically adjust the date and time of your system with your local workstation date and time.

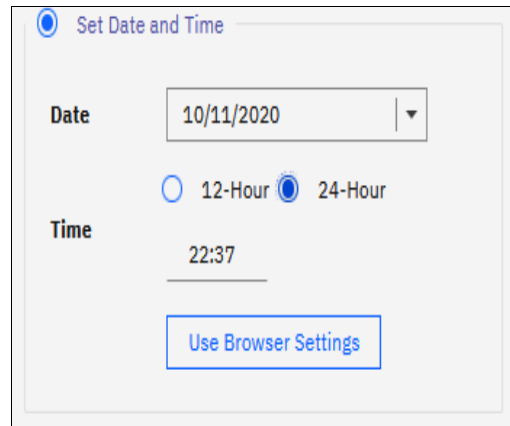


Figure 4-88 Set Date and Time window

- If you are using an NTP server, select **Set NTP Server IP Address**, and then enter the IP address of the NTP server, as shown in Figure 4-89.

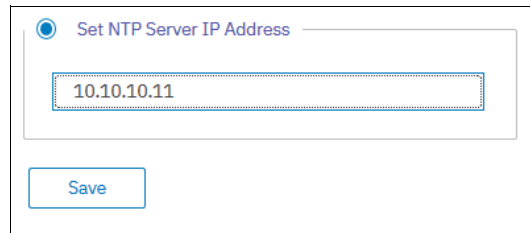


Figure 4-89 Set NTP Server IP Address window

4. Click **Save**.

Licensed Functions

The base license that is provided with your system includes the use of its basic functions. However, the extra licenses can be purchased to expand the capabilities of your system. Administrators are responsible for purchasing extra licenses and configuring the systems within the license agreement, which includes configuring the settings of each licensed function on the system.

Depending on the platform, different license schemes may be used:

- ▶ The IBM FlashSystem 5010 and IBM FlashSystem 5030 system licenses are Licensed Internal Code (LIC). All licenses are controller-based.
- ▶ The IBM FlashSystem 5100 system uses enclosure-based licensing, which allows the use of certain licensed functions that are based on the number of enclosures that are indicated in the license.
- ▶ IBM FlashSystem 7200, IBM FlashSystem 9100, and IBM FlashSystem 9200 systems use differential licensing for external virtualization, and capacity-based licensing for other functions.

Differential licensing charges different rates for different types of virtualized storage, which provides cost-effective management of capacity across multiple tiers of storage. It is based on the number of storage capacity units (SCUs) that are purchased.

Each SCU corresponds to a different amount of usable capacity based on the type of storage.

Table 4-1 shows the different storage types and the associated SCU ratios.

Table 4-1 SCU ratio per storage type

License	Drive classes	SCU ratio
Storage-class memory (SCM)	SCM devices	One SCU equates to 1 TiB of usable Category 1 storage.
Flash	All flash devices, other than SCM drives	One SCU equates to 1 TiB of usable Category 1 storage.

License	Drive classes	SCU ratio
Enterprise	10 K or 15 K RPM drives	One SCU equates to 1.18 TiB of usable Category 2 storage.
Nearline (NL)	NL Serial Advanced Technology Attachment (SATA) drives	One SCU equates to 4.00 TiB of usable Category 3 storage.

License settings are initially entered in to a system initialization wizard. They can be changed later.

To view or configure the licensing settings, complete the following steps:

1. From the main Settings window, click **Settings** and select **System**.
2. In the left column, click **Licensed Functions**. The example in Figure 4-90 shows the License Functions window of an IBM FlashSystem 9100 system, which uses differential licensing for External Virtualization.

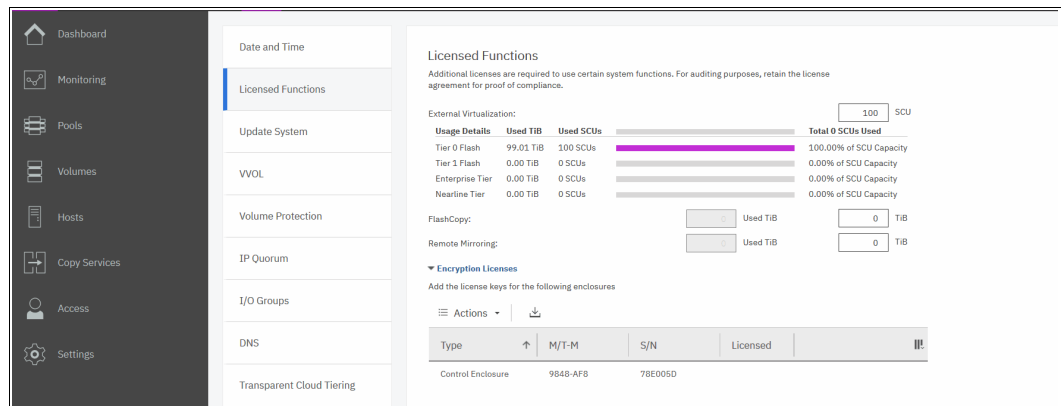


Figure 4-90 Licensing window

3. In the Licensed Functions window, you can view or set the licensing options for the IBM Storage System for the following elements:

- External Virtualization

You can enter the number of SCU units that are licensed for External Virtualization.

When monitoring External Virtualization license usage, consider the following items:

- The license accounts for usable MDisk capacity. For example, one SCU is used for one 1 TB Tier0 MDisk when it is assigned to a storage pool, independently of the amount of the actual data written on the MDisk.
- SCU is used for complete and incomplete chunks of MDisk capacity. For example, if a combined capacity of all NL Tier MDisk in your system is 5 TB, two SCUs are needed: one SCU for 4 TB of NL storage, and one SCU for another 4 TB (even if 1 TB is used).
- If your system uses enclosure-based licensing, specify the number of enclosures of external storage systems that are attached to your IBM Storage System. Data can be migrated from storage systems to your systems that use the External Virtualization function within 90 days of purchase of the system without purchase of a license. After 90 days, any ongoing use of the External Virtualization function requires a license for each enclosure in each external system.

- FlashCopy (if required on the platform)

The FlashCopy function copies the contents of a source volume to a target volume. It is also used to create cloud snapshots of volumes in systems that have TCT enabled.

FlashCopy can be licensed in terabytes (TB). In this case, the used capacity for FlashCopy mappings is the sum of all of the volumes that are the source volumes of a FlashCopy mapping and volumes with cloud snapshots.

If licensed in enclosures, FlashCopy can be used on a total number of internal enclosures and virtualized (external) enclosures.

- Remote mirroring

The remote mirroring function configures a relationship between two volumes. This function mirrors updates that are made to one volume to another volume. The volumes can be in the same system or on two different systems.

If a remote mirroring function is licensed per enclosure, you can use the remote mirroring functions on the total number of enclosures that are licensed. The total number of enclosures must include the enclosures on external storage systems that are licensed for virtualization and the number of control and expansion enclosures that are part of your local system.

If licensed by capacity, the function specifies the amount of data that can be replicated. The used capacity for remote mirroring is the sum of the capacities of all the volumes that are in an MM or GM relationship. Both master and auxiliary volumes are counted.

The license settings apply only to the system on which you are configuring license settings. For RC partnerships (includes HyperSwap), a license is also required on any remote systems that are in the partnership.

- Compression (if required on a platform)

With the compression function, data is compressed as it is written to disk, which saves extra capacity for the system. If a compression license is not included in a base license for your platform, it must be purchased separately for each enclosure.

- Encryption license

In addition to these enclosure-based licensed functions, the system also supports encryption through a key-based license. Key-based licensing requires an authorization code to activate encryption on the system. Only certain models of the control enclosures support encryption.

During initial setup, you can select to activate the license with the authorization code. The authorization code is sent with the licensed function authorization documents that you receive after purchasing the license. These documents contain the authorization codes that are required to obtain keys for the encryption function that you purchased for your system.

Encryption is activated on a per system basis, and an active license is required for each control enclosure that uses encryption. During system setup, the system detects any SAS attached enclosures and applies the license to these enclosures. If control enclosures are added and require encryption, more encryption licenses must be purchased and activated.

Note: To monitor license usage, run the `lslicense` CLI command, as described in [IBM Documentation](#).

Updating your storage system

For more information about the update procedure that uses the GUI, see Chapter 13, “Reliability, availability, and serviceability, monitoring and logging, and troubleshooting” on page 793.

VMware vSphere Virtual Volumes

IBM Spectrum Virtualize can manage VVOLs directly in cooperation with VMware. It enables VMware virtual machines (VMs) to get the assigned disk capacity directly rather than from the Elastic Sky X Integrated (ESXi) data store. This technique enables storage administrators to control the appropriate usage of storage capacity, and to enable enhanced features of storage virtualization directly to the VM (such as replication, thin-provisioning, compression, and encryption).

VVOL management is enabled in the System section, as shown in Figure 4-91. The NTP server must be configured before enabling VVOL management. As a best practice, use the same NTP server for ESXi and your system.

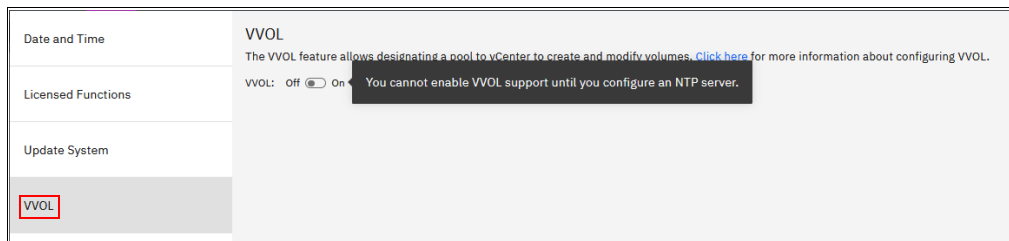


Figure 4-91 Enabling VVOLs management

Restriction: You cannot enable VVOL support until the NTP server is configured in the system.

For more information about VVOLs, see the following publications:

- ▶ *Quick-start Guide to Configuring VMware Virtual Volumes for Systems Powered by IBM Spectrum Virtualize*, REDP-5321
- ▶ *Configuring VMware Virtual Volumes for Systems Powered by IBM Spectrum Virtualize*, SG24-8328

Volume protection

Volume protection prevents active volumes or host mappings from being deleted inadvertently if the system detects recent I/O activity.

Note: This global setting is enabled by default on new systems. You can either set this value to apply to all volumes that are configured on your system, or control whether the system-level volume protection is enabled or disabled on specific pools.

To prevent an active volume from being deleted unintentionally, administrators can use the system-wide setting to enable volume protection. They can also specify a period that the volume must be idle before it can be deleted. If volume protection is enabled and the period is not expired, the volume deletion fails even if the **-force** parameter is used.

Note: The system-wide volume protection and the pool-level protection must both be enabled for protection to be active on a pool. The pool-level protection depends on the system-level setting to ensure that protection is applied consistently for volumes within that pool.

If system-level protection is enabled but pool-level protection is not enabled, any volumes in the pool can be deleted even when the setting is configured at the system level. When you delete a volume, the system verifies whether it is a part of a host mapping, FlashCopy mapping, or RC relationship. For a volume that contains these dependencies, the volume cannot be deleted unless the **-force** parameter is specified on the corresponding remove commands. However, the **-force** parameter does not delete a volume if it has recent I/O activity and volume protection is enabled. The **-force** parameter overrides the volume dependencies, not the volume protection setting.

The following actions are affected by this setting:

- ▶ Deleting a volume
- ▶ Deleting a volume copy
- ▶ Deleting a host or a host cluster mapping
- ▶ Deleting a storage pool
- ▶ Deleting a host from an I/O group
- ▶ Deleting a host or host cluster
- ▶ Deleting a defined host port
- ▶ Creating an RC relationship

For more information, see Figure 4-92.

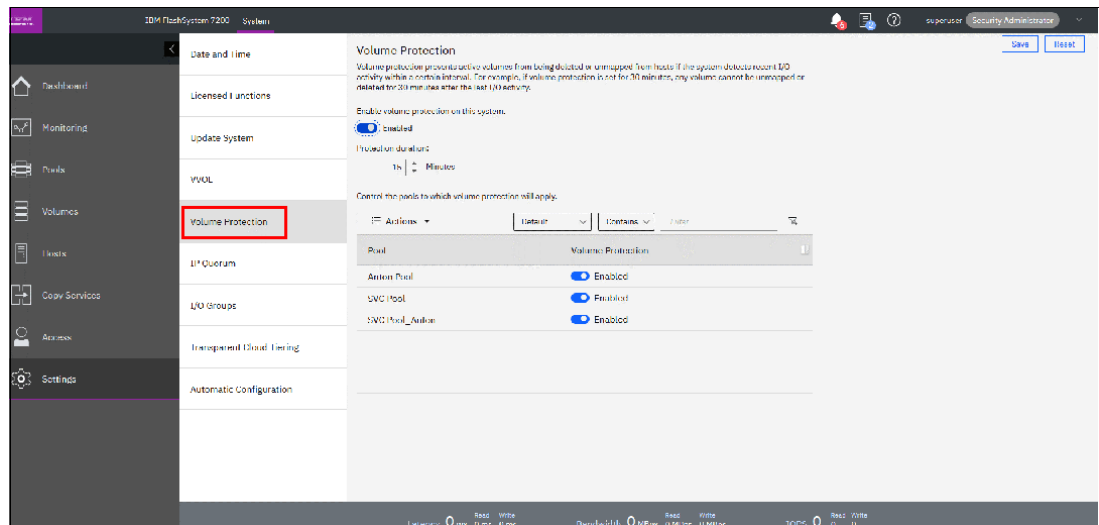


Figure 4-92 Volume protection

IP quorum

IBM Spectrum Virtualize also supports an IP quorum application. By using an IP-based quorum application as the quorum device for the third site, a FICON is not required. Java applications run on hosts at the third site.

To install the IP quorum device, complete the following steps:

1. If your IBM Storage System is configured for IPv4, click **Download IPv4 Application**. If it is configured for IPv6, select **Download IPv6 Application**. In our example, IPv4 is the option, as shown in Figure 4-93 on page 219.

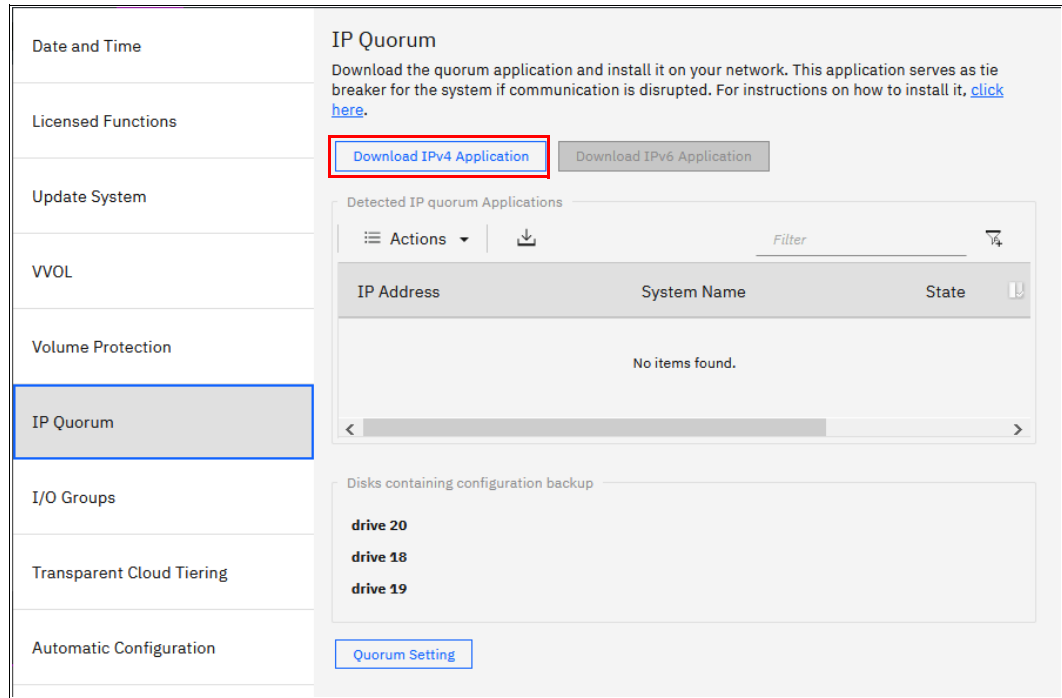


Figure 4-93 IP Quorum settings

- When you select **Download IPv4 Application**, you are prompted whether you want to download the IP quorum application with or without recovery metadata, as shown in Figure 4-94. IP quorum applications are used to resolve communication problems between nodes and store metadata, which restores system configuration during failure scenarios. If you have a third-site quorum disk that stores recovery metadata, you can download the IP quorum application without the recovery metadata.

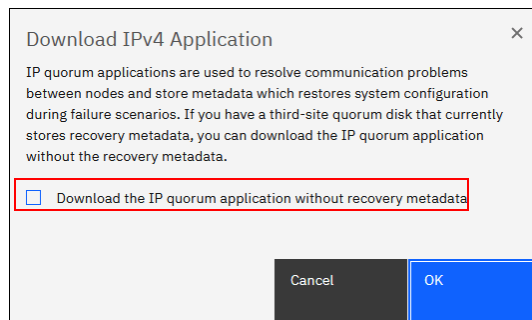


Figure 4-94 IP quorum application metadata

3. After you select your correct IP configuration, IBM Spectrum Virtualize generates an IP Quorum Java application, as shown in Figure 4-95. The application can be saved and installed in a host that is to run the IP quorum application.

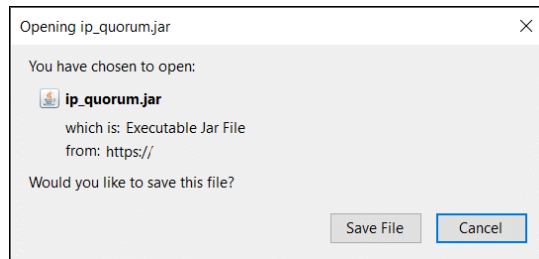


Figure 4-95 IP Quorum Java application

4. After you download the IP quorum application, you must save the application on a separate host or server.
5. If you change the configuration by adding a node, changing a service IP address, or changing Secure Sockets Layer (SSL) certificates, you must download and install the IP quorum application again.
6. On the host, you must use the Java command line to initialize the IP quorum application. On the server or host on which you plan to run the IP quorum application, create a separate directory that is dedicated to the IP quorum application.
7. Run the **ping** command on the host server to verify that it can establish a connection with the service IP address of each node in the system.
8. Change to the folder where the application is, and run the following command:

```
java -jar ip_quorum.jar
```

Note: The IP quorum application always must be running.

9. To verify that the IP quorum application is installed and active, select **Settings** → **System** → **IP Quorum**. The new IP quorum application is displayed in the table of detected applications. The system automatically selects MDisks for quorum disks.

An IP quorum application can also act as the quorum device for systems that are configured with a single-site or standard topology that does not have any external storage configured. The IP quorum mode is set to **Standard** when the system is configured for standard topology. The quorum mode of **Preferred** or **Winner** is available only if the system topology is not set to **Standard**. To change the quorum mode for the IP quorum application, select **Settings** → **System** → **IP Quorum** and set the mode to **Preferred** or **Winner**, or run the **chsystem** command. This configuration provides a system tie-break capability, automatically resuming I/O processing if half of the system's nodes or enclosures are inaccessible.

For specific quorum settings, see Figure 4-96.

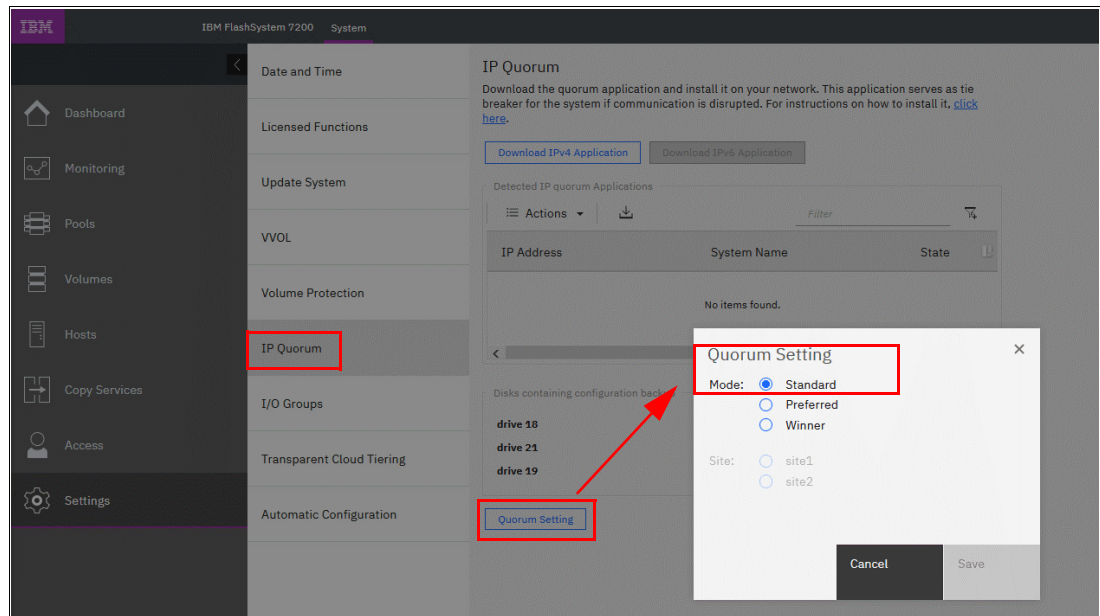


Figure 4-96 Quorum settings

On systems that support multiple-site topologies, you can specify which site resumes I/O after a disruption based on the applications that run on the site or other factors like whether the environment uses a third site for quorum management. For example, you can specify whether a selected site is preferred for resuming I/O or if the site automatically “wins” in tie-break scenarios. If only one site runs critical applications, you can configure this site as preferred.

During a disruption, the system delays processing tie-break operations on other sites that are not specified as preferred. The designated preferred site becomes more apt to resume I/O, and critical applications remain online. If the preferred site is the site that is disrupted, the other site continues to win the tie-breaks and continue I/O.

This feature applies only to IP quorum applications. It does not apply to FC-based third-site quorum management. In stretched configurations or HyperSwap configurations, an IP quorum application can be used at the third site as an alternative to third-site quorum disks. No FC connectivity at the third site is required to use an IP quorum application as the quorum device. If you have a third-site quorum disk, you must remove the third site before you use an IP quorum application.

Note: The maximum number of IP quorum applications that can be deployed on a single system is *five*. Only one instance of the IP quorum application per host or server is supported. IP quorum applications on multiple hosts or servers can be configured to provide redundancy. If you have multiple IBM Spectrum Virtualize systems in your environment, more than one IP quorum application is allowed per host, but each IP quorum instance must be dedicated to a single IBM Spectrum Virtualize system within the environment. In addition, the host or server requires available bandwidth to support multiple IP quorum instances.

Use the network requirements that are shown in “I/O groups” on page 222 to determine bandwidth and latency needs in these types of environments. The recommended configuration remains a single IP quorum application per host or server.

I/O groups

For ports within an I/O group, you can enable virtualization of FC ports that are used for host I/O operations. With N_Port ID Virtualization (NPIV), the FC port consists of both a physical port and a virtual port. When port virtualization is enabled, ports do not come up until they are ready to handle I/O, which improves host behavior. In addition, path failures due to an offline node are masked from hosts.

The target port mode on the I/O group indicates the current state of port virtualization:

- ▶ Enabled: The I/O group contains virtual ports that are available to use.
- ▶ Disabled: The I/O group does not contain any virtualized ports.
- ▶ Transitional: The I/O group contains physical FC and virtual ports that are being used. You cannot change the target port mode directly from Enabled to Disabled states, or vice versa. The target port mode must be in a transitional state before it can be changed to Disabled or Enabled states.

The system can be in the transitional state for an indefinite period while the system configuration is changed. However, system performance can be affected because the number of paths from the system to the host doubled. To avoid increasing the number of paths substantially, use zoning or other means to temporarily remove some of the paths until the state of the target port mode is enabled.

The port virtualization settings of I/O groups are available by selecting **Settings** → **System** → **I/O Groups**, as shown in Figure 4-97.

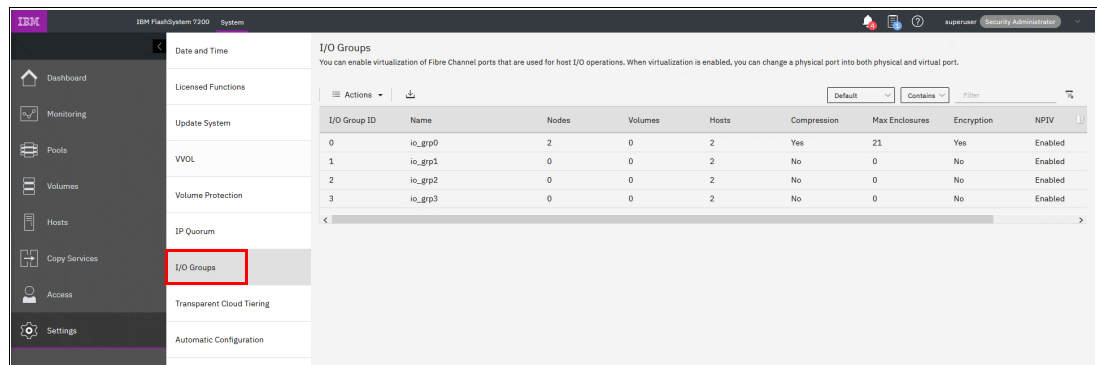


Figure 4-97 I/O group port virtualization

You can change the status of the port by right-clicking the I/O group and selecting **Change NPIV Settings**, as shown in Figure 4-98.

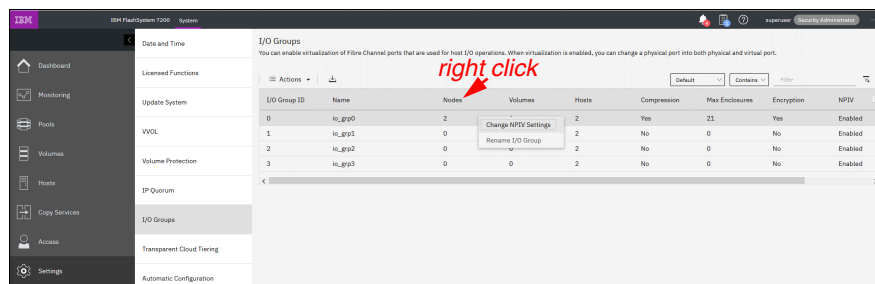


Figure 4-98 Changing NPIV settings

Transparent Cloud Tiering

TCT is a licensed function that enables volume data to be copied and transferred to cloud storage. The system supports creating connections to cloud service providers (CSPs) to store copies of volume data in private or public cloud storage.

With TCT, administrators can move older data to cloud storage to free up capacity on the system. PiT snapshots of data can be created on the system and then copied and stored on the cloud storage. An external CSP manages the cloud storage, which reduces storage costs for the system. Before data can be copied to cloud storage, a connection to the CSP must be created from the system.

A cloud account is an object on the system that represents a connection to a CSP by using a particular set of credentials. These credentials differ depending on the type of CSP that is being specified. Most CSPs require the hostname of the CSP and an associated password, and some CSPs also require certificates to authenticate users of the cloud storage.

Public clouds use certificates that are signed by well-known CAs. Private CSPs can use a self-signed certificate or a certificate that is signed by a trusted CA. These credentials are defined on the CSP and passed to the system through the administrators of the CSP. A cloud account defines whether the system can successfully communicate and authenticate with the CSP by using the account credentials.

If the system is authenticated, it can then access cloud storage to copy data to the cloud storage or restore data that is copied to cloud storage back to the system. The system supports one cloud account to a single CSP. Migration between providers is not supported.

Note: Before enabling TCT, consider the following requirements:

- ▶ Ensure that the DNS is configured on your system and accessible.
- ▶ Determine whether your company's security policies require enabled encryption. If yes, ensure that the encryption licenses are properly installed and that encryption is enabled.

The system supports connections to various CSPs. Some CSPs require connections over external networks, and others can be created on a private network.

Each CSP requires different configuration options. The system supports the following CSPs:

- ▶ IBM Cloud

The system can connect to IBM Cloud, which is a cloud computing platform that combines platform as a service (PaaS) with infrastructure as a service (IaaS).

- ▶ OpenStack Swift

OpenStack Swift is a standard cloud computing architecture from which administrators can manage storage and networking resources in a single private cloud environment. Standard APIs can be used to build customizable solutions for a private cloud solution.

- ▶ Amazon Simple Storage Service (Amazon S3)

Amazon S3 provides programmers and storage administrators with flexible and secure public cloud storage. Amazon S3 is also based on Object Storage standards and provides a web-based interface to manage, back up, and restore data over the web.

To view your IBM Spectrum Virtualize cloud provider settings, from the IBM Storage System Settings window, click **Settings** and select **System**. Then, select **Transparent Cloud Tiering**, as shown in Figure 4-99.

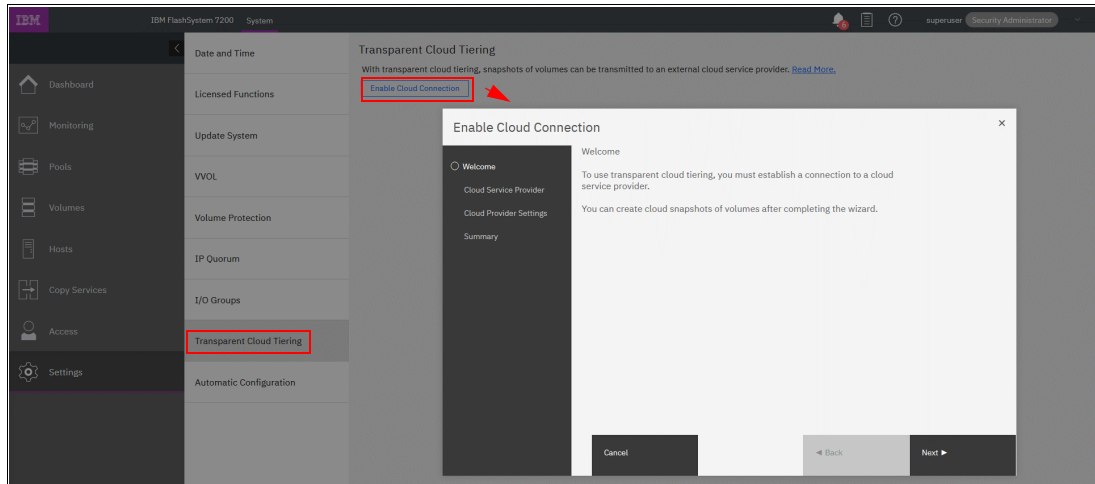


Figure 4-99 Transparent Cloud Tiering settings

By using this view, you can enable and disable features of your TCT and update the system information concerning your CSP. This window allows you to set the following options:

- ▶ CSP
- ▶ Cloud Object Storage Uniform Resource Locator (URL)
- ▶ The tenant or the container information that is associated to your Cloud Object Storage
- ▶ Username of the cloud object account
- ▶ API key
- ▶ The container prefix or location of your object
- ▶ Encryption
- ▶ Bandwidth

For more information about how to configure and enable TCT, see 10.3, “Transparent Cloud Tiering” on page 621.

Automatic Configuration for Virtualization

If you are using this system as external storage for an IBM SAN Volume Controller (SVC) that uses FC, this system can be automatically configured following best practices for usage as external storage behind an SVC.

This process completes the following actions automatically:

- ▶ Creates the appropriate redundant array of independent disks (RAID) arrays based on the technology type of the drives.
- ▶ Creates a pool for each array.
- ▶ Provisions all capacity in each pool to volumes based on best practices.
- ▶ Maps all volumes to the SVC system for virtualization as MDisks.

Figure 4-100 shows how to enable Automatic Configuration for Virtualization.

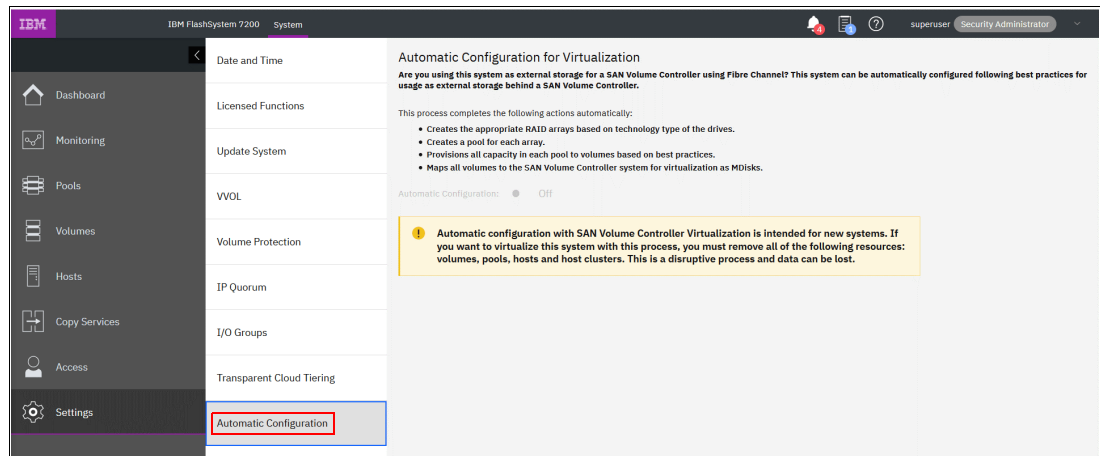


Figure 4-100 Automatic Configuration for Virtualization

4.10.6 Support menu

Use the Support window to configure and manage connections and upload support packages to the IBM Support Center.

The following options are available from the menu:

► Call Home

The Call Home feature transmits operational and event-related data to you and IBM through a Simple Mail Transfer Protocol (SMTP) server connection in the form of an event notification email. When configured, this function alerts IBM Support personnel about hardware failures and potentially serious configuration or environmental issues.

This view provides the following useful information about email notification and Call Home information (among others), as shown in Figure 4-101:

- IP of the email server (SMTP server) and port.
- Call Home email address.
- Email of one or more users set to receive one or more email notifications.
- Contact information of the person in the organization that is responsible for the system.

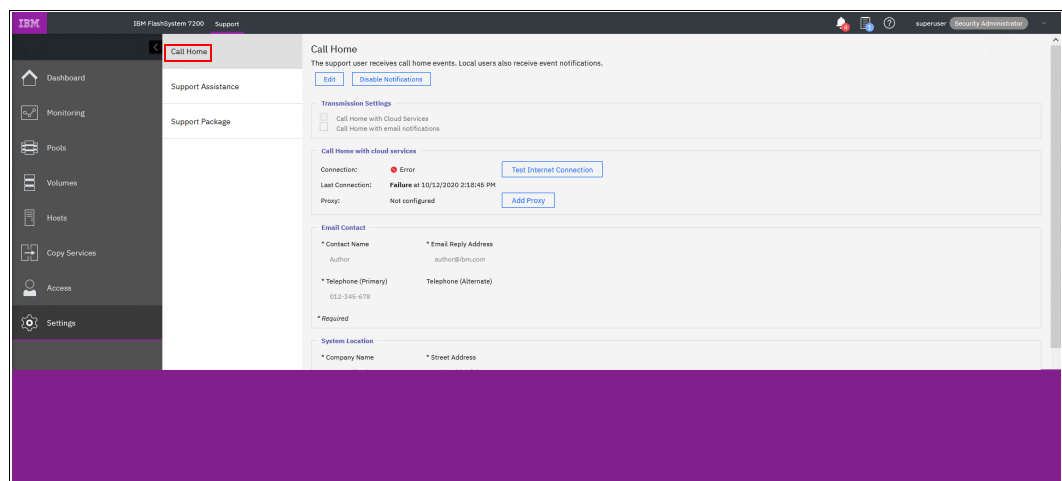


Figure 4-101 Call Home settings

► Support Assistance

This option enables IBM Support personnel to access the system to complete troubleshooting and maintenance tasks. You can configure local Support Assistance, where IBM Support personnel visit your site to fix problems with the system, or Remote Support Assistance. Both local and Remote Support Assistance use secure connections to protect data exchange between the IBM Support Center and the system. More access controls can be added by the system administrator, as shown in Figure 4-102.

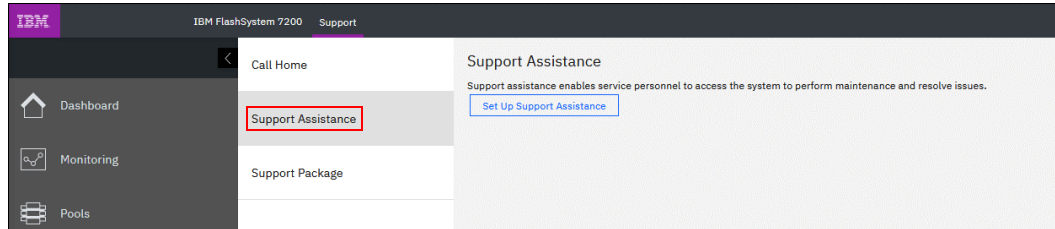


Figure 4-102 Support assistance

► Support Package

If Support Assistance is configured on your systems, you can automatically or manually upload new support packages to the IBM Support Center to help analyze and resolve errors on the system.

The menus are available by selecting **Settings** → **Support** → **Support package**, as shown in Figure 4-103.

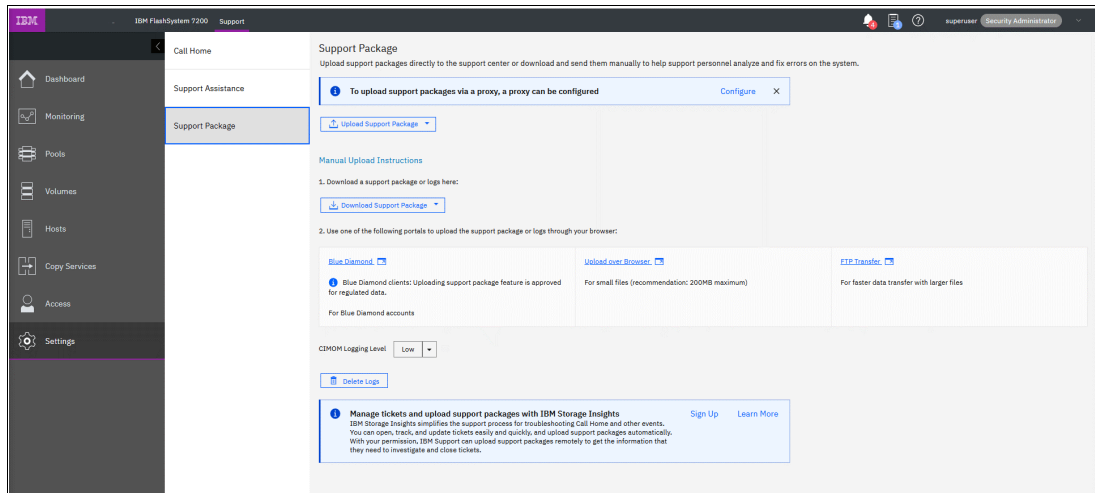


Figure 4-103 Support package menu

For more information about how the Support menu helps with troubleshooting your system or how to back up your systems, see Chapter 13, “Reliability, availability, and serviceability, monitoring and logging, and troubleshooting” on page 793.

4.10.7 GUI Preferences menu

The **GUI Preferences** menu consists of the following options:

- Login
- General

Figure 4-104 shows the GUI Preferences selection window.

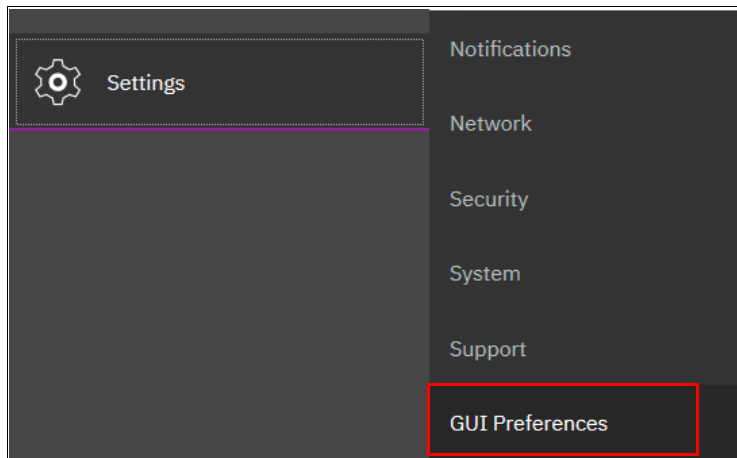


Figure 4-104 GUI Preferences selection window

Login message

IBM Spectrum Virtualize enables administrators to configure the welcome banner (login message). This message is a text message that appears in the GUI login window or at the CLI login prompt.

The content of the welcome message is helpful when you need to notify users about some important information about the system, such as security warnings or a location description. To define and enable the welcome message by using the GUI, edit the text area with the message content and click **Save** (see Figure 4-105).

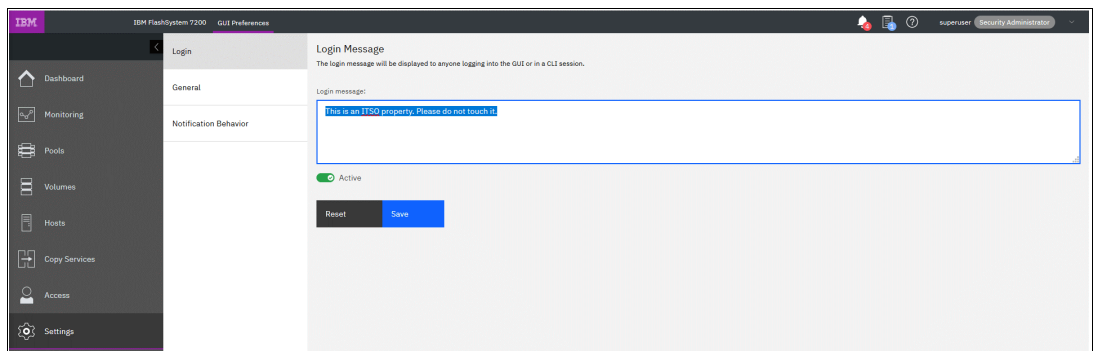


Figure 4-105 Enabling the login message

The resulting login dialog box is shown in Figure 4-106.

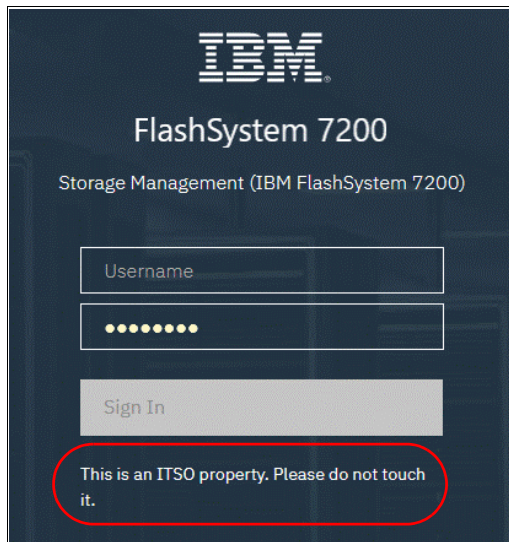


Figure 4-106 Welcome message in the GUI

The banner message also appears in the CLI login prompt window, as shown in Figure 4-107.

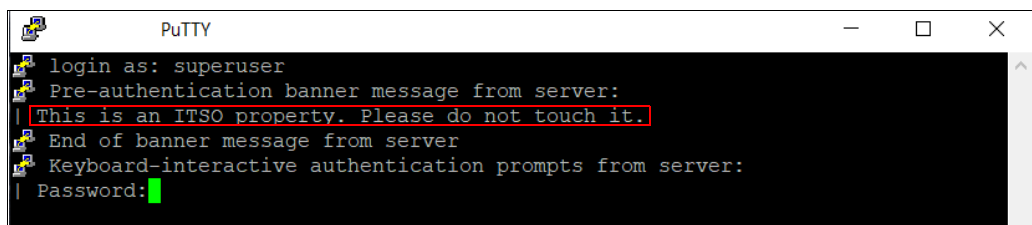


Figure 4-107 Welcome message in CLI

General Settings

With the **General Settings** menu, you can refresh the GUI cache, set the low graphics mode option, and enable advanced pools settings.

To configure general GUI preferences, complete the following steps:

1. From the Settings window, click **Settings** and select **GUI Preferences** → **General** (see Figure 4-108 on page 229).

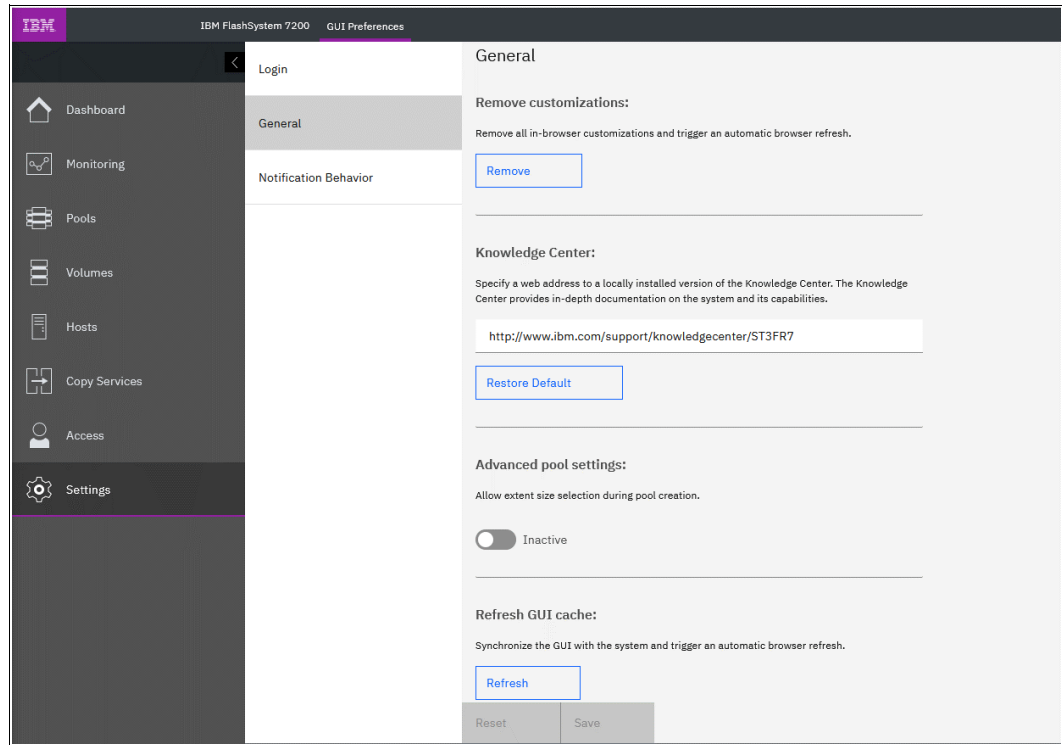


Figure 4-108 General GUI Preferences window

2. You can configure the following elements:

- Clear customizations

This option deletes all GUI preferences that are stored in the browser and restores the default preferences.

- IBM Documentation

You can change the URL of IBM Documentation for IBM Spectrum Virtualize.

- Advanced pool settings

You can select the extent size during storage pool creation.

- Refresh GUI cache

This option causes the GUI to refresh all its views and clears the GUI cache. The GUI looks up every object again. This option is useful if a value or object that is shown in the CLI is not being reflected in the GUI.

Notification Behavior

Figure 4-109 shows that you can allow certain notifications to remain on screen until they are manually dismissed.

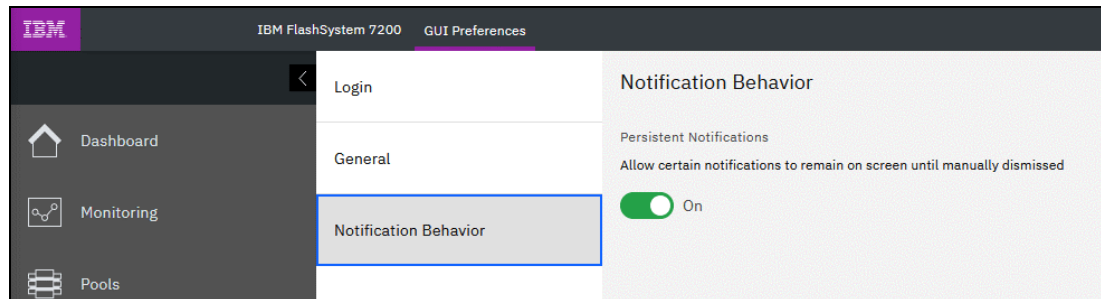


Figure 4-109 Notification Behavior

4.11 Additional frequent tasks in the GUI

This section describes additional options and tasks that are available in the GUI of your system that are frequently used by administrators.

4.11.1 Renaming components

These sections provide guidance about how to rename your system and canisters.

Renaming your storage system

All objects in the system have names that are user-defined or system-generated. Choose a meaningful name when you create an object. If you do not choose a name for the object, the system generates a name for you.

A well-chosen name serves both as a label for an object and as a tool for tracking and managing the object. Choosing a meaningful name is important if you decide to use configuration backup and restore.

When you choose a name for an object, apply the following naming rules:

- ▶ Names must begin with a letter.

Important: Do not start names by using an underscore (`_`) character even though it is possible. Using an underscore as the first character of a name is a reserved naming convention that is used by the system configuration restore process.

- ▶ The first character cannot be numeric.
- ▶ The name can be a maximum of 63 characters, but there are exceptions. The name can be a maximum of 15 characters for RC relationships and groups. The `lsfabric` command displays long object names that are truncated to 15 characters for nodes and systems. (`lsrelationshipcandidate` or `lsrelationship` commands).
- ▶ Valid characters are uppercase letters (A - Z), lowercase letters (a - z), digits (0 - 9), the underscore (`_`) character, a period (`.`), a hyphen (`-`), and a space.
- ▶ Names must not begin or end with a space.

- ▶ Object names must be unique within the object type. For example, you can have a volume that is called ABC and an MDisk called ABC, but you cannot have two volumes that are called ABC.
- ▶ The default object name is valid (an object prefix with an integer).
- ▶ Objects can be renamed to their current names.

To rename the system from the System window, complete the following steps:

1. Select **Monitoring** → **System Hardware - Overview**, and click **System Actions** in the upper right of the window, as shown in Figure 4-110.

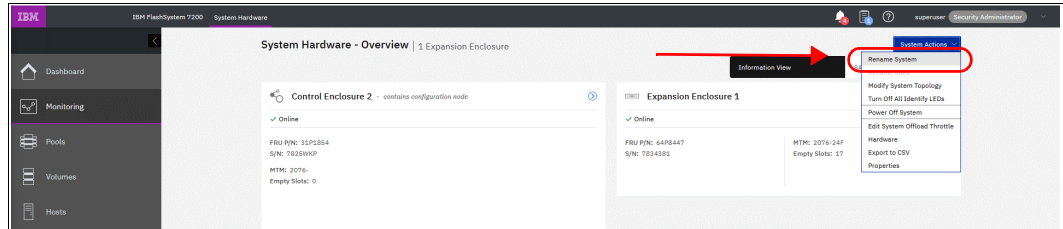


Figure 4-110 Actions menu in the System Hardware - Overview window

2. The Rename System window opens (see Figure 4-111). Specify a new name for the system and click **Rename**.

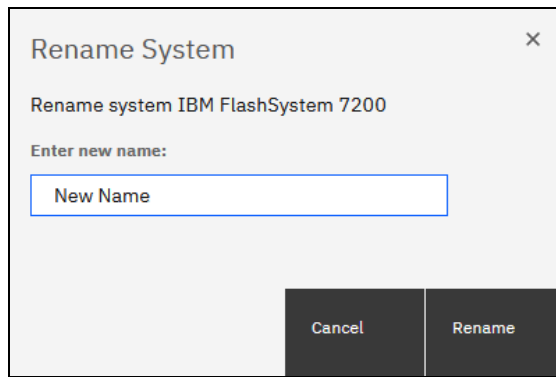


Figure 4-111 Renaming the system

System name: You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The clustered system name can be 1 - 63 characters.

Warning: When you rename your system, the iSCSI name automatically changes because it includes the system name by default. Therefore, this change needs more actions on iSCSI-attached hosts.

Renaming a node canister

To rename a node canister, complete the following steps:

1. Go to the **System Hardware - Overview** window and right-click the node that you want to rename, as shown in Figure 4-112. Click **Rename**.

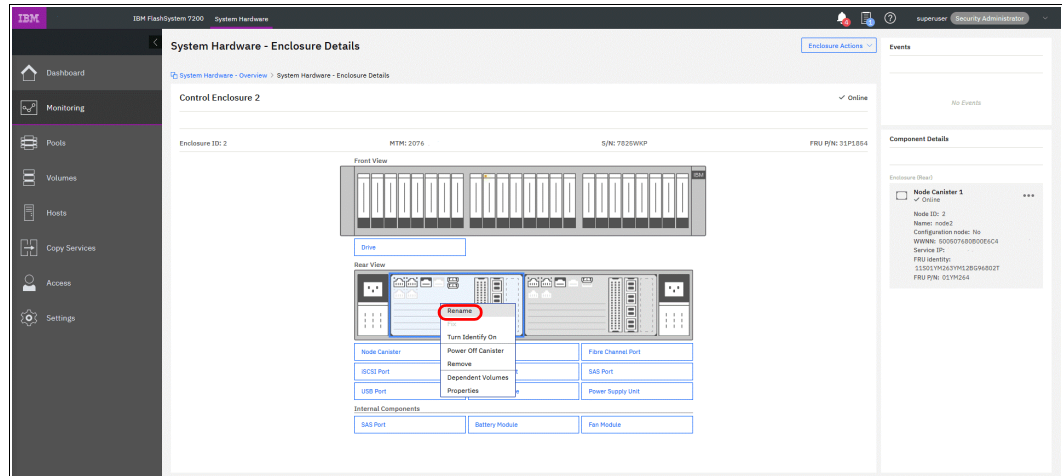


Figure 4-112 Renaming a node on the System Hardware - Overview window

2. Enter the new name of the node and click **Rename** (see Figure 4-113).

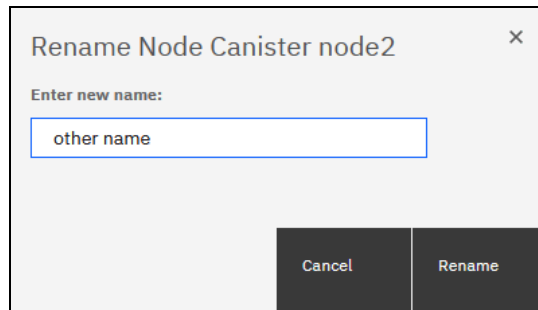


Figure 4-113 Entering the new name of the node

Warning: Changing the node canister name causes an automatic IQN update and requires the reconfiguration of all iSCSI-attached hosts.

4.11.2 Working with enclosures

The following section describes how to add or remove expansion enclosure to or from your IBM Storage System.

Adding an enclosure

After the expansion enclosure is properly attached and powered on, complete the following steps to activate it in the system:

1. In the System window that is available from the **Monitoring** menu, select **SAS Chain View**. Only correctly attached and powered on enclosures appear in the window, as shown in Figure 4-114 on page 233. The new enclosure is showing as unmanaged, which means it is not part of the system.

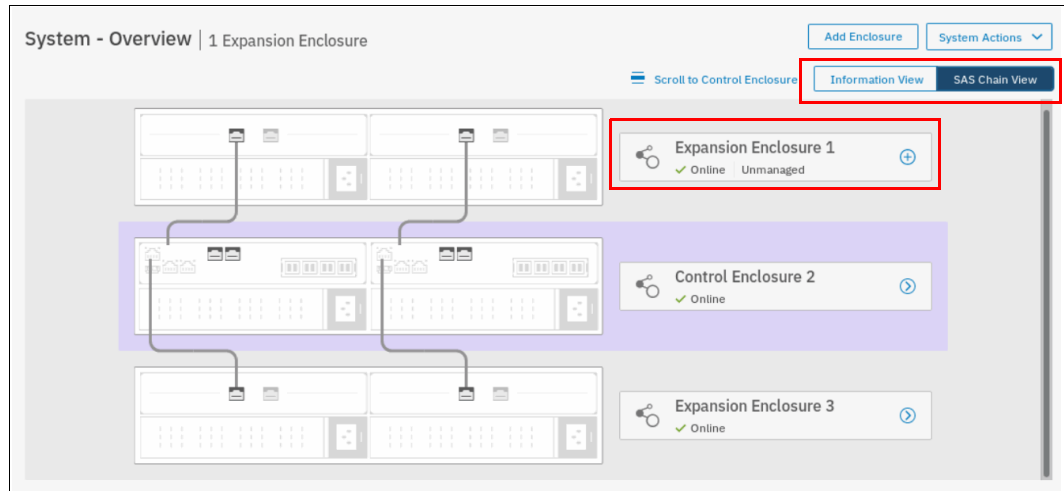


Figure 4-114 Newly detected expansion enclosure

2. Select the + next to the enclosure that you want to add or click **Add Enclosure** at the top. These buttons appear only if there is an unmanaged enclosure that is eligible to be added to the system. After they are selected, a window opens, on which you need to select the enclosure you want to add. Expansion enclosures that are directly cabled do not need to be selected, as shown in Figure 4-115.

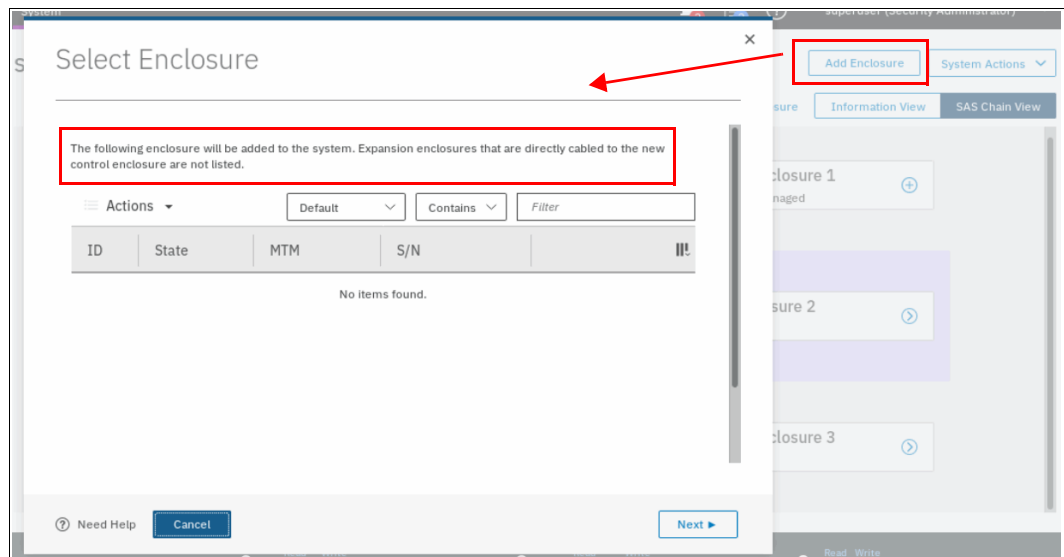


Figure 4-115 Adding an enclosure

3. Select **Next** and then **Finish** after you are satisfied with your selections. The enclosures are then added to the system and appear as managed. Instead of the + button, you see a >, which allows you to view details about the enclosure because it is now part of the system, as shown in Figure 4-116.

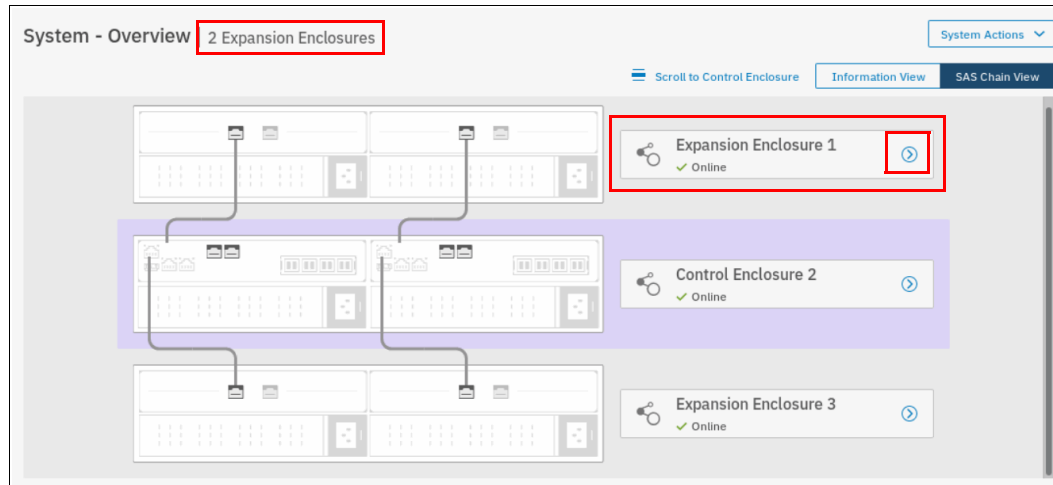


Figure 4-116 Enclosure successfully added

Removing an enclosure

The enclosure removal procedure includes its logical detachment from the system by using a GUI and physically unmounting the systems from the rack. The IBM Storage System guides you through this process. Complete the following steps:

1. In the System window that is available from the **Monitoring** menu, select > next to the enclosure that you want to remove. The Enclosure Details pane opens. You can then click **Enclosure Actions** and select **Remove**, as shown in Figure 4-117.

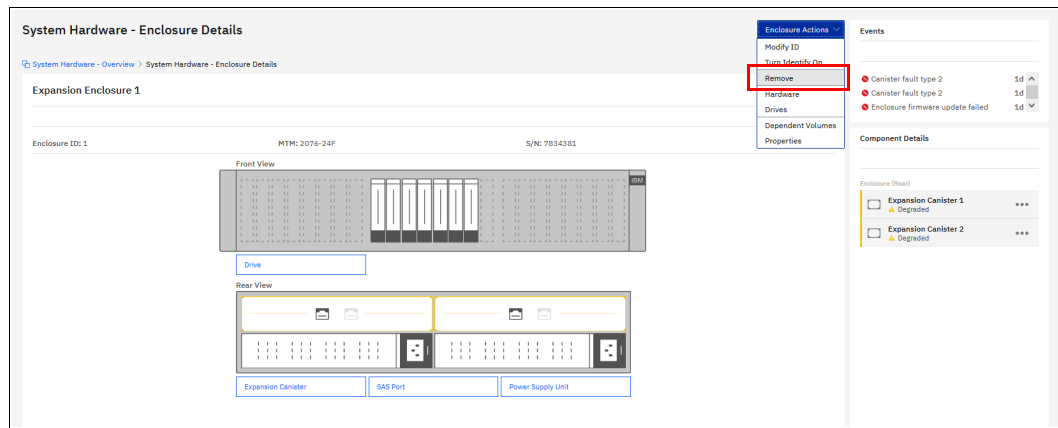


Figure 4-117 Selecting an enclosure for removal

2. The system prompts you to remove the enclosure. All disk drives in the removed enclosure must be in the *Unused* state. Otherwise, the removal process fails (see Figure 4-118 on page 235).

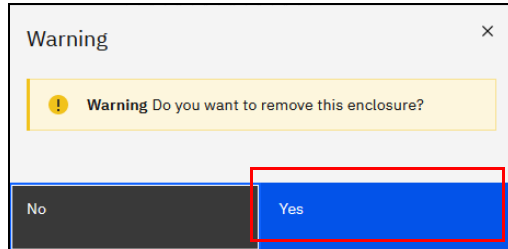


Figure 4-118 Confirming the removal

3. After the enclosure is logically removed from the system (set to the *Unmanaged* state), the system reminds you about the steps that are necessary for physical removal, such as power off, uncabling, dismantling from the rack, and secure handling (see Figure 4-119).

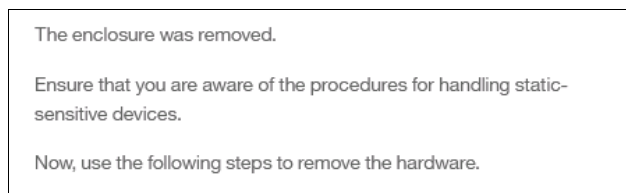


Figure 4-119 Enclosure removed

As part of the enclosure removal process, see your company security policies about how to handle sensitive data on removed storage devices before they leave the secure data center. Most companies require data to be encrypted or logically shredded.

4.11.3 Restarting the GUI service

The service that runs that GUI operates from the configuration node. Occasionally, you might need to restart this service if the GUI is not performing to your expectation (or you cannot connect). To do this task, complete the following steps:

1. Log in to the Service Assistant and identify the configuration node, as shown in Figure 4-120.

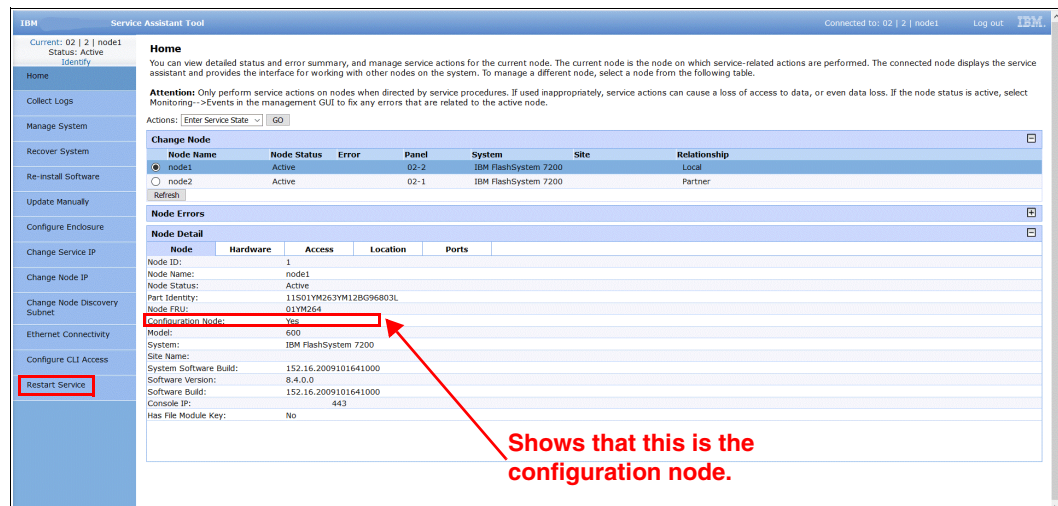


Figure 4-120 Identifying the configuration node on the Service Assistant

2. After the process completes, go to **Restart Service**, as shown in Figure 4-121.

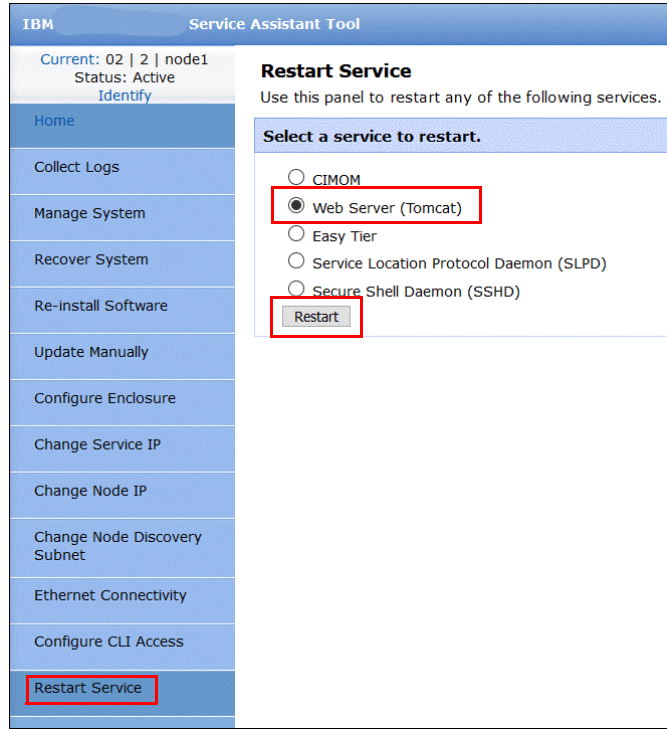


Figure 4-121 Restarting the Tomcat web server

3. Select **Web Server (Tomcat)**. Click **Restart**, and the web server that runs the GUI restarts. This task is a concurrent action, but the cluster GUI is unavailable while the server is restarting (the Service Assistant and CLI are not affected). After 5 minutes, check to see whether GUI access was restored.



Storage pools

This chapter describes how the storage system manages physical storage resources. All storage resources that are under system control are managed by using *storage pools* or *managed disk (MDisk) groups*.

Storage pools aggregate internal and external capacity and provide the containers in which you can create volumes. Storage pools make it easier to dynamically allocate resources, maximize productivity, and reduce costs.

You can configure storage pools through the management GUI, either during initial configuration or later. Alternatively, you can configure the storage to your own requirements by using the command-line interface (CLI).

This chapter includes the following topics:

- ▶ 5.1, “Working with storage pools” on page 238
- ▶ 5.2, “Working with internal drives and arrays” on page 257
- ▶ 5.3, “Working with external controllers and MDisks” on page 286

5.1 Working with storage pools

Storage pools act as containers for MDisks, which provide storage capacity to the pool, and volumes that are provisioned from this capacity, which can be mapped to host systems. The system organizes storage in this fashion to ease storage management and make it more efficient.

MDisks can either be redundant array of independent disks (RAID) arrays that are created by using internal storage, such as drives and flash modules, or logical units (LUs) that are provided by external storage systems. A single storage pool can contain both types of MDisks, but a single MDisk can be part of only one storage pool. MDisks themselves are not visible to host systems.

Figure 5-1 provides an overview of how storage pools, MDisks, and volumes are related.

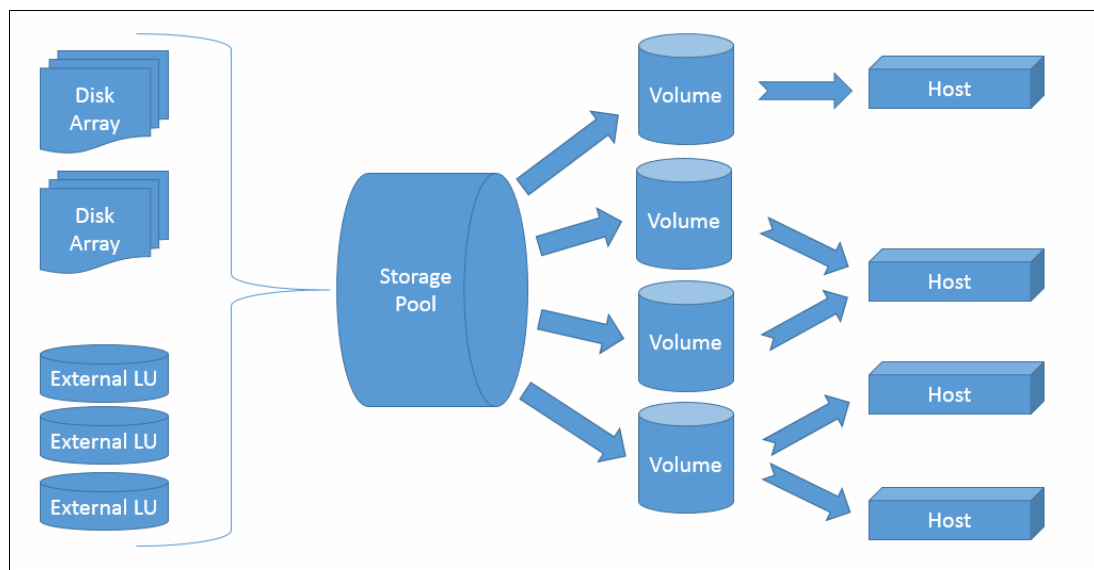


Figure 5-1 Relationship between MDisks, storage pools, and volumes

All MDisks in a pool are split into chunks of the same size, which are called *extents*. Volumes are created from the set of available extents in the pool. The extent size is a property of the storage pool and cannot be changed after the pool is created. The choice of extent size affects the total amount of storage that can be managed by the system.

It is possible to add MDisks to an existing pool to provide more usable capacity in the form of extents. The system automatically balances volume extents between the MDisks to provide the best performance to the volumes. It is also possible to remove extents from the pool by deleting an MDisk. The system automatically migrates extents that are in use by volumes to other MDisks in the same pool to make sure that the data on the extents is preserved.

A storage pool represents a failure domain. If one or more MDisks in a pool become inaccessible, all volumes (except for image mode volumes) in that pool are affected. Volumes in other pools are unaffected.

The system supports *standard pools* and *Data Reduction Pools (DRPs)*. Both support parent pools and child pools.

Child pools are created from existing capacity that is assigned to a parent pool instead of being created directly from MDisks. When the child pool is created from a standard pool, the capacity for a child pool is reserved from the parent pool. This capacity is no longer reported as available capacity of the parent pool. In terms of volume creation and management, child pools are similar to parent pools.

DRPs use a set of techniques that can be used to reduce the amount of usable capacity that is required to store data, such as compression and deduplication. Data reduction can increase storage efficiency and performance, and reduce storage costs, especially for flash storage. DRPs automatically reclaim capacity that is no longer needed by host systems. This reclaimed capacity is given back to the pool as usable capacity and can be reused by other volumes. Child pools that are created from DRPs are quotaless and can use the entire parent pool capacity.

For more information about DRP planning and implementation, see Chapter 9, “Advanced features for storage efficiency” on page 509 and *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430.

In general, you manage storage as follows:

1. Create storage pools (standard or DRP), depending on your requirements and sizing.
2. Assign storage to these pools by using one or more of the following options:
 - Create array MDisks from internal drives or flash modules.
 - Add MDisks provisioned from external storage systems.
3. Create volumes in these pools and map them to hosts or host clusters.

You manage storage pools either in the Pools window of the GUI or by using the CLI. To access the Pools pane, select **Pools** → **Pools**, as shown in Figure 5-2.

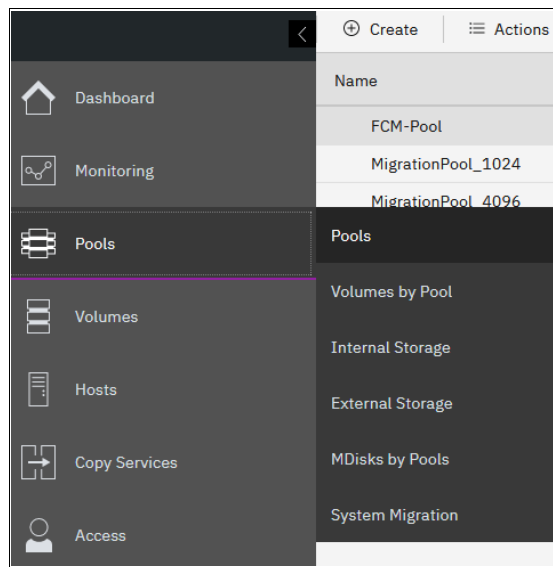


Figure 5-2 Accessing the Pools window

The window lists all storage pools and their major parameters. If a storage pool has child pools, they are also shown.

To see a list of configured storage pools by using the CLI, run the `lsmdiskgrp` command without any parameters, as shown in Example 5-1.

Example 5-1 The `lsmdiskgrp` output (some columns are not shown)

```
IBM_IBM FlashSystem_7200:superuser>lsmdiskgrp
id name          status mdisk_count vdisk_count capacity extent_size free_capacity
0 NVMe-Pool0    online 13          76          6.71TB  2048 634.00GB
2 FCM-Pool      online 1           71          178.81TB 2048 170.04TB
```

5.1.1 Creating storage pools

To create a storage pool, complete the following steps:

1. Select **Pools** → **MDisks by Pools** and click **Create Pool**, or select **Pools** → **Pools** and click **Create**, as shown in Figure 5-3.

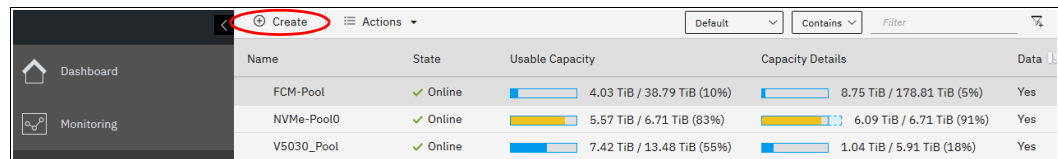


Figure 5-3 Option to create a storage pool in the Pools window

Both alternatives open the dialog box that is shown in Figure 5-4.

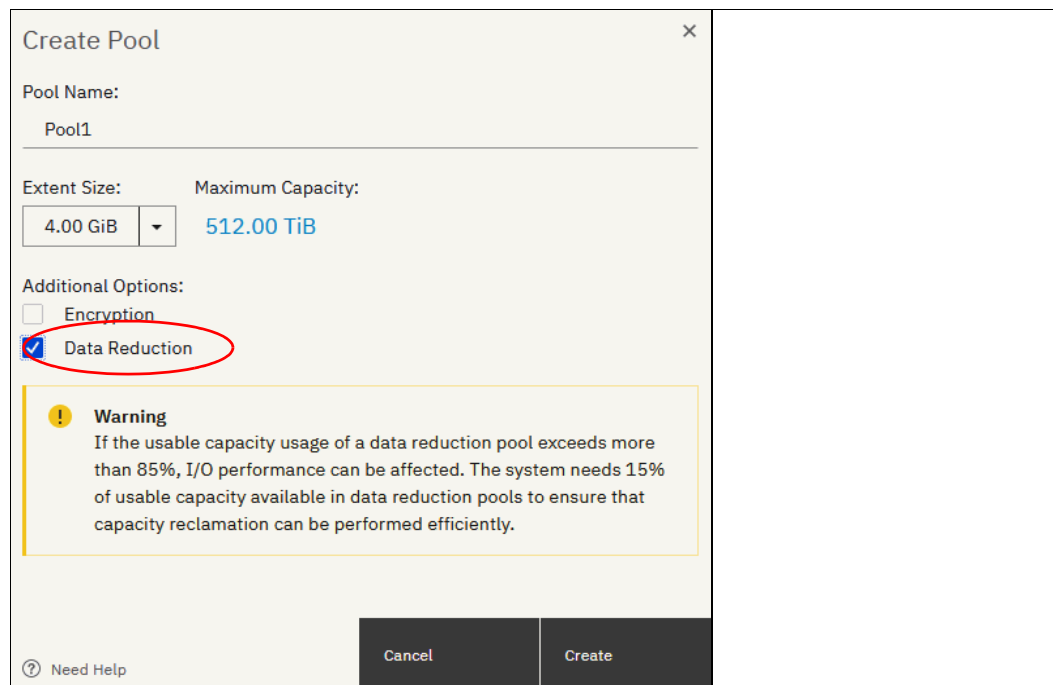


Figure 5-4 Create Pool dialog box

2. Select the **Data reduction** check box to create a DRP. Leaving it clear creates a standard storage pool.

Note: DRPs require careful planning and sizing. Limitations and performance characteristics of DRPs are different from standard pools.

A standard storage pool that is created by using the GUI has a default extent size of 1 GB. DRPs have a default extent size of 4 GB. The size of the extents is selected at creation time and cannot be changed later. The extent size controls the maximum total storage capacity that is manageable per system (across all pools). For DRPs, the extent size also controls the maximum capacity after reduction in the pool itself.

For more information about the differences between standard pools and DRPs and for extent size planning, see Chapter 2, “Planning” on page 71 and Chapter 9, “Advanced features for storage efficiency” on page 509.

Note: Do not create DRPs with small extent sizes. For more information, see this [IBM Support alert](#).

When creating a standard pool, you cannot change the extent size by using the GUI by default. If you want to specify a different extent size, enable this option by selecting **Settings** → **GUI Preferences** → **General** and checking **Advanced pool settings**, as shown in Figure 5-5. Click **Save**.

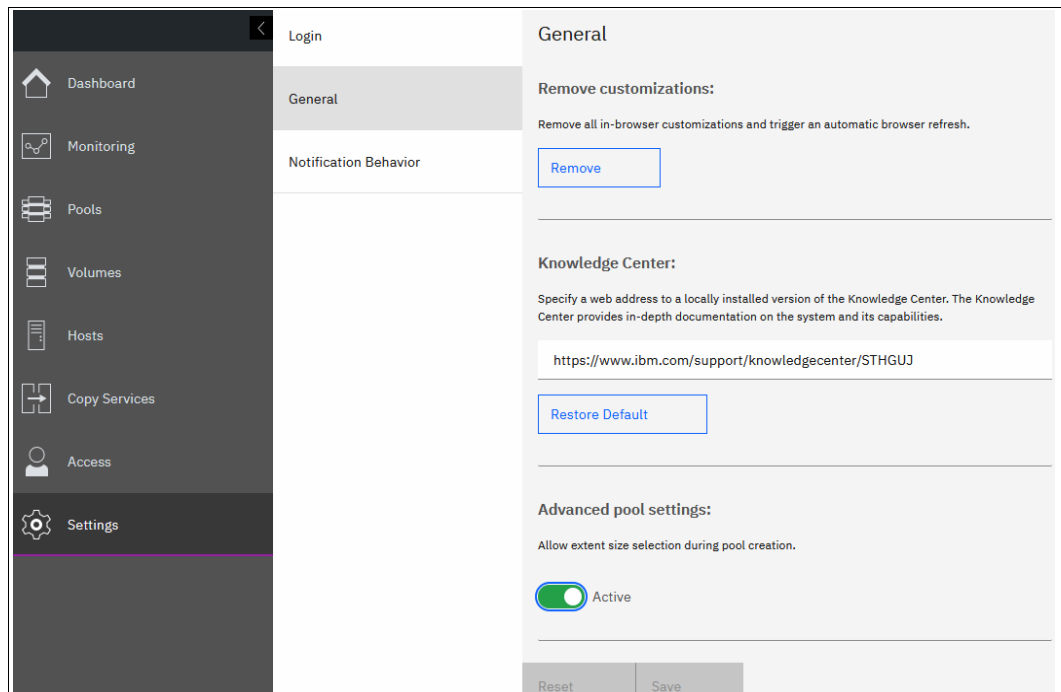


Figure 5-5 Advanced pool settings

When the **Advanced pool settings** option is enabled, you can also select an extent size for standard pools at creation time, as shown in Figure 5-6.

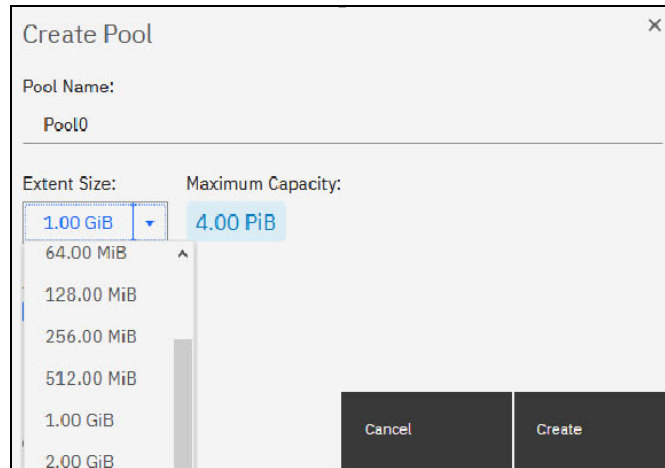


Figure 5-6 Creating a standard pool with Advanced pool settings selected

If an encryption license is installed and enabled, you can select whether the storage pool is encrypted, as shown in Figure 5-7. The encryption setting of a storage pool is selected at creation time and cannot be changed later. By default, if encryption is enabled, encryption is selected. For more information about encryption and encrypted storage pools, see Chapter 12, “Encryption” on page 735.

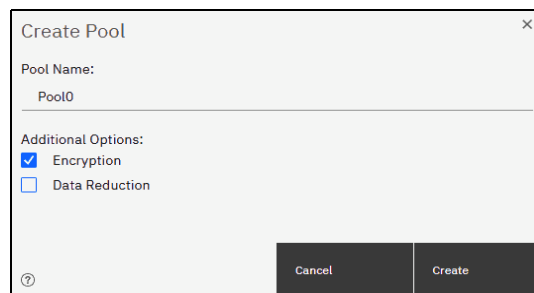


Figure 5-7 Creating a pool with encryption enabled

3. Enter the name for the pool and click **Create**.

Naming rules: When you choose a name for a pool, the following rules apply:

- ▶ Names must begin with a letter.
- ▶ The first character cannot be numeric.
- ▶ The name can be a maximum of 63 characters.
- ▶ Valid characters are uppercase letters (A - Z), lowercase letters (a - z), digits (0 - 9), underscore (_), period (.), hyphen (-), and space.
- ▶ Names must not begin or end with a space.
- ▶ Object names must be unique within the object type. For example, you can have a volume that is named *ABC* and a storage pool that is called *ABC*, but not two storage pools that are both called *ABC*.
- ▶ The default object name is valid (object prefix with an integer).
- ▶ Objects can be renamed at a later stage.

The new pool is created and is included in the list of storage pools with zero bytes, as shown in Figure 5-8.

Name	State	Usable Capacity	Capacity Details	Data
FCM-Pool	Online	4.03 TiB / 38.79 TiB (10%)	8.76 TiB / 178.81 TiB (5%)	Yes
NVMe-Pool0	Online	5.57 TiB / 6.71 TiB (83%)	6.09 TiB / 6.71 TiB (91%)	Yes
V5030_Pool	Online	7.42 TiB / 13.48 TiB (55%)	1.04 TiB / 5.91 TiB (18%)	Yes
Pool0	Online	0 bytes	0 bytes	No

Figure 5-8 Newly created empty pool

To perform this task by using the CLI, run the `mkmdiskgrp` command. The only required parameter is the extent size, which is specified by the `-ext` parameter and must have one of the following values: 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, or 8192 (MB). To create a DRP, specify `-datareduction yes`. The minimum extent size of DRPs is 1024, and attempting to use a smaller extent size sets the extent size to 1024.

In Example 5-2, the command creates a DRP that is named `Pool0` with no MDisks in it.

Example 5-2 The `mkmdiskgrp` command

```
IBM_IBM FlashSystem 7200:superuser>mkmdiskgrp -name Pool0 -datareduction yes -ext 8192
MDisk Group, id [3], successfully created
```

5.1.2 Managed disks in a storage pool

A storage pool is created as an empty container with no storage that is assigned to it. Storage is then added in the form of *MDisks*. An MDisk can be either a RAID array from internal storage (as an array of drives) or a LU from an external storage system. The same storage pool can include both internal and external MDisks.

Arrays are assigned to storage pools at creation time. Arrays cannot exist outside of a storage pool and they cannot be moved between storage pools. It is only possible to destroy an array by removing it from a pool and re-creating it within a new pool.

External MDisks can exist within a pool or outside of a pool. The MDisk object remains on a system if it is visible from external storage, but its access mode changes depending on whether it is assigned to a pool or not.

MDisks are managed by using the MDisks by Pools window. To access the MDisks by Pools window, select **Pools** → **MDisks by Pools**, as shown in Figure 5-9.

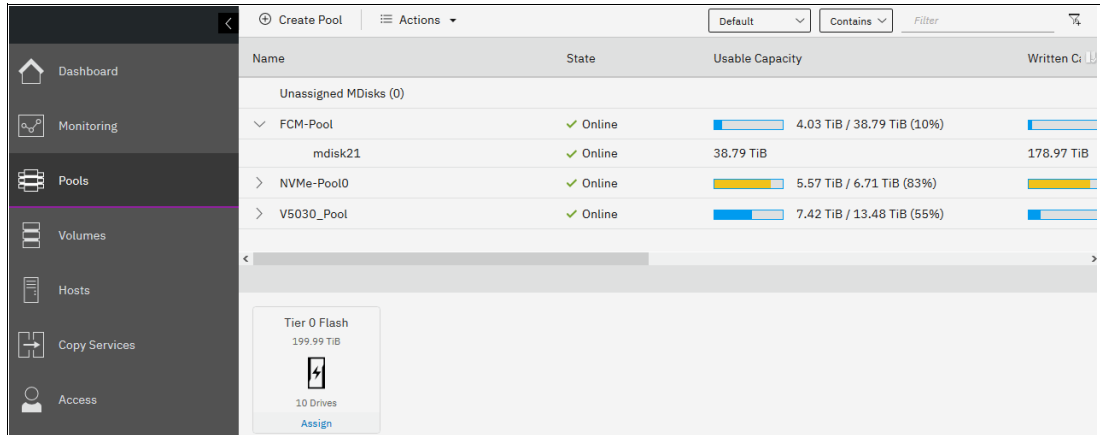


Figure 5-9 MDisks by Pools

The window lists all the MDisks that are available in the system under the storage pool to which they belong. Unassigned MDisks are listed separately at the top. Both arrays and external MDisks are listed. For more information about operations with array MDisks, see 5.2, “Working with internal drives and arrays” on page 257. To implement a solution with external MDisks, see 5.3, “Working with external controllers and MDisks” on page 286.

To list all MDisks that are visible by the system by using the CLI, run the `lsmdisk` command without any parameters. If required, you can filter output to include only external or only array type MDisks.

5.1.3 Actions on storage pools

A number of actions can be performed on storage pools. To select an action, select the storage pool and click **Actions**, as shown in Figure 5-10. Alternatively, right-click the storage pool.

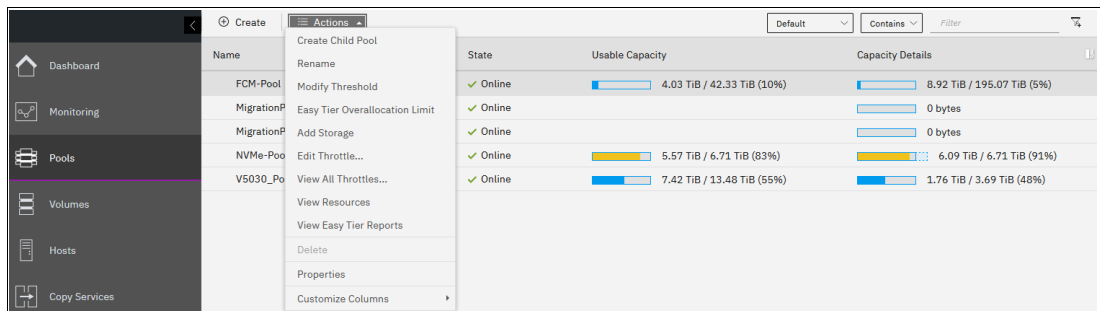


Figure 5-10 Pools actions menu

Create Child Pool window

To create a child storage pool, click **Create Child Pool**. For more information about child storage pools and a detailed description of this wizard, see 5.1.4, “Child pools” on page 252. It is not possible to create a child pool from an empty pool.

Rename window

To modify the name of a storage pool, click **Rename**. Enter the new name and click **Rename** in the dialog window.

To do this task by using the CLI, run the **chmdiskgrp** command. Example 5-3 shows how to rename Pool2 to StandardStoragePool. If successful, the command returns no output.

Example 5-3 Using chmdiskgrp to rename a storage pool

```
IBM__IBM FlashSystem 7200:superuser>chmdiskgrp -name StandardStoragePool Pool2
IBM__IBM FlashSystem 7200:superuser>
```

Modify Threshold window

A warning event is generated when the amount of used capacity in the pool exceeds the warning threshold. When you use thin-provisioned volumes that auto-expand (automatically use available extents from the pool), monitor the capacity usage and get warnings before the pool runs out of free extents so that you can add storage before running out of space.

Note: The warning is generated only the first time that the threshold is exceeded by the used capacity in the storage pool.

To modify the threshold, select **Modify Threshold** and enter the new value. The default threshold is 80%. To disable warnings, set the threshold to 0%.

The threshold is visible in the pool properties and indicated by a red bar, as shown in Figure 5-11.

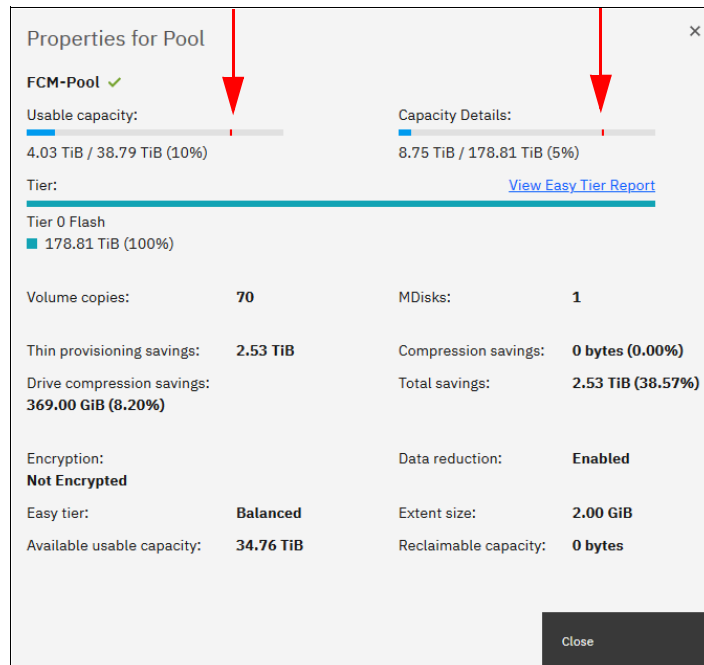


Figure 5-11 Pool properties with a warning threshold

To do the task by using the CLI, run the **chmdiskgrp** command. You can specify the threshold by using a percentage. You can also set an exact value and specify a unit.

Example 5-4 shows the warning threshold set to 750 GB for FCM-Pool.

Example 5-4 Changing the warning threshold level by using the CLI

```
IBM_IBM FlashSystem 7200:superuser>chmdiskgrp -warning 750 -unit gb FCM-Pool
IBM_IBM FlashSystem 7200:superuser>
```

Easy Tier Overallocation Limit window

If the system contains self-compressing drives (IBM FlashCore Module (FCM) drives) in the top tier of storage in a pool with multiple tiers and Easy Tier is in use, consider setting an overallocation limit within these pools.

Easy Tier migrates storage only at a slow rate, which might not keep up with changes to the compression ratio within the tier. This situation might result in the tier running out of space, which can cause a loss of access to data until the condition is resolved.

Therefore, the user might specify the maximum overallocation ratio for pools that contain self-compressing arrays to prevent out-of-space scenarios, as shown in Figure 5-12.

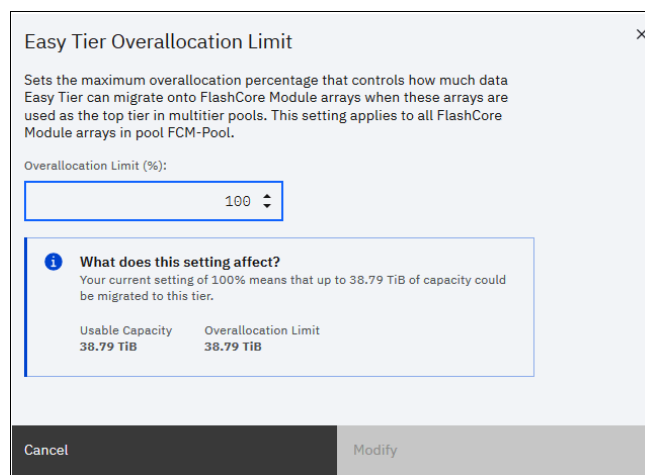


Figure 5-12 Modifying the pool overallocation limit

The value acts as a multiplier of the physically available space in self-compressing arrays. The allowed values are a percentage 100% (default) - 400%, or off. The default setting prevents overallocation of new pools. Setting the value to off disables this feature.

On the CLI, run the **chmdiskgrp** command with the **-etfcmoverallocationmax** parameter to set a percentage or use **off** to disable the limit.

For more information and a more detailed explanation, see Chapter 9, “Advanced features for storage efficiency” on page 509.

Add Storage to Pool window

This action starts the configuration wizard, which assigns storage to the pool, as shown in Figure 5-13.

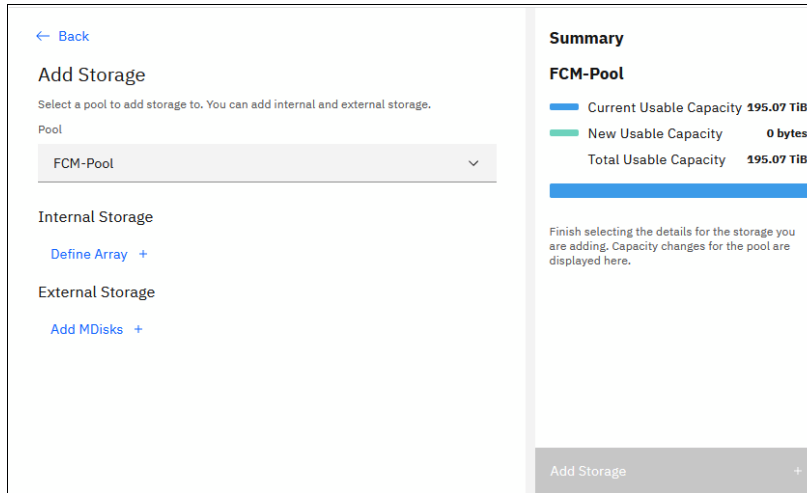


Figure 5-13 Add Storage to Pool wizard

If **Internal Storage** is chosen, the system guides you through array MDisk creation by using internal drives. If **External Storage** is selected, the system guides you through the selection of external storage MDisks. If no external storage is attached or the External Virtualization license is zero, the **External Storage** option is not shown. You can add internal and external storage for a single pool in the configuration dialog.

Figure 5-14 shows an example for internal and external selection.

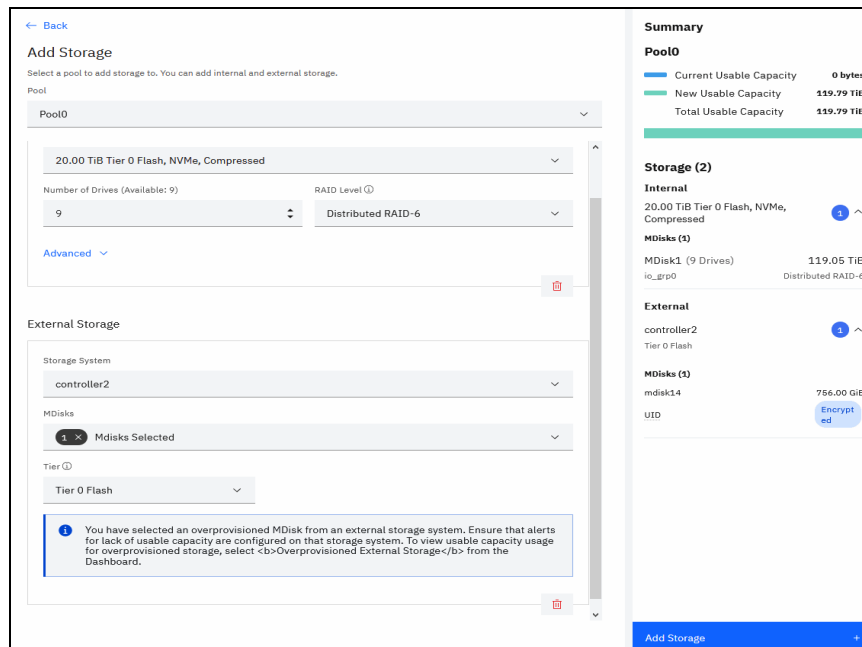


Figure 5-14 Add Storage dialog with internal and external storage selection

Edit Throttle for Pool window

Click this option to access the window where you set the pool's throttle configuration. Throttles can be defined for storage pools to control I/O operations. If a throttle limit is defined, the system either processes the I/O for that object or delays the processing of the I/O. Resources become free for more critical I/O operations.

You can use storage pool throttles to avoid overwhelming the back-end storage. Only parent pools support throttles because only parent pools contain MDisks from internal or external back-end storage. For volumes in child pools, the throttle of the parent pool is applied.

You can define a throttle for input/output operations per second (IOPS), bandwidth, or both, as shown in Figure 5-15:

- ▶ **IOPS limit** indicates the limit of configured IOPS (for both reads and writes combined).
- ▶ **Bandwidth limit** indicates the bandwidth limit in megabytes per second (MBps). You can also specify the limit in gigabits per second (Gbps) or terabytes per second (TBps).

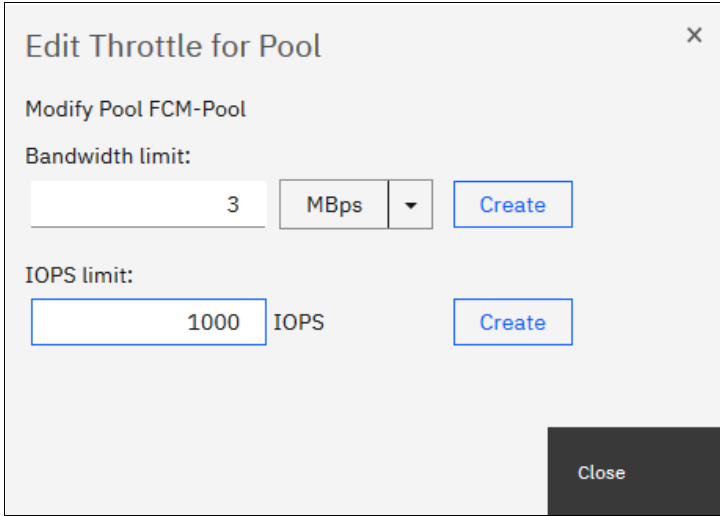


Figure 5-15 Edit Throttle for Pool window

If more than one throttle applies to an I/O operation, the lowest and most stringent throttle is used. For example, if a throttle of 100 MBps is defined on a pool and a throttle of 200 MBps is defined on a volume of that pool, the I/O operations are limited to 100 MBps.

The throttle limit is a per node limit. For example, if a throttle limit is set for a volume at 100 IOPS, each node on the system that has access to the volume allows 100 IOPS for that volume. Any I/O operation that exceeds the throttle limit is queued at the receiving nodes. The multipath policies on the host determine how many nodes receive I/O operations and the effective throttle limit.

If a throttle exists for the storage pool, the dialog checkbox that is shown in Figure 5-15 also shows the **Remove** button that is used to delete the throttle.

To set a storage pool throttle by using the CLI, run the **mkthrottle** command. Example 5-5 shows a storage pool throttle, named `iops_bw_limit`, that is set to 3 megabits per second (Mbps) and 1000 IOPS on Pool0.

Example 5-5 Setting a storage pool throttle by using the CLI

```
IBM_IBM FlashSystem 7200:superuser>mkthrottle -type mdiskgrp -iops 1000 -bandwidth 3 -mdiskgrp FCM-Pool
Throttle, id [0], successfully created.
```

To remove a throttle by using the CLI, run the **rmthrottle** command. The command uses the throttle ID or throttle name as an argument, as shown in Example 5-6. The command returns no feedback if it runs successfully.

Example 5-6 Removing a pool throttle by using a CLI

```
IBM_IBM FlashSystem 7200:superuser>rmthrottle 0
IBM__IBM FlashSystem 7200:superuser>
```

View All Throttles window

You can display the defined throttles by using the Pools window. Right-click a pool and select **View all Throttles** to display the list of the pool's throttles. If you want to view the throttle of other elements (like **Volumes** or **Hosts**, for example), you can select **All Throttles** in the drop-down list, as shown in Figure 5-16.

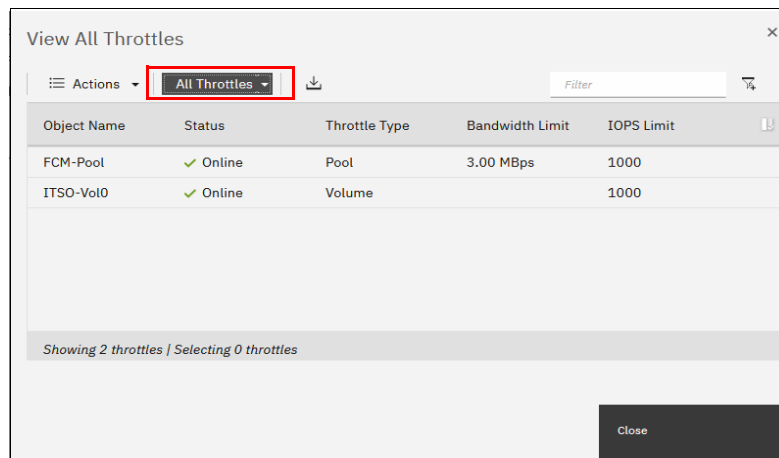
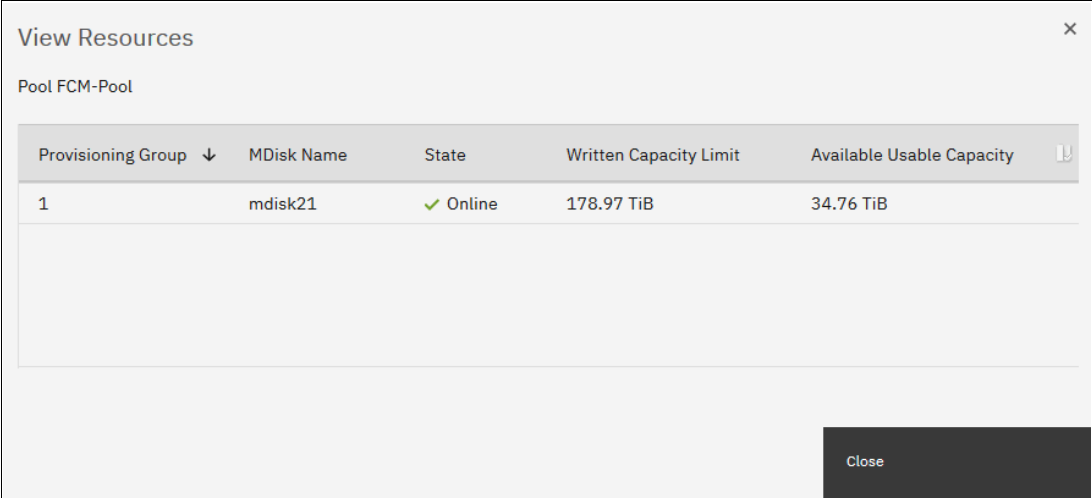


Figure 5-16 Viewing all throttles

To see a list of created throttles by using the CLI, run the **lsthrottle** command. When you run the command without arguments, it displays a list of all throttles on the system. To list only storage pool throttles, specify the **-filtervalue throttle_type=mdiskgrp** parameter.

View Resources window

To browse a list of MDisks that are part of the storage pool, click **View Resources**, which opens the window that is shown in Figure 5-17.



The screenshot shows a window titled "View Resources" with a close button in the top right corner. Below the title bar, it says "Pool FCM-Pool". There is a table with the following columns: "Provisioning Group" (with a dropdown arrow), "MDisk Name", "State", "Written Capacity Limit", and "Available Usable Capacity" (with a sort icon). The table contains one row with the following data: Provisioning Group: 1, MDisk Name: mdisk21, State: Online (with a green checkmark icon), Written Capacity Limit: 178.97 TiB, and Available Usable Capacity: 34.76 TiB. At the bottom right of the window is a "Close" button.

Provisioning Group ↓	MDisk Name	State	Written Capacity Limit	Available Usable Capacity
1	mdisk21	✓ Online	178.97 TiB	34.76 TiB

Figure 5-17 List of resources in the storage pool

To list storage pool resources by using the CLI, run the **lsmdisk** command. You can filter the output to display MDisk objects that belong only to a single MDisk group (storage pool), as shown in Example 5-7.

Example 5-7 Using lsmdisk (some columns are not shown)

```
IBM_IBM FlashSystem 7200:superuser>lsmdisk -filtervalue mdisk_grp_name=FCM-Pool
id name status mode mdisk_grp_id mdisk_grp_name capacity ctrl_LUN_#
32 mdisk21 online array 2 FCM-Pool 179TB
```

View Easy Tier Reports

View the most recent Easy Tier statistics. For more information about Easy Tier Reports, see 9.1.3, “Monitoring Easy Tier activity” on page 523.

Deleting a storage pool

A storage pool can be deleted by using the GUI only if no volumes are associated with it. Select **Delete** to delete the pool immediately without any additional confirmation.

If there are volumes in the pool, the **Delete** option is inactive and cannot be selected. Delete the volumes or migrate them to another storage pool before proceeding. For more information about volume migration and volume mirroring, see Chapter 6, “Volumes” on page 299.

After you delete a pool, the following actions occur:

- ▶ All the external MDisks in the pool return to a mode of *Unmanaged*.
- ▶ All the array mode MDisks in the pool are deleted and all member drives return to a status of *Candidate*.

To delete a storage pool by using the CLI, run the **rmmdiskgrp** command.

Note: Be *extremely* careful when you run the `rmmdiskgrp` command with the `-force` parameter. Unlike the GUI, it does not prevent you from deleting a storage pool with volumes. This command deletes all volumes and host mappings on a storage pool, and they *cannot be recovered*.

Properties for Pool window

Select **Properties** to display information about the storage pool. By hovering your cursor over the elements of the window and clicking **[?]**, you see a short description of each property, as shown in Figure 5-18.

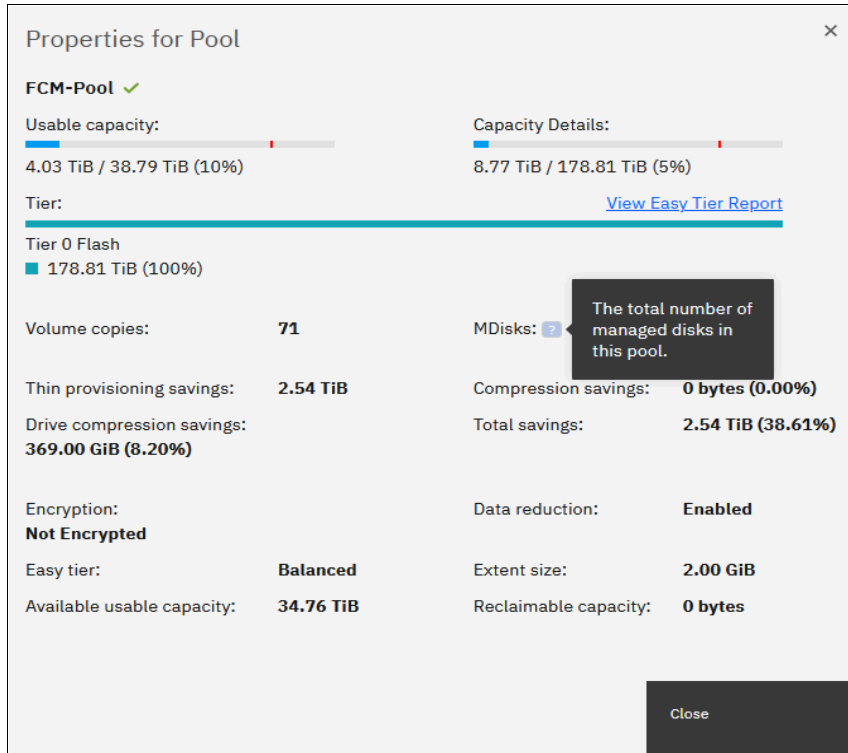


Figure 5-18 Pool properties and details

To display detailed information about the properties by using the CLI, run the `lsmdiskgrp` command with a storage pool name or ID as a parameter, as shown in Example 5-8.

Example 5-8 The `lsmdiskgrp` output (partially shown)

```
IBM_IBM FlashSystem 7200:superuser>lsmdiskgrp FCM-Pool
id 2
name FCM-Pool
status online
mdisk_count 1
vdisk_count 71
capacity 178.81TB
extent_size 2024
free_capacity 6.57TB
<...>
```

5.1.4 Child pools

A *child pool* is a storage pool that is created within another storage pool. The storage pool in which the child storage pool is created is called the *parent storage pool*.

A storage pool type is parent, child_thick, or child_quotaless.

Unlike a parent pool, a child pool does not contain MDisks. Its capacity is provided by the parent pool. A child pool from a standard parent pool is the child_thick type. The capacity of a child pool from a standard pool is set at creation time, but can be modified later nondisruptively. The capacity must be a multiple of the parent pool extent size and smaller than the free capacity of the parent pool. Capacity that is assigned to a child pool of the child_thick type is taken away from the capacity of the parent pool.

A child pool from a data reduction parent pool is the child_quotaless type. It is not possible to set the capacity for a child pool of the child_quotaless type. A child pool of the child_quotaless type can use the whole capacity of the parent pool due to the nature of DRPs.

Creating a child pool within another child pool also is not possible.

Child pools of the child_thick type are useful when the capacity that is allocated to a specific set of volumes must be controlled. For example, child pools of the child_thick type can be used with VMware vSphere Virtual Volumes (VVOLs). Storage administrators can restrict the access of VMware administrators to only a part of the storage pool and prevent volume creation from affecting the rest of the parent storage pool.

Ownership groups can be used to restrict access to storage resources to a specific set of users, as described in Chapter 11, “Ownership groups” on page 723.

Child pools of the child_thick type also can be useful when strict control over thin-provisioned volume expansion is needed. For example, you might create a child pool with no volumes in it to act as an emergency set of extents so that if the parent pool ever runs out of free extents, you can use the ones from the child pool.

On systems with encryption enabled, child pools of the child_thick type can be created to migrate existing volumes in a non-encrypted pool to encrypted child pools. When you create a child pool of the child_thick type after encryption is enabled, an encryption key is created for the child pool even when the parent pool is not encrypted. You can then use volume mirroring to migrate the volumes from the non-encrypted parent pool to the encrypted child pool. Encrypted child pools of the quotaless type can be created only if the parent pool is encrypted. The data reduction child pool inherits an encryption key from the parent pool.

Child pools inherit most properties from their parent pools, and these properties cannot be changed. The inherited properties include:

- ▶ Extent size
- ▶ Easy Tier setting

Also, a child data reduction pool inherits the encryption setting and encryption key from a parent data reduction pool.

Creating a child storage pool

To create a child pool, complete the following steps:

1. Select **Pools** → **Pools**, right-click the parent pool that you want to create a child pool from, and select **Create Child Pool**, as shown in Figure 5-19 on page 253.

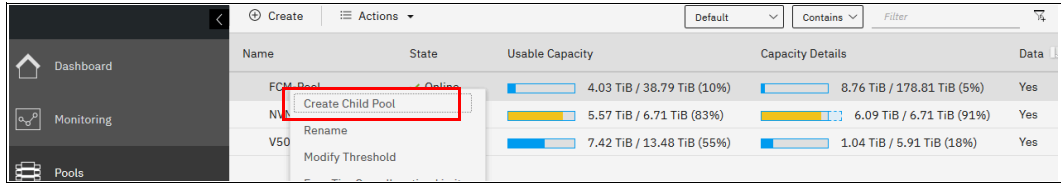


Figure 5-19 Creating a child pool

- When the dialog box opens, enter the name of the child pool and click **Create**. Figure 5-20 shows the dialog for pool type `child_quotaless`.

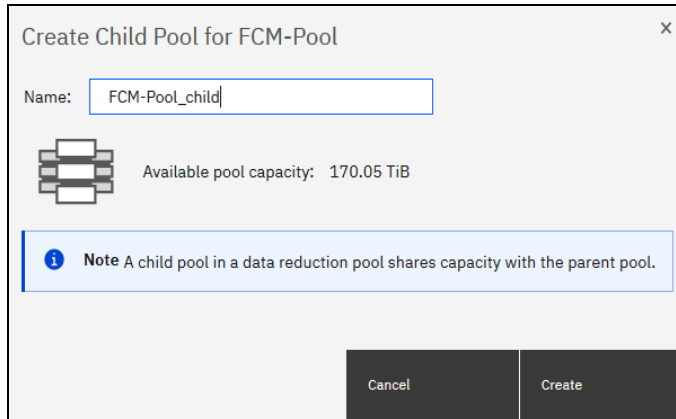


Figure 5-20 Creating a child pool type `child_quotaless`

- Figure 5-21 shows the Create Child Pool for Pool0 dialog box. For pool type `child_thick`, enter the pool capacity.

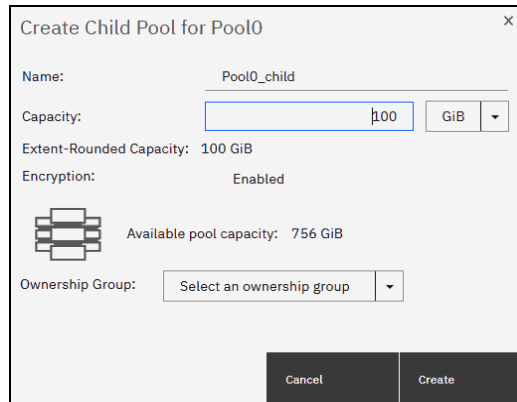


Figure 5-21 Creating a child pool type `child_thick`

After the child pool is created, it is listed in the Pools window under its parent pool. Toggle the sign to the left of the storage pool icon to either show or hide the child pools, as shown in Figure 5-22. The capacity that is assigned to the child pools is not usable in the parent pool, as shown by the gray area on the capacity details bar of the parent pool.

Name	State	Usable Capacity	Capacity Details
FCM-Pool	Online	4.03 TiB / 38.79 TiB (10%)	8.77 TiB / 178.81 TiB (5%)
FCM-Pool_child	Online		
FCM-Pool_child0	Online		
NVMe-Pool0	Online	5.57 TiB / 6.71 TiB (83%)	6.09 TiB / 6.71 TiB (91%)
Pool0	Online	7.42 TiB / 43.58 TiB (17%)	50.00 TiB / 141.93 TiB (35%)
V5030_Pool	Online	7.42 TiB / 13.48 TiB (55%)	1.01 TiB / 2.95 TiB (34%)

Figure 5-22 Listing parent and child pools

To create a child pool by using the CLI, run the `mkmdiskgrp` command. You must specify the parent pool for your new child pool and its size for pool type `child_thick`, as shown in Example 5-9. The size is in megabytes by default (unless the `-unit` parameter is used) and must be a multiple of the parent pool's extent size. In this case, it is $100 * 1024 \text{ MB} = 100 \text{ GB}$.

Example 5-9 The `mkmdiskgrp` command to create child pools

```
IBM_IBM FlashSystem 7200:superuser>mkmdiskgrp -parentmdiskgrp Pool0 -size 102400
-name Pool0_child0
MDisk Group, id [4], successfully created
```

Actions for child storage pools

You can rename, resize (only child pools type `child_thick`), or delete a child pool. Also, it is possible to modify its warning threshold and assign it to an ownership group. To select an action, for example, **Resize**, complete the following steps:

1. Right-click the child storage pool, as shown in Figure 5-23. Alternatively, select the storage pool and click **Actions**.

Name	State	Usable Capacity	Capacity Details	Data Reduction
FCM-Pool	Degraded	4.03 TiB / 38.79 TiB (10%)	8.76 TiB / 178.81 TiB (5%)	Yes
FCM-Pool_child	Degraded			Yes
NVMe-Pool0	Online	5.57 TiB / 6.71 TiB (83%)	6.09 TiB / 6.71 TiB (91%)	Yes
Pool0	Online	7.42 TiB / 43.58 TiB (17%)	50.00 TiB / 141.93 TiB (35%)	No
Pool0_child0	Online		0 bytes / 50.00 TiB (0%)	No
V5030_Pool	Online	7.42 TiB / 13.48 TiB (55%)	1.01 TiB / 2.95 TiB (34%)	Yes

Figure 5-23 Actions for child storage pools

2. Select **Resize** to increase or decrease the capacity of the child storage pool type `child_thick`, as shown in Figure 5-24 on page 255. Enter the new pool capacity and click **Resize**.

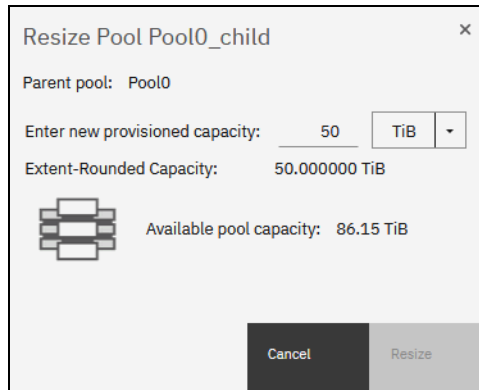


Figure 5-24 Resizing a child pool

Note: You cannot shrink a child pool below its real capacity. Thus, the new size of a child pool must be larger than the capacity that is used by its volumes.

When the child pool is shrunk, the system resets the warning threshold and issues a warning if the threshold is reached.

To rename and resize child pool by using the CLI, run the `chmdiskgrp` command. Example 5-10 renames the child pool `Pool0_child0` to `Pool0_child_new` and reduces its size to 44 GB. If successful, the command returns no feedback.

Example 5-10 Running the `chmdiskgrp` command to rename a child pool

```
IBM_IBM FlashSystem 7200:superuser>chmdiskgrp -name Pool0_child_new -size 45056
Pool0_child0
IBM_Storwize:ITS0V7K:superuser>
```

Deleting a child pool is a task that is like deleting a parent pool. As with a parent pool, the **Delete** action is disabled if the child pool contains volumes, as shown in Figure 5-25.

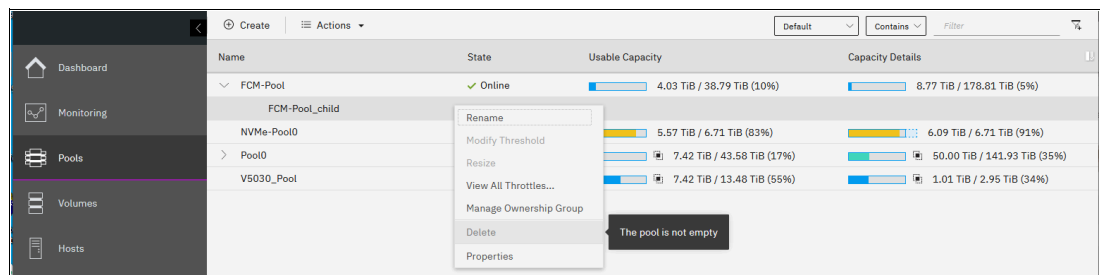


Figure 5-25 Deleting a child pool

After you delete a child pool type `child_thick`, the extents that it occupied return to the parent pool as free capacity.

To delete a child pool by using the CLI, run the `rmmdiskgrp` command.

To assign an existing ownership group to a child pool, click **Manage Ownership Group**, as shown in Figure 5-26. All volumes that are created in the child pool inherit the ownership group of the child pool. For more information, see Chapter 11, “Ownership groups” on page 723.

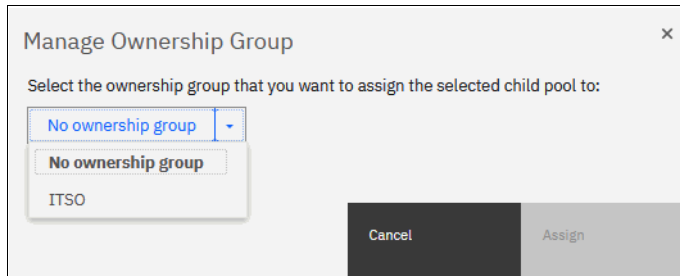


Figure 5-26 Managing the ownership group of a child pool

Migrating volumes to and from child pools

To move a volume to another pool, you can use migration or volume mirroring in the same way that you use them for parent pools. For more information about volume migration and volume mirroring, see Chapter 6, “Volumes” on page 299.

The system supports migration of volumes between child pools within the same parent pool or migration of a volume between a child pool and its parent pool. Migrations between a source and target child pool with different parent pools are not supported. However, you can migrate the volume from the source child pool to its parent pool. Then, the volume can be migrated from the parent pool to the parent pool of the target child pool. Finally, the volume can be migrated from the target parent pool to the target child pool.

During a volume migration within a parent pool (between a child and its parent or between children with the same parent), there is no data movement, but there are extent reassignments.

Volume migration between a child storage pool and its parent storage pool can be performed by going to the **window** page and clicking **Volumes**. Right-click a volume and select it to migrate it into a suitable pool.

In the example in Figure 5-27, the volume `child_volume` was created in child pool `Pool0_child0`. The child pools appear exactly like the parent pools in the Volumes by Pool window.

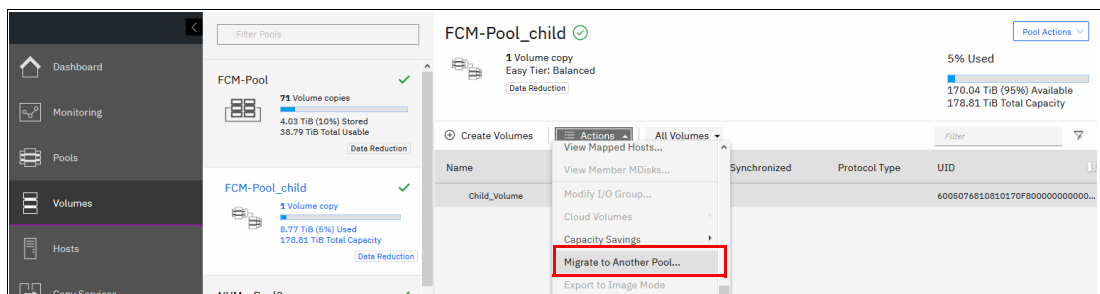


Figure 5-27 Actions menu in Volumes by Pool

For more information about the CLI commands for migrating volumes to and from child pools, see Chapter 6, “Volumes” on page 299.

5.1.5 Encrypted storage pools

The system supports two types of encryption: hardware encryption and software encryption.

Hardware encryption is implemented at an array level, and software encryption is implemented at a storage pool level. For more information about encryption and encrypted storage pools, see Chapter 12, “Encryption” on page 735.

5.2 Working with internal drives and arrays

An array is a type of MDisk that is made up of disk drives (or flash modules); these drives are members of the array. RAID is a method of configuring member drives to create high availability (HA) and high-performance groupings of drives. The system supports nondistributed (traditional) and distributed redundant array of independent disks (DRAID) configurations.

5.2.1 Working with drives

This section describes how to manage internal storage disk drives and configure them to be used in arrays.

Listing disk drives

The system provides an Internal Storage window for managing all internal drives. To access the Internal Storage window, select **Pools** → **Internal Storage**, as shown in Figure 5-28.

The screenshot displays the 'All Internal Storage' window. On the left, a navigation menu includes Dashboard, Monitoring, Pools, Volumes, Hosts, Copy Services, Access, and Settings. The 'Pools' menu is expanded, showing 'Internal Storage' selected. The main area shows a summary of internal storage usage: 241.44 TiB (55%) Assigned, 441.43 TiB Total Written Capacity. A red box highlights the following statistics: 55% Assigned, 241.44 TiB (55%) Assigned to MDisks, 0 bytes (0%) Assigned to Spares, 199.99 TiB (45%) Available, and 441.43 TiB Total Written Capacity Limit. Below the summary is a table of internal drives:

Drive ID	Written C...	Use	Status	MDisk Name	Member ID
0	20.00 TiB	Member	Online	mdisk21	0
1	744.21 GiB	Member	Online	mdisk0	1
2	744.21 GiB	Member	Online	mdisk0	0
3	20.00 TiB	Member	Online	mdisk21	1
4	20.00 TiB	Member	Online	mdisk21	2
5	20.00 TiB	Member	Online	mdisk21	3
6	20.00 TiB	Member	Online	mdisk21	4
7	20.00 TiB	Member	Online	mdisk21	5
8	20.00 TiB	Member	Online	mdisk21	6
9	20.00 TiB	Member	Online	mdisk21	7
10	20.00 TiB	Member	Online	mdisk21	8
11	20.00 TiB	Member	Online	mdisk21	9
12	20.00 TiB	Member	Online	mdisk21	10
13	20.00 TiB	Member	Online	mdisk21	11
14	20.00 TiB	Candidate	Online		

Showing 24 drives | Selected 0 drives

Figure 5-28 Internal Storage window

This pane gives an overview of the internal drives in the system. To display all drives that are managed in the system, including all I/O groups and expansion enclosures, click **All Internal Storage** in the **Drive Class** filter.

Alternatively, you can filter the drives by their type or class. For example, you can choose to show only enterprise drives, nearline (NL) drives, or flash drives. Select the class on the left side of the window to filter the list and display only the drives of the selected class.

You can find information about the capacity allocation of each drive class in the upper right, as shown in Figure 5-28 on page 257:

Assigned to MDisks	Shows the storage capacity of the selected drive class that is assigned to MDisks.
Assigned to Spares	Shows the storage capacity of the selected drive class that is used for spare drives.
Available	Shows the storage capacity of the selected drive class that is not yet assigned to either MDisks or Spares.
Total Written Capacity Limit	Shows the total amount of storage capacity of the drives in the selected class.

If **All Internal Storage** is selected under the Drive Class filter, the values that are shown refer to the entire internal storage.

The percentage bar indicates how much of the total written capacity limit is assigned to MDisks and spares. MDisk capacity is represented by the solid portion, and spare capacity by the shaded portion of the bar.

To list all internal drives that are available in the system, run the `lsdrive` command. If needed, you can filter output to list only drives that belong to particular enclosure, that have specific capacity, or by other attributes. For an example, see Example 5-11.

Example 5-11 The lsdrive output (some lines and columns are not shown)

```

IBM_IBM FlashSystem 7200:superuser>lsdrive
id status error_sequence_number use tech_type capacity mdisk_id
0 online member tier0_flash 20TB 32
1 online member tier0_flash 744.21GB 0
2 online member tier0_flash 744.21GB 0
3 online member tier0_flash 20TB 32
4 online member tier0_flash 20TB 32
5 online member tier0_flash 20TB 32
<...>

```

The drive list shows the Status of each drive. A drive can be `Online`, which means that the drive is fully accessible by both nodes in the I/O group. A `Degraded` drive is only accessible by one of the two nodes. A drive status of `Offline` indicates that the drive is not accessible by any of the nodes, for example, because it was physically removed from the enclosure or it is unresponsive or failing.

The drive Use attribute describes the role that it plays in the system. The values and meanings are:

Unused	The system has access to the drive but was not told to take ownership of it. Most actions on the drive are not permitted. This state is a safe state for newly added hardware.
Candidate	The drive is owned by the system, and is not part of the RAID configuration. It is available to be used in an array MDisk.

Spare	The drive is a hot spare protecting nondistributed (traditional) RAID arrays. If any member of such an array fails, a spare drive is taken and becomes a Member for rebuilding the array.
Member	The drive is part of a RAID array.
Failed	The drive is owned by the system and was diagnosed as faulty. It is waiting for a service action.

The Use attribute can change to different values, but not all changes are valid, as shown in Figure 5-29.

		To				
		unused	candidate	failed	member	spare
From	unused	yes	yes	no	no	no
	candidate	yes	yes	yes	no	yes
	failed	yes	yes	yes	no	no
	member	no	no	yes	no	no
	spare	no	yes	yes	no	yes

Figure 5-29 Drive use changes

The system automatically sets the Use to Member when it creates a RAID array. Changing Use from Member to Failed is possible only if the array does not depend on the drive, and additional confirmation is required when taking a drive offline when no spare is available. Changing a Candidate drive to Failed is possible only by using the CLI.

Note: To start configuring arrays in a new system, all Unused drives must be configured as Candidates. The Initial Setup or Assign Storage wizards do that automatically.

A number of actions can be performed on internal drives. To perform any action, select one or more drives and right-click the selection, as shown in Figure 5-30. Alternatively, select the drives and click **Actions**.

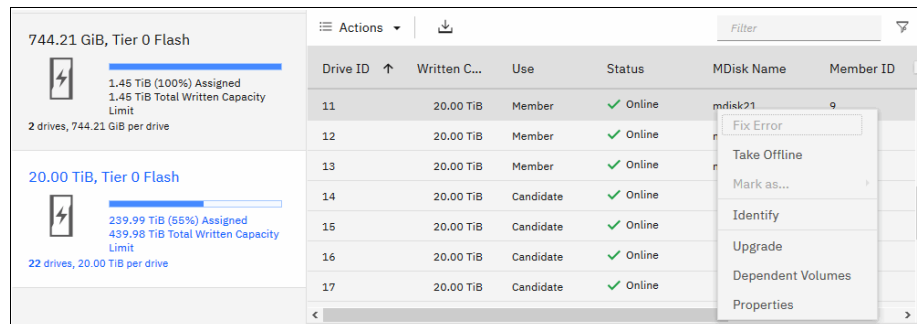


Figure 5-30 Actions for internal storage

The actions that are available in the drop-down menu depend on the status and usage of the drive or drives that are selected. Some actions can be performed only on drives in a certain state, and some are possible only when a single drive is selected.

Action: Fix Error

This action is available only if the drive that is selected has an error event that is associated with it. Select **Fix Error** to start the directed maintenance procedure (DMP) for the selected drive. For more information about DMPs, see Chapter 13, “Reliability, availability, and serviceability, monitoring and logging, and troubleshooting” on page 793.

Action: Take Offline

If a problem is identified with a specific drive, you can select **Take Offline** to take the drive offline. You must confirm the action, as shown in Figure 5-31.

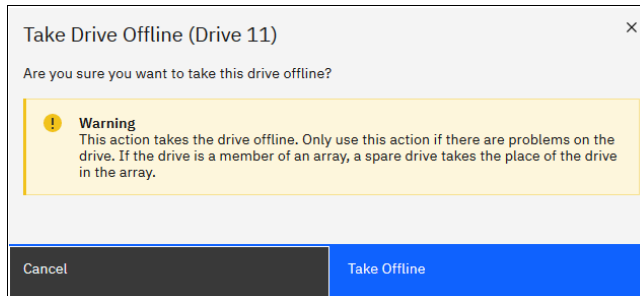


Figure 5-31 Taking a drive offline if a spare or rebuild area is available

Figure 5-32 shows the message that appears if the action results in a degraded array status.

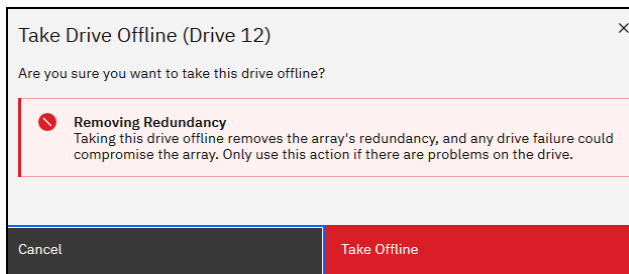


Figure 5-32 Taking drive offline leading to degraded redundancy

The system prevents you from taking the drive offline if taking the drive offline results in a loss of access to data.

If a spare is available and the drive is taken offline, the associated MDisk remains **Online** and the RAID array starts a rebuild by using a suitable spare. If no spare is available and the drive is taken offline, the status of the associated MDisk becomes **Degraded**. The status of the storage pool to which the MDisk belongs becomes **Degraded** too.

A drive that is taken offline is considered **Failed**, as shown in Figure 5-33.

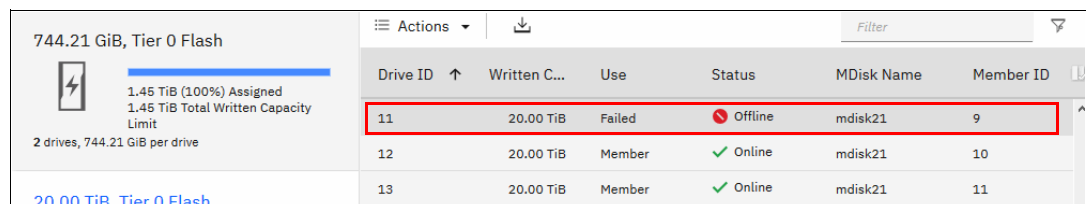


Figure 5-33 An offline drive is marked as Failed

To take a drive offline by using the CLI, run the `chdrive` command, as shown in Example 5-12. This command returns no feedback. Use the `-allowdegraded` parameter to set a member drive offline even if no suitable spare is available.

Example 5-12 Setting a drive offline by using the CLI

```
IBM_IBM FlashSystem 7200:superuser>chdrive -use failed 11
IBM_IBM FlashSystem 7200:superuser>
```

The system prevents you from taking a drive offline if the RAID array depends on that drive and doing so would result in a loss of access to data, as shown in Figure 5-34.

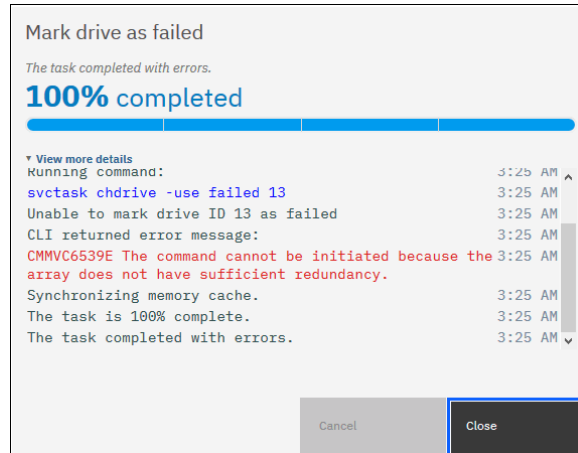


Figure 5-34 Taking a drive offline fails if it would result in a loss of access to data

Action: Mark as

Select **Mark as** to change the use that is assigned to the drive, as shown in Figure 5-35. The list of available options depends on the current drive use and state. For more information, see the allowed state transitions that are shown in Figure 5-35.

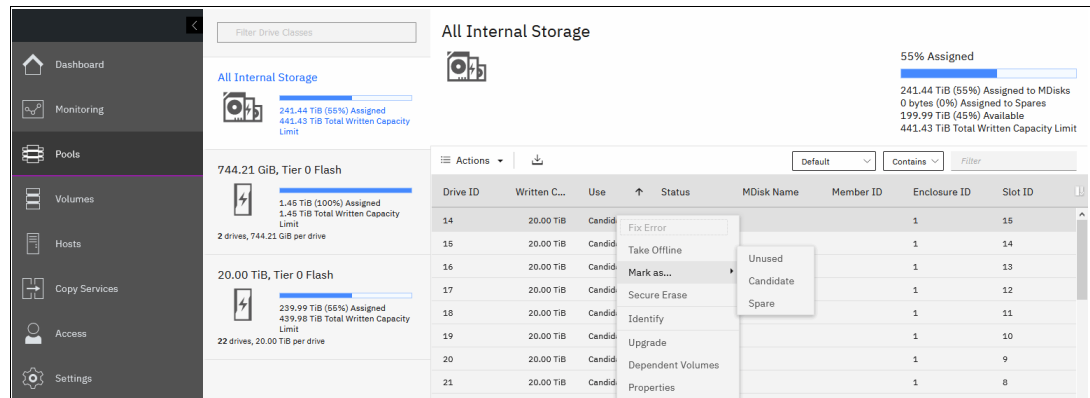


Figure 5-35 A drive can be marked as Unused, Candidate, or Spare

To change the drive role by using the CLI, run the **chdrive** command, as shown in Example 5-13. It shows that the drive that was set offline by a previous command is set as a spare. The drive cannot go from Failed to Spare in one step. Instead, the drive must be assigned to a Candidate role before it is set to the Spare role. DRAIDs do not use spares. It is not possible to mark drives as spares that are supported only in DRAIDs.

Example 5-13 Changing the drive role by using the CLI

```
IBM_IBM FlashSystem 7200:superuser>lsdrive -filtervalue status=offline
id status error_sequence_number use tech_type capacity mdisk_id
3 offline                               failed tier_enterprise 1.1TB
IBM_IBM FlashSystem 7200:superuser>chdrive -use spare 3
CMMVC6537E The command cannot be initiated because the drive that you have specified has a Use
property that is not supported for the task.
IBM_IBM FlashSystem 7200:superuser>chdrive -use candidate 3
IBM_IBM FlashSystem 7200:superuser>chdrive -use spare 3
IBM_IBM FlashSystem 7200:superuser>
```

Note: Marking a compressed drive to the Candidate role causes the drive to perform a format. The format must complete before the drive goes online and is available for use.

Action: Identify

Select **Identify** to turn on the light-emitting diode (LED) light of the enclosure slot of the selected drive. With this action, you can easily find a drive that must be replaced or that you want to troubleshoot. A dialog box opens so that you can confirm that the LED was turned on, as shown in Figure 5-36.

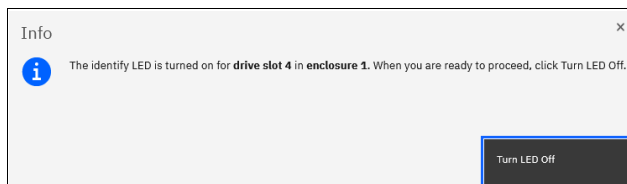


Figure 5-36 Identifying an internal drive

Your action makes an amber LED flash (turn on and off continuously) for the drive that you want to identify.

Click **Turn LED Off** when you are finished. The LED returns to its initial state.

On the CLI, run the **chenclosureslot** command to turn on the LED. Example 5-14 shows the commands to find the enclosure and slot for drive 1 and to turn on and off the identification LED of slot 3 in enclosure 1.

Example 5-14 Changing a slot LED to identification mode by using the CLI

```
IBM_IBM FlashSystem 7200:superuser>lsdrive 1
id 21
<...>
enclosure_id 1
slot_id 4
<...>
IBM_IBM FlashSystem 7200:superuser>chenclosureslot -identify yes -slot 4 1
IBM_IBM FlashSystem 7200:superuser>lenclosureslot -slot 4 1
enclosure_id 1
slot_id 4
fault_LED slow_flashing
```

```
powered yes
drive_present yes
drive_id 1
IBM_IBM FlashSystem 7200:superuser>chenclosureslot -identify no -slot 4 1
```

Action: Upgrade

Select **Upgrade** to update the drive firmware, as shown in Figure 5-37. You can choose to update an individual drive, selected drives, or all the drives in the system.

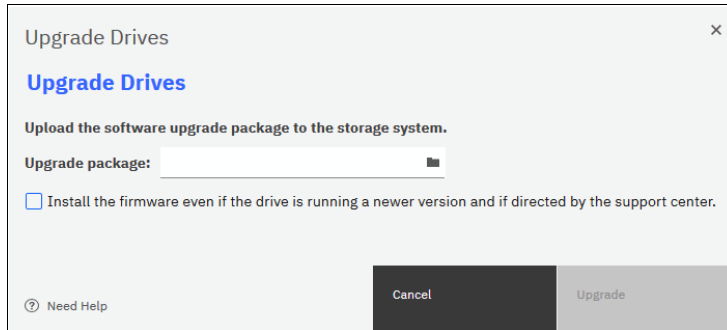


Figure 5-37 Upgrading a drive or a set of drives

For information about updating the drive firmware, see Chapter 13, “Reliability, availability, and serviceability, monitoring and logging, and troubleshooting” on page 793.

Action: Dependent Volumes

Select **Dependent Volumes** to list the volumes that depend on the selected drives. A volume depends on a drive or a set of drives when removal or failure of that drive or set of drives results in a loss of access or a loss of data for that volume. Use this option before you do maintenance operations to confirm which volumes (if any) will be affected.

Figure 5-38 shows the list of volumes that depend on a set of three drives that belong to the same MDisk.

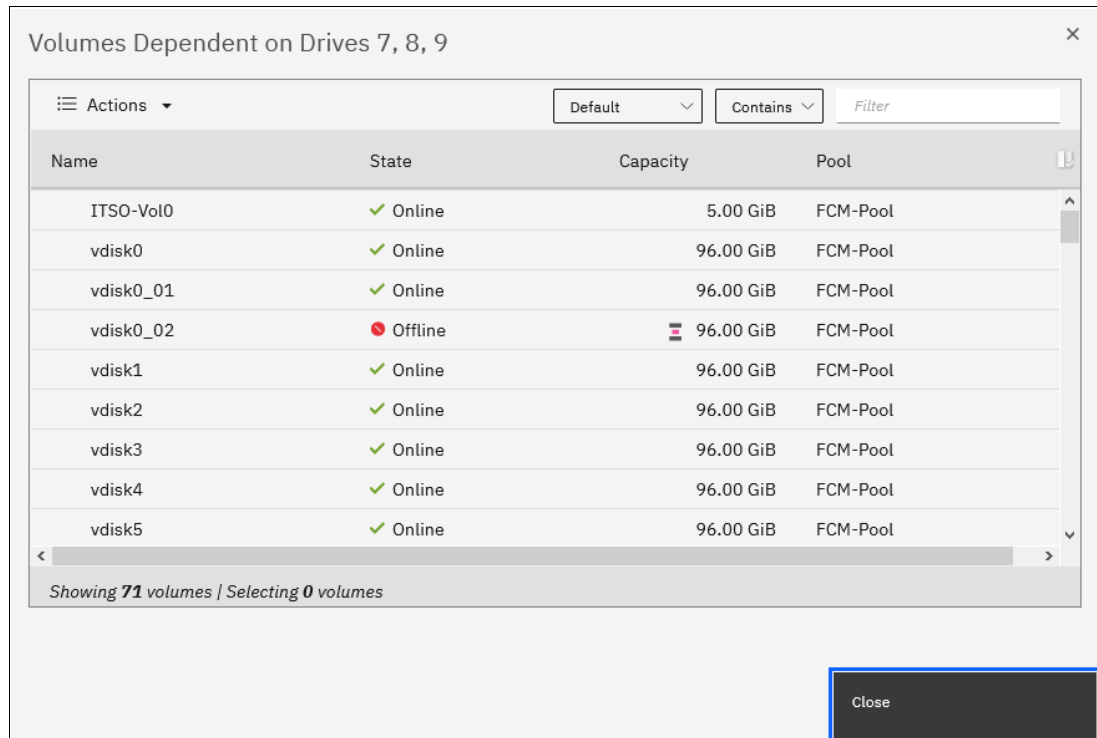


Figure 5-38 List of volumes that depend on disks 7, 8, and 9

All listed volumes go offline if *all* selected drives go offline concurrently. This situation does *not* mean that volumes go offline if a single drive or two of the three drives go offline.

Whether there are dependent volumes depends on the redundancy of the RAID array at a certain point. The redundancy is based on the RAID level, state of the array, and state of the other member drives in the array. For example, it takes three or more drives going offline concurrently in a healthy RAID 6 array to have dependent volumes.

Note: A lack of dependent volumes does not imply that there are no volumes that use the drive. Volume dependency shows the list of volumes that become unavailable if the drive or the set of selected drives becomes unavailable.

You can get the same information by running the `lsdependentvdisks` command. Use the parameter `-drive` with the list of drive IDs that you are checking, separated with a colon (:), as shown in Example 5-15.

Example 5-15 Listing volumes that depend on drives

```
IBM_IBM FlashSystem 7200:superuser>lsdependentvdisks -drive 7:8:9
vdisk_id vdisk_name
0        vdisk0
1        vdisk1
2        vdisk2
3        vdisk3
4        vdisk4
5        vdisk5
...
```


Action: Properties

Select **Properties** to view more information about the drive, as shown in Figure 5-39.

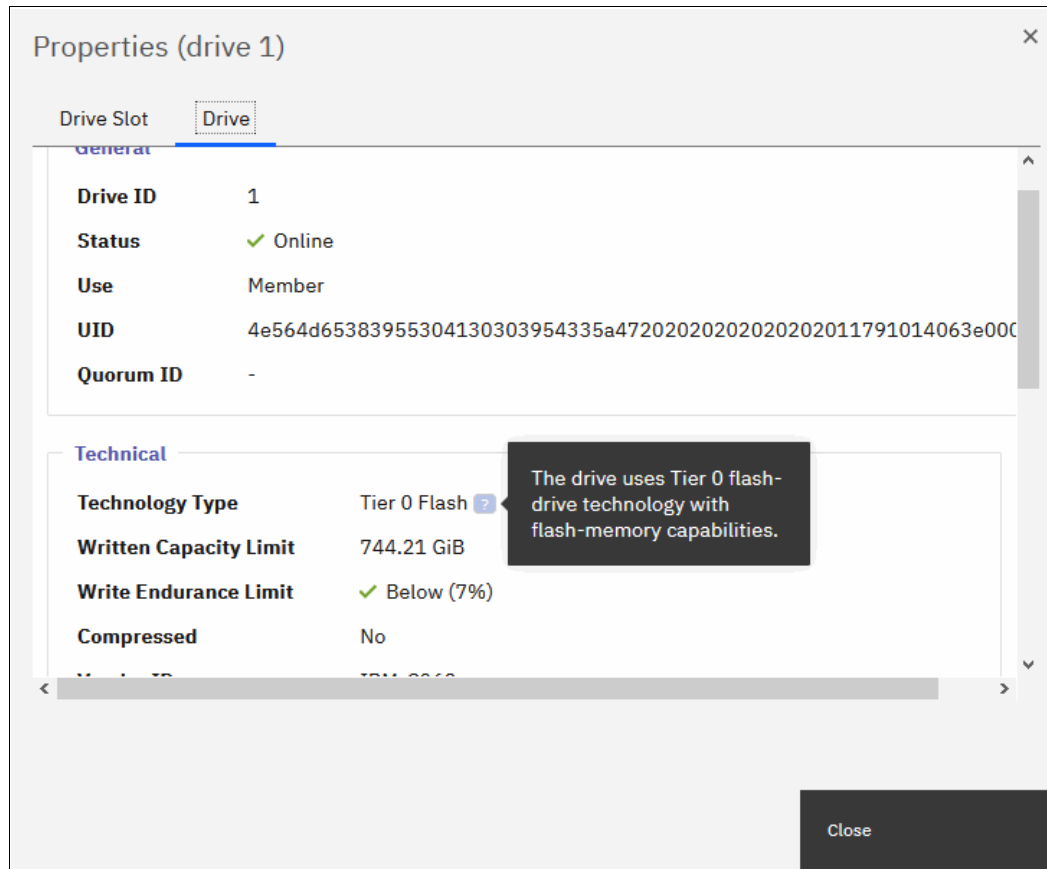


Figure 5-39 Drive properties

You can find a short description of each drive property by hovering your cursor over it and clicking [?]. You can also display drive slot details by clicking the **Drive Slot** tab.

To get all available information about the particular drive, run the `lsdrive` command with the drive ID as the parameter. To get slot information, run the `lsclosureslot` command.

5.2.2 RAID and distributed redundant array of independent disks

To use internal disks in storage pools, you must join them as RAID arrays to form array mode MDisks.

RAID provides two key design goals:

- ▶ Increased data reliability
- ▶ Increased input/output (I/O) performance

Introduction to RAID technology

RAID technology can provide better performance for data access, HA for the data, or a combination of both. RAID levels define a tradeoff between HA, performance, and cost.

When multiple physical disks are set up to use RAID, they are in a *RAID array*. The system provides multiple, traditional RAID (TRAIID) levels:

- ▶ RAID 0
- ▶ RAID 1
- ▶ RAID 10

Note: RAID 5 and RAID 6 are available only in DRAID configurations.

RAID 0 does not provide any redundancy. A single drive failure in a RAID 0 array causes data loss.

In a TRAIID approach, data is spread among up to 16 drives in an array. There are separate spare drives that do not belong to an array, and they can potentially protect multiple arrays. When one of the drives within the array fails, the system rebuilds the array by using a spare drive.

For example, in RAID 10 all data is read from the mirrored copy and then written to a spare drive. The spare becomes a member of the array when the rebuild starts. After the rebuild is complete and the failed drive is replaced, a member exchange is performed to add the replacement drive to the array and restore the spare to its original state so it can act as a hot spare again for another drive failure in the future.

During a rebuild of a TRAIID array, writes are submitted to a single spare drive, which can become a bottleneck and might impact I/O performance. With increasing drive capacity, the rebuild time increases significantly. Additionally, the probability of a second failure during the rebuild process also becomes more likely. Outside of any rebuild activity, the spare drives are idle and do not process I/O requests for the system.

DRAID addresses these shortcomings.

Distributed redundant array of independent disks

In DRAID, there are no dedicated spare drives that are idle most of the time. All 2 - 128 drives in the array process I/O requests always, which improves the overall I/O performance. Spare capacity is spread across all member drives to form one or more *rebuild areas*. During a rebuild, the write workload is distributed across all drives, removing the single drive bottleneck of traditional arrays.

Using this approach, DRAID reduces the rebuild time, the impact on I/O performance during the rebuild, and the probability of a second failure during the rebuild. Like TRAIID, a DRAID 6 array can tolerate two drive failures and survive. If another drive fails in the same array before the array is rebuilt, the MDisk and the storage pool go offline. In other words, DRAID has the same redundancy characteristics as TRAIID.

A rebuild after a drive failure reconstructs the data on the failed drive and distributes it across all drives in the array by using a rebuild area. After the failed drive is replaced, a copyback process copies the data to the replacement drive and to free the rebuild area so that it can be used for another drive failure in the future.

The following DRAID types are available:

- ▶ DRAID 1
- ▶ DRAID 5
- ▶ DRAID 6

Table 5-1 on page 267 shows the summary of supported drives, array types, and RAID levels

Note: DRAID 1 is supported only on IBM FlashSystem 7200, IBM FlashSystem 9200, or newer platforms.

Table 5-1 Summary of supported drives, array types, and RAID levels

Drive type	Non-DRAID	DRAID		
	RAID 1	DRAID 1	DRAID 5	DRAID 6
Industry standard Non-Volatile Memory Express (NVMe) drives or serial-attached Small Computer System Interface (SCSI) (SAS) drives (expansion enclosure)	x	x	x	x
FCM NVMe drives		x	x	x
Storage-class memory (SCM)		x	x	x

Note: DRAID 1 is not recommended for FCM drives larger than 8 TB. You cannot use the GUI to create DRAID 1 arrays on XL FCM drives (80 TB).

Understanding DRAID 6

Figure 5-40 shows an example of a DRAID 6 with 10 disks. The capacity on the drives is divided into many packs. The reserved spare capacity (marked in yellow) is equivalent to two spare drives, but the capacity is distributed across all of the drives (depending on the pack number) to form two rebuild areas. The data is striped like a TRAIID array, but the number of drives in the array can be larger than the stripe width.

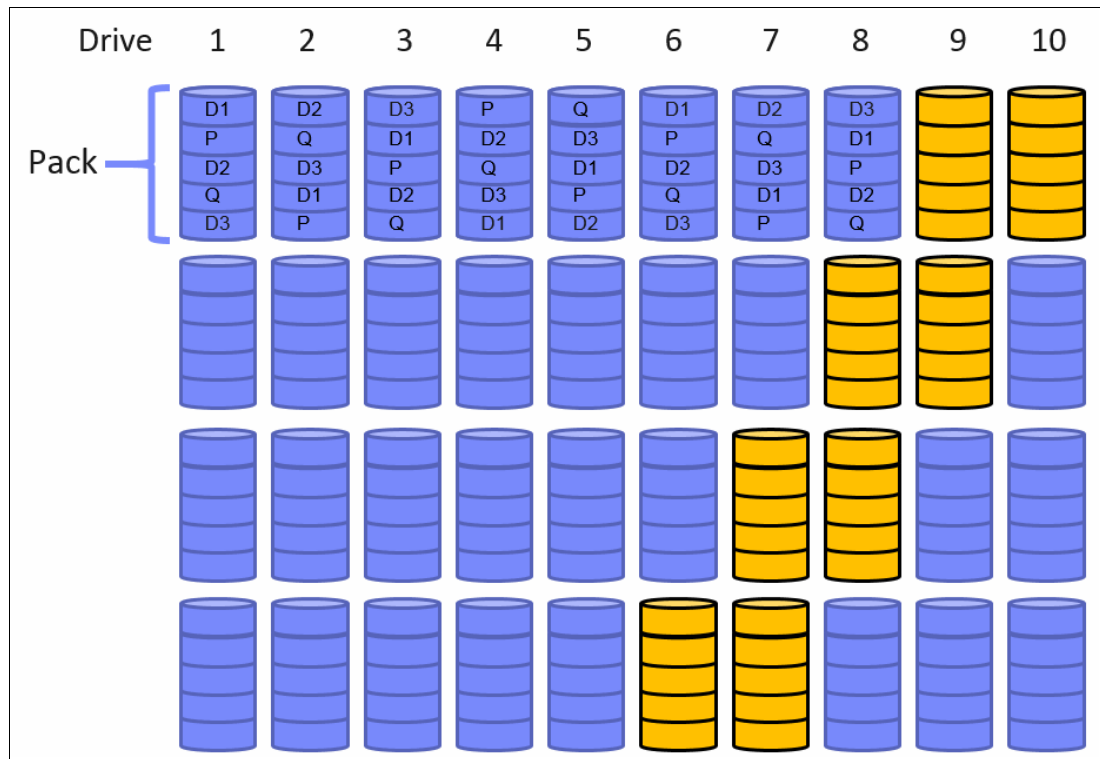


Figure 5-40 DRAID 6 (for simplification, not all packs are shown)

Figure 5-41 shows what happens after a single drive failure in this DRAID 6. Drive 3 failed and the array is using half of the spare capacity in each pack (marked in green) to rebuild the data of the failed drive. All drives are involved in the rebuild process, which reduces the rebuild time. One of the two distributed rebuild areas is in use, but the second rebuild area can be used to rebuild the array once more after another failure.

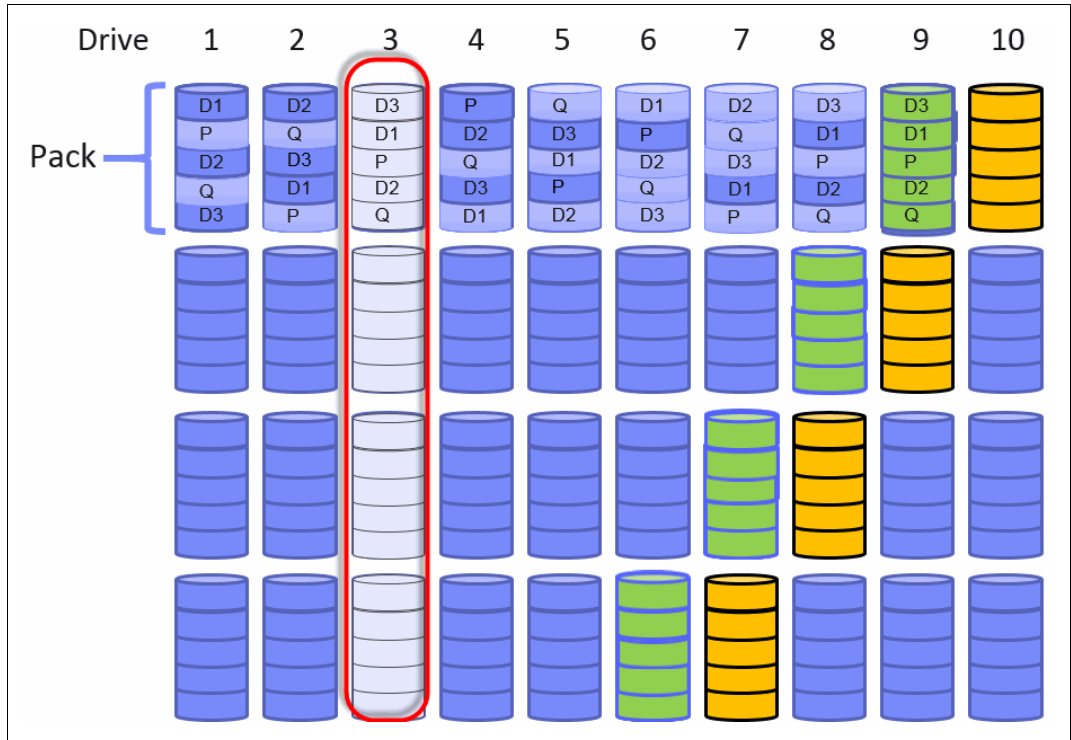


Figure 5-41 Single drive failure with DRAID 6 (for simplification, not all packs are shown)

After the rebuild completes, the array can sustain two more drive failures even before drive 3 is replaced. If no rebuild area is available to perform a rebuild after another drive failure, the array becomes Degraded until a rebuild area is available again and the rebuild can start.

After drive 3 is replaced, a copyback process copies the data from the occupied rebuild area to the replacement drive to empty the rebuild area and make sure that it can be used again for a new rebuild.

DRAID addresses the main disadvantages of TRAIT while providing the same redundancy characteristics:

- ▶ In a drive failure, data is read from many drives and written to many drives. This process minimizes the impact on performance during the rebuild process. Also, it reduces rebuild time. Depending on the DRAID configuration and drive sizes, the rebuild process can be up to 10 times faster.
- ▶ Spare space is distributed throughout the array, which means more drives are processing I/O and no dedicated spare drives are idling.

The DRAID implementation has the following extra advantages:

- ▶ Arrays can be much larger than before and can span many more drives, which improves the performance of the array. The maximum number of drives a DRAID can contain is 128.
- ▶ Existing DRAIDs can be expanded by adding one or more drives. Traditional arrays cannot be expanded.

- ▶ DRAIDs use all the node CPU cores to improve performance, especially in configurations with few arrays.

Here is the minimum number of drives that are needed to build a DRAID array:

- ▶ Two drives for a DRAID 1 array
- ▶ Six drives for a DRAID 6 array
- ▶ Four drives for a DRAID 5 array

Understanding DRAID 1

DRAID 1 can contain only 2 - 6 drives initially and can be expanded up to 16 drives of the same capacity. DRAID 1 arrays consist of two mirrored strips that are distributed across all member drives. Unlike DRAID 5 and 6, DRAID 1 does not contain any parity strips.

DRAID 1 arrays can support:

- ▶ Two drives with no rebuild area. The minimum extent size is 1024 MB.
- ▶ Three to sixteen drives with a single rebuild area. DRAID 1 arrays can tolerate a single failed member drive when a rebuild area is in place.

DRAID 1 array with two drives

If a member drive fails in a DRAID 1 array that contains only two member drives, the array becomes degraded. Degraded storage arrays with only two member drives use the rebuild-in-place process. The *rebuild-in-place process* restores data redundancy by copying or reconstructing the data directly back into the replaced member drive by using the original data distribution. Solid-state drives (SSDs), SAS flash drives, NVMe flash drives, NVMe FCM drives, and SCMs with a maximum capacity of 8 TB support DRAID 1 with two members.

Figure 5-42 shows an example of a DRAID that is configured as a DRAID 1 with two member drives and no rebuild area. Both of the drives in the array are active.

1. (Minimum) Two active drives and stripe width
2. Pack, with a depth of two strips

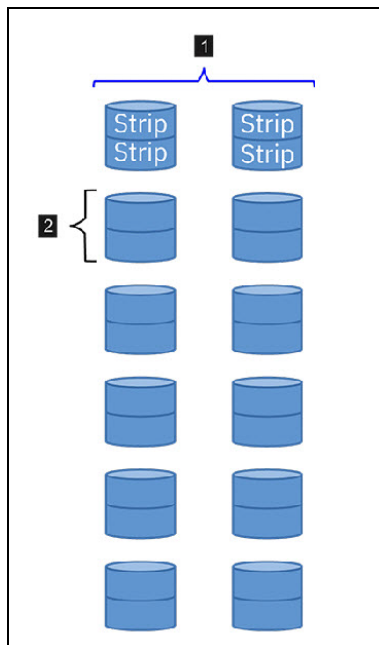


Figure 5-42 DRAID 1 with no rebuild area

Figure 5-43 shows a DRAID with two member drives that contains a failed drive. To recover data, the data copies the strip from the active drive to the new or previously failed drive.

1. Active drive
2. Failed drive

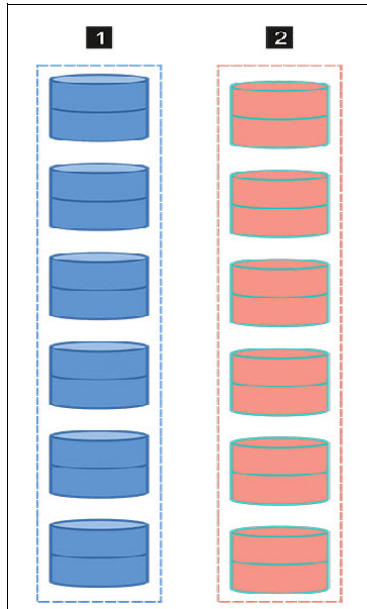


Figure 5-43 DRAID with a failed drive

Figure 5-44 on page 271 shows the rebuild-in-place process, with the new data being copied directly into the replaced drive.

1. Active drive
2. Data being copied directly into the replaced drive

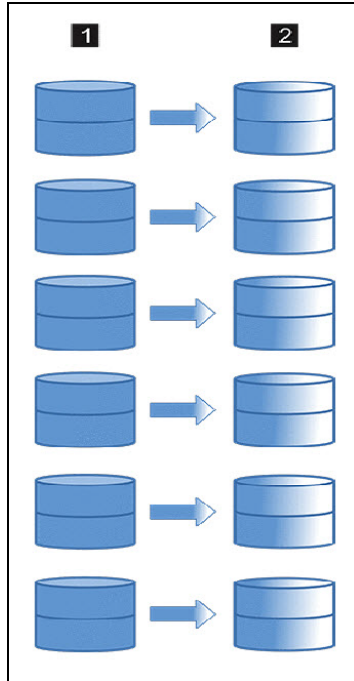


Figure 5-44 Rebuild-in-place process on DRAID 1

DRAID 1 array with three or more drives and a single rebuild area

The rebuild process starts if a member drive fails in a DRAID 1 array with three or more drives. To recover data, the data is read from multiple drives. The recovered data is written to the rebuild areas, which are distributed across all of the drives in the array. All drives are involved in the rebuild process, which reduces the rebuild time. After the drive is replaced, the copyback process starts. Data is copied from the rebuild area to the original location. SSDs, SAS flash drives, NVMe flash drives, NVMe FCM drives, SCMs with maximum capacity of 8 TB, and hard disk drives (HDDs) (up to 8 TB) support DRAID 1 with 3 - 16 members. An array of FCM XL drives (80 TB) is limited to nine drives and cannot be created through the GUI.

Figure 5-45 shows an example of a DRAID that is configured with DRAID 1 with five member drives and a single rebuild area that is marked in yellow.

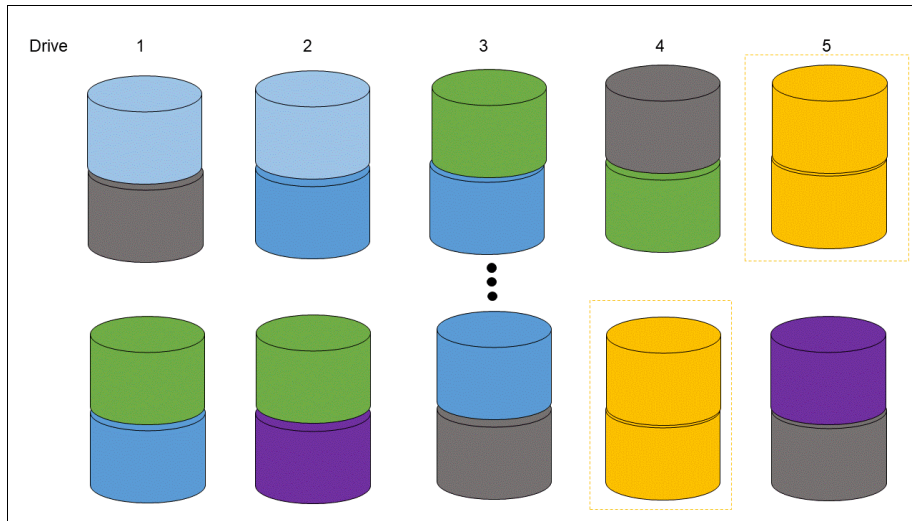


Figure 5-45 Distributed RAID 1 array with five members and a single rebuild area

Figure 5-46 shows that drive 2 failed, which triggers the rebuild process from all drives to the rebuild areas.

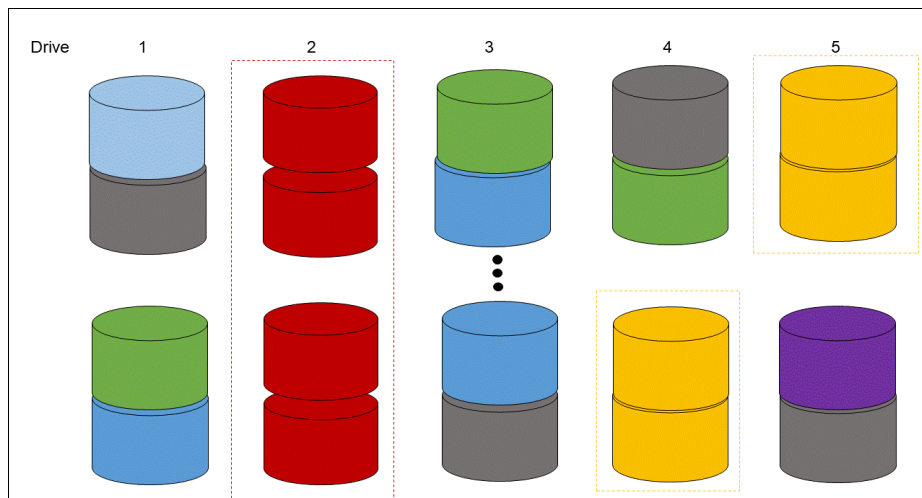


Figure 5-46 Single drive failure in distributed RAID 1 triggers a rebuild area

Figure 5-47 on page 273 shows the copyback process after a drive replacement.

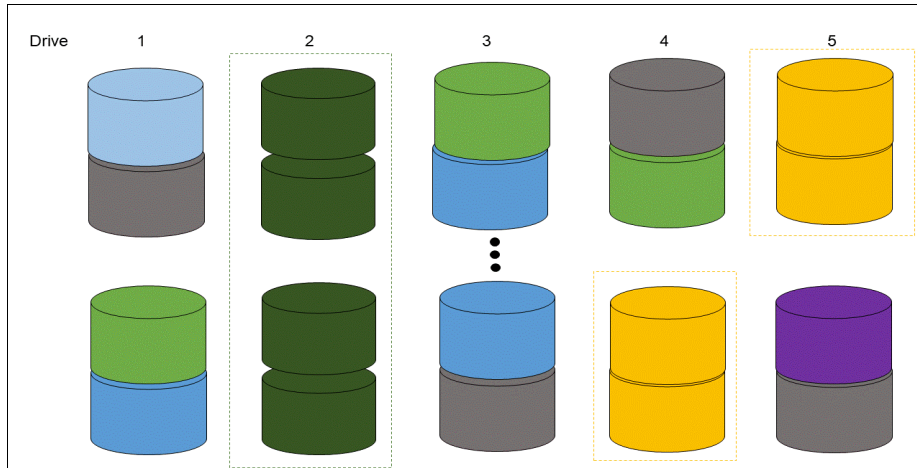


Figure 5-47 Copyback process from a rebuild area to the original location.

5.2.3 Creating arrays

Only RAID arrays (array mode MDisks) can be added to a storage pool. It is not possible to add just a bunch of disks (JBOD) or a single drive. It is also not possible to create a RAID array without assigning it to a storage pool.

Note: As a best practice, use DRAID 6 whenever possible. DRAID technology dramatically reduces rebuild times, decreases the exposure of volumes to the extra load of recovering redundancy, and improves performance. For six drives or less, DRAID 1 is the recommended DRAID type for all supporting platforms.

To create a RAID array from internal storage, select **Pools** → **Pools**, then **Actions**, and then **Add Storage**, or right-click the storage pool to which you want to add arrays and select **Add Storage**, as shown in Figure 5-48.

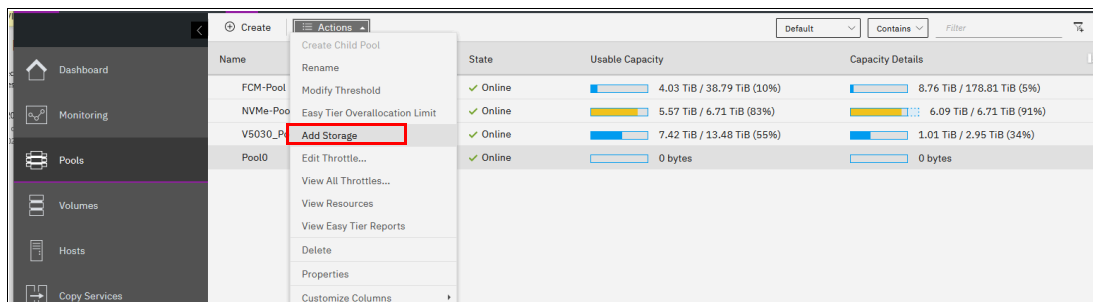


Figure 5-48 Adding storage to a pool

Alternatively, select **Pool** → **Mdisk by pools** and click **Assign** for the drive class that you want to open the configuration dialog, as shown in Figure 5-49.

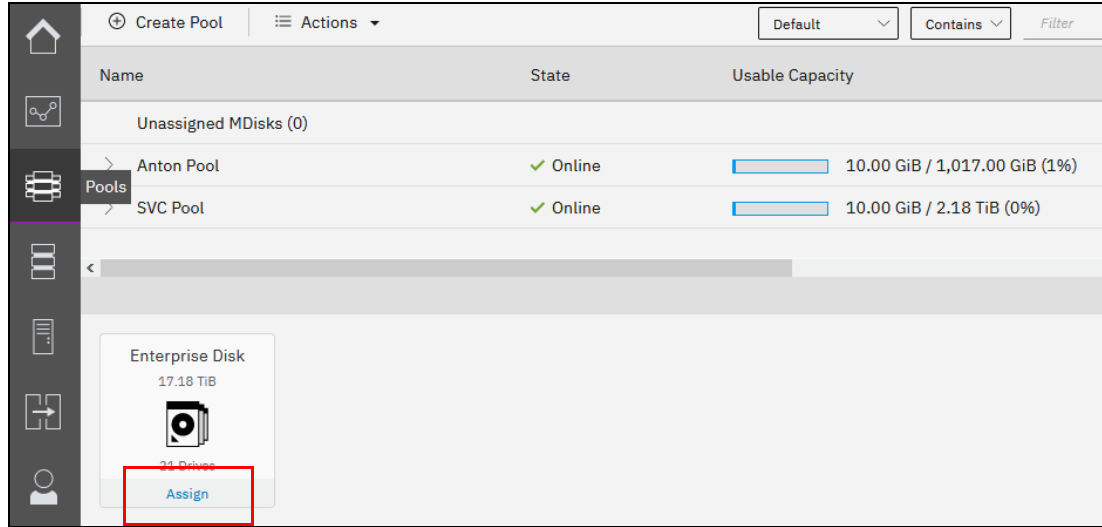


Figure 5-49 Assigning storage for a drive class

This action opens the configuration box that is shown in Figure 5-50. If any of the drives have the Unused role, reconfigure them as Candidates to be included into configuration.

If **Internal-Storage** is chosen, then the system guides you through array Mdisk creation. If **External-Storage** is selected, the system guides you through the selection of external storage. Select the pool from the drop-down menu if no pool is selected yet. The summary view at the right pane shows the Current Usable Capacity of the selected pool. After you define either internal or external storage or both, click **Add storage**.

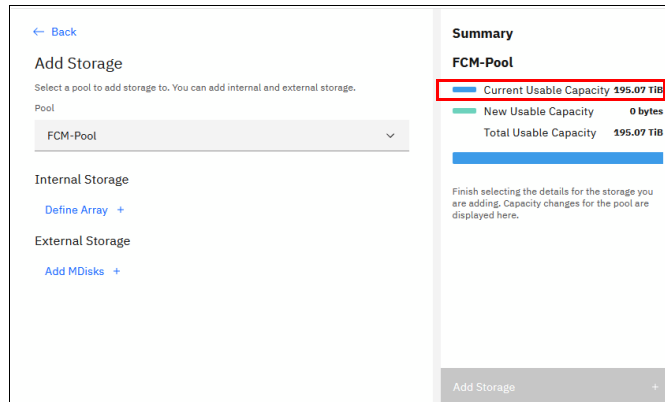


Figure 5-50 Assigning storage to a pool

Internal storage

Select **Define Array**, and choose the drive class for the array from the drop-down menu. Only drive classes for which candidate drives exist are displayed. The system automatically recommends a RAID type and level based on the available candidate drives.

If you are adding storage to a pool that already has storage that is assigned to the pool, the existing storage configuration is considered for the recommendation. The system aims to achieve a balanced configuration, so some properties are inherited from existing arrays in the pool for a specific drive class.

It is not possible to add RAID arrays that are different from existing arrays in a pool by using the GUI. Select **Advanced** to adjust the number of spares or rebuild areas, the stripe width, and the array width before the array is created. Depending on the adjustments that are made, the system might select a different RAID type and level. The summary view at the right pane can be expanded to preview the details of the arrays that are going to be created.

Note: It is not possible to change the RAID level or stripe width of an existing array. You also cannot change the drive count of a traditional array. If you must change these properties, you must delete the array MDisk and re-create it with the required settings.

In Figure 5-51, the dialog box recommends that you create one DRAID 6 with all fifteen 10 K enterprise drives. The summary view reflects the new usable capacity based on your selection.

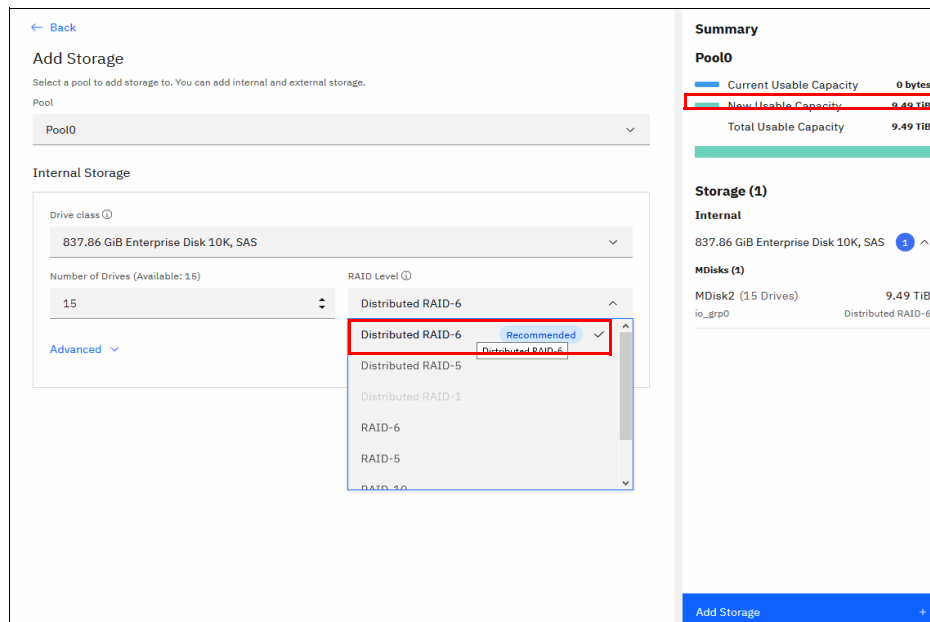


Figure 5-51 DRAID recommendation

Figure 5-52 shows rebuild areas, stripe width, and array width settings.

The screenshot shows a web interface for configuring storage. At the top, there is a 'Back' link and a title 'Add Storage'. Below the title, a message states: 'Select a pool to add storage to. You can add internal and external storage.' The 'Pool' dropdown menu is set to 'Pool0'. Under the 'Internal Storage' section, the 'Drive class' is set to '837.86 GiB Enterprise Disk 10K, SAS'. The 'Number of Drives (Available: 23)' is set to '15', and the 'RAID Level' is set to 'Distributed RAID-6'. An 'Advanced' section is expanded, showing three settings: 'Rebuild Areas' set to '1', 'Stripe width' set to '12', and 'Array width' set to '15'. These three settings are enclosed in a red rectangular box.

Figure 5-52 Advanced selection for rebuild areas, stripe width, and array width

The stripe width indicates the number of strips of data that can be written at one time when data is rebuilt after a drive fails. This value is also referred to as the *redundancy unit width*.

A stripe, which can also be referred to as a redundancy unit, is the smallest amount of data that can be addressed. The DRAID strip size is 256 KB. By default, the system recommends DRAID 6 when possible.

In Figure 5-53 on page 277, if the system has multiple drive classes (for example, flash and enterprise drives), use the plus symbol to create an extra array from other drive classes to take advantage from Easy Tier. The plus symbol is displayed only if multiple drive classes are on the system. For more information about Easy Tier, see Chapter 9, “Advanced features for storage efficiency” on page 509.

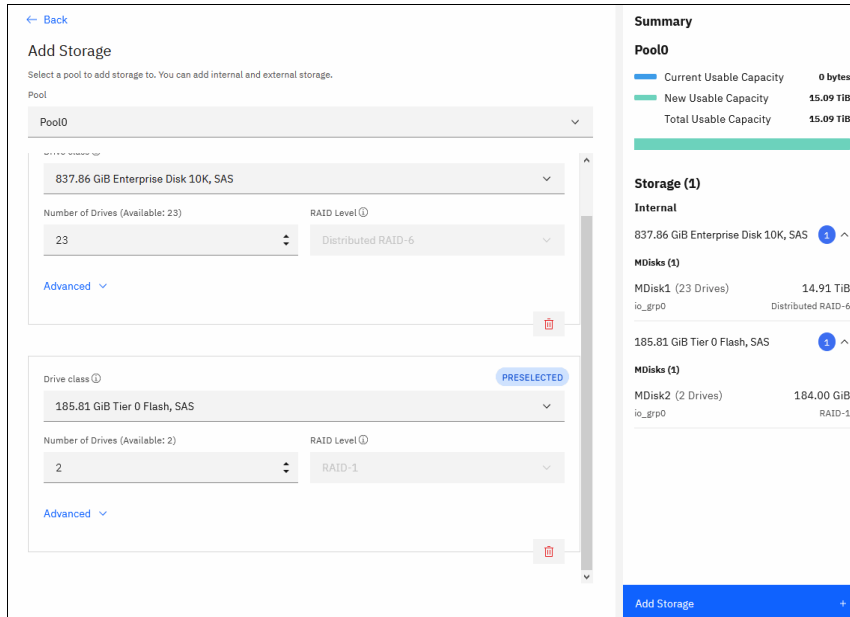


Figure 5-53 Creating arrays from different drive classes.

If the pool has an existing DRAID 6 array of 16 drives, you cannot add a two-drive RAID 1 array to the same pool from the same drive class because this configuration creates an imbalanced storage pool. You can still add any array of any configuration to an existing pool by using the CLI if the platform supports the RAID level.

When you are satisfied with the configuration, click **Add Storage**. The RAID arrays are created, added as array mode MDisks to the pool, and initialized in the background.

If you used self-compressing drives to create the array, the system might prompt you to modify the overallocation limit of the pool. For more information, see “Easy Tier Overallocation Limit window” on page 246.

You can monitor the progress of the initialization by selecting the corresponding task under **Running Tasks** in the upper right of the GUI, as shown in Figure 5-54. The array is available for I/O during this process, so you do not need to wait for it to complete.

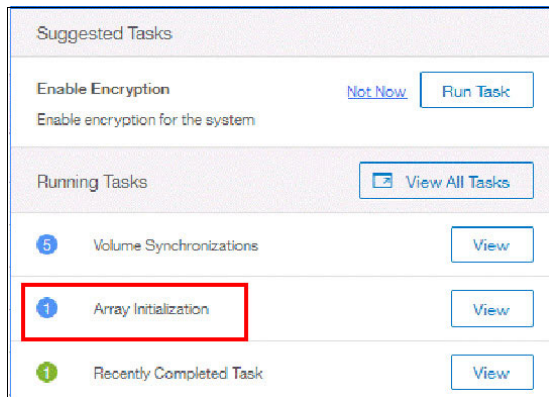


Figure 5-54 Array Initialization task

Click **View** in the **Running Tasks** list to see the initialization progress and the time remaining, as shown in Figure 5-55. The time that it takes to initialize an array depends on the type of drives that is in it. For example, an array of flash drives is much quicker to initialize than NL-serial-attached SCSI (SAS) drives.

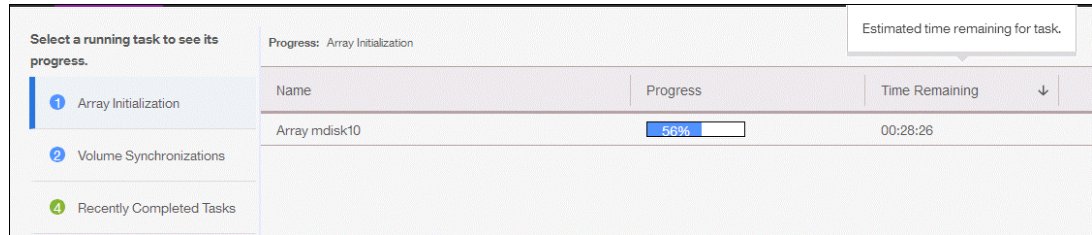


Figure 5-55 Array Initialization task progress information

Configuring arrays with the CLI

When you work with the CLI, run the **mkarray** command to create TRAIT arrays, and run the **mkdistributedarray** command to create DRAID arrays. First, retrieve a list of drives that are ready to become array members. To find information about how to list all available drives, read their use roles, and change those use roles, see 5.2.1, “Working with drives” on page 257.

To get the recommended array configuration by using the CLI, run the **lsdriveclass** command to list the available drive classes, and then use **lsarrayrecommendation** commands, as shown in Example 5-16. The recommendations are listed in the order of preference.

Example 5-16 Listing array recommendations by using the CLI

```

IBM_IBM FlashSystem 7200:superuser>lsdriveclass
id RPM    capacity tech_type      block_size candidate_count
0  10000  1.1TB   tier_enterprise 512          10
1  7200   931.0GB tier_nearline  512          5
2  15000  136.2GB tier_enterprise 512          1
IBM_Storwize:ITS0V7K:superuser>lsarrayrecommendation -driveclass 0 -drivecount 10
Pool2

raid_level distributed stripe_width rebuild_areas drive_count array_count capacity
RAID 6      yes        9            1            10           1           7.6TB
RAID 6      no         10           0            10           1           8.7TB
RAID 5      yes        9            1            10           1           8.7TB
RAID 5      no         9            0            9            1           8.7TB
RAID 10     no         8            0            8            1           4.4TB
RAID 1      no         2            0            2            5           5.5TB

```

To create the recommended DRAID 6 array, specify the RAID level, drive class, number of drives, stripe width, number of rebuild areas, and the storage pool. The system automatically chooses drives for the array from the available drives in the class. In Example 5-17, you create a DRAID 6 array out of 10 drives of class 0 by using a stripe width of 9 and a single rebuild area, and you add it to Pool2.

Example 5-17 Creating a DRAID by running the **mkdistributedarray** command

```

IBM_IBM FlashSystem 7200:superuser>mkdistributedarray -level RAID 6 -driveclass 0
-drivecount 10 -stripewidth 9 -rebuildareas 1 Pool2
MDisk, id [0], successfully created

```

There are default values for the stripe width and the number of rebuild areas, which depend on the RAID level and the drive count. In this example, you had to specify the stripe width because for DRAID 6 it is 12 by default. The drive count value must equal or be greater than the sum of the stripe width and the number of rebuild areas.

To create a RAID 10 MDisk instead, you must specify a list of drives that you want to add as members, the RAID level, and the storage pool name or ID to which you want to add this array.

Example 5-18 creates a RAID 10 array and adds it to Pool2. It also designates a spare drive.

Example 5-18 Creating a RAID by running the mkarray command

```
IBM_IBM FlashSystem 7200:superuser>mkarray -level RAID 10 -drive 0:1:2:3:4:5:6:7
Pool2
MDisk, id [0], successfully created
IBM_IBM FlashSystem 7200:superuser>chdrive -use spare 8
```

Note: Do not forget to designate some of the drives as spares when creating traditional arrays. Spare drives are required to perform a rebuild immediately after a drive failure.

The storage pool must exist. To create a storage pool, see 5.1.1, “Creating storage pools” on page 240. To check the array initialization progress by using the CLI, run the `lsarrayinfo progress` command.

5.2.4 Actions on arrays

MDisks that are created from internal storage support specific actions that external MDisks do not support. Some actions that are supported on TRAIID arrays are not supported on DRAID arrays, and vice versa.

To select an action, select **Pools** → **MDisks by Pools**, select the array (MDisk), and click **Actions**. Alternatively, right-click the array, as shown in Figure 5-56.

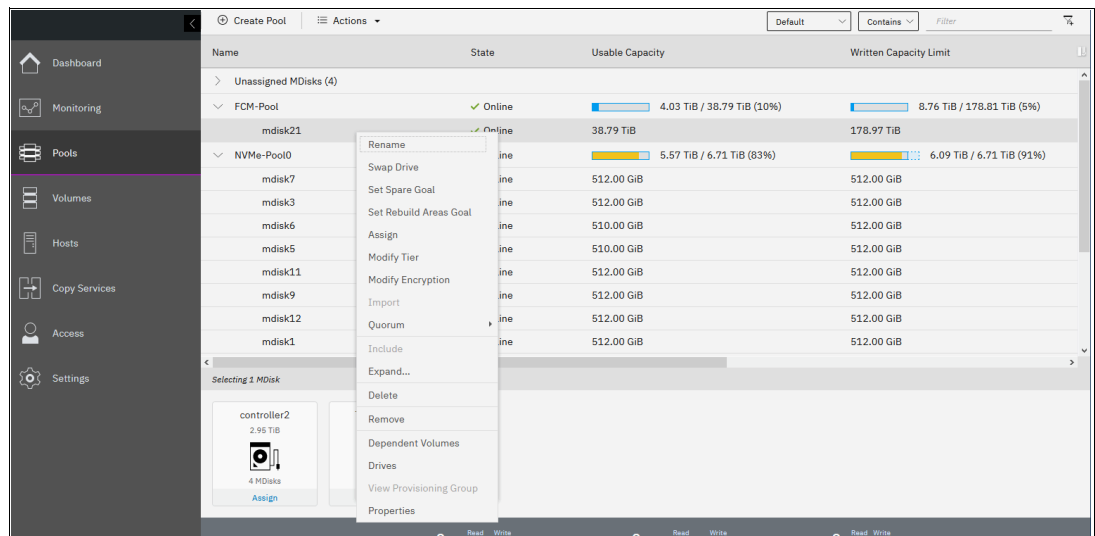


Figure 5-56 Actions on arrays

Rename

To change the name of an MDisk, select this option.

The CLI command for this operation is **charray**, as shown in Example 5-19. No feedback is returned.

Example 5-19 Renaming an array MDisk by running the *charray* command

```
IBM_IBM FlashSystem 7200:superuser>charray -name Distributed_array mdisk21
IBM_IBM FlashSystem 7200:superuser>
```

Swap Drive

To replace a drive in the array with another drive, select **Swap Drive**. The other drive must have the Candidate or Spare role. Use this action to perform proactive drive replacement or replace a drive that has not failed but is expected to fail soon, for example, as indicated by an error message in the event log.

Figure 5-57 shows the dialog box that opens. Select the member drive that you want to replace and the replacement drive, and click **Swap**.

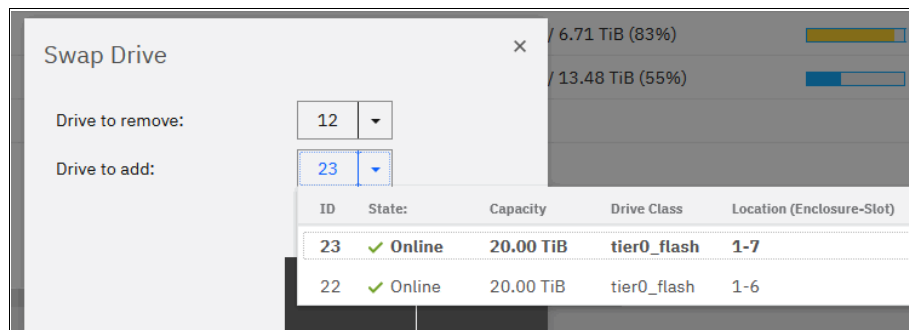


Figure 5-57 Swapping an array member with a candidate or spare drive

The exchange of the drives runs in the background. The volumes on the affected MDisk remain accessible during the process.

Swapping a drive in a traditional array performs a concurrent member exchange, which does not reduce the redundancy of the array. The data of the old member is copied to the new member, and after the process is complete, the old member is removed from the array.

In a DRAID, the system immediately removes the old member from the array and performs a rebuild. After the rebuild completes, a copyback is initiated to copy the data to the new member drive. This process is non-disruptive, but reduces the redundancy of the array during the rebuild process.

You can run the **charraymember** command to do this task. Example 5-20 shows the replacement of array member ID 7 that was assigned to drive ID 12 with drive ID 17. The **-immediate** parameter is required for DRAIDs to acknowledge that a rebuild will start.

Example 5-20 Replacing an array member by using the CLI (some columns are not shown)

```
IBM_IBM FlashSystem 7200:superuser>lsarraymember 16
mdisk_id mdisk_name      member_id drive_id new_drive_id spare_protection
16       Distributed_array 6        18        17          1
16       Distributed_array 7        12        17          1
16       Distributed_array 8        15        17          1
<...>
```



```

IBM_IBM FlashSystem 7200:superuser>lsdrive
id status error_sequence_number use tech_type capacity mdisk_id
16 online member tier_enterprise 558.4GB 16
17 online spare tier_enterprise 558.4GB
18 online member tier_enterprise 558.4GB 16
<...>
IBM_IBM FlashSystem 7200:superuser>charraymember -immediate -member 7 -newdrive 17
Distributed_array
IBM_IBM FlashSystem 7200:superuser>

```

Set Spare Goal or Set Rebuild Areas Goal

Select this option to set the number of spare drives (on a RAID) or rebuild areas (on a DRAID) that are expected to protect the array from drive failures.

If the number of rebuild areas that are available does not meet the configured goal, an error is logged in the event log, as shown in Figure 5-58. This error can be fixed by replacing failed drives in the DRAID array.

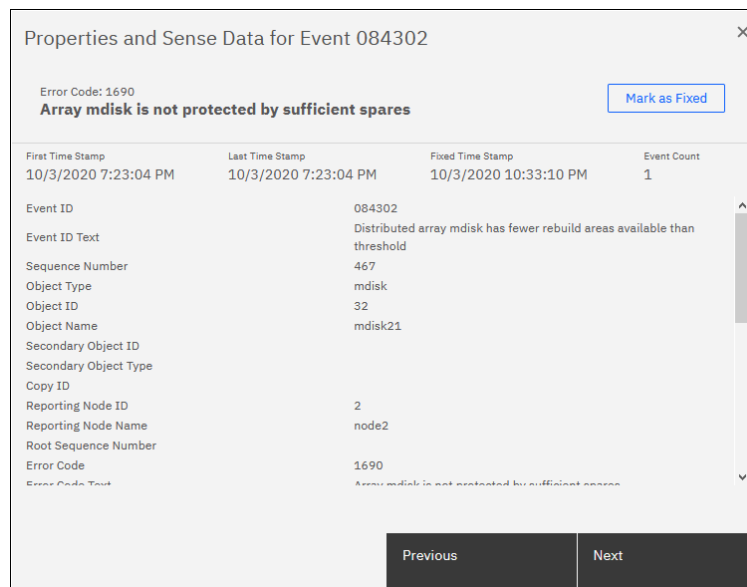


Figure 5-58 Error 1690 for insufficient rebuild areas available

Note: This option does not change the actual number of rebuild areas or spares that are available to the array, but specifies only at which point a warning event is generated. Setting the goal to 0 does not prevent the array from rebuilding.

On the CLI, this task is performed with the **charray** command (see Example 5-21).

Example 5-21 Adjusting array goals by running the charray command (some columns are not shown)

```

IBM_IBM FlashSystem 7200:superuser>lsarray
mdisk_id mdisk_name status mdisk_grp_id mdisk_grp_name distributed
0 mdisk0 online 0 mdiskgrp0 no
16 Distributed_array online 1 mdiskgrp1 yes
IBM_IBM FlashSystem 7200:superuser>charray -sparegoal 2 mdisk0
IBM_IBM FlashSystem 7200:superuser>charray -rebuildareasgoal 2 Distributed_array

```

Expand

Select **Expand** to expand the array by adding more drives to it to increase the available capacity of the array or create more rebuild areas. Only DRAIDs can be expanded because the option is not available for traditional arrays.

Candidate drives of a drive class that is compatible with the drive class of the array must be available in the system or an error message is shown and the array cannot be expanded. A drive class is compatible with another one if its characteristics, such as capacity and class, are an exact match or are superior. In most cases, drives of the same class should be used to expand an array.

The dialog box that is shown in Figure 5-59 shows an overview of the size of the array, the number of available candidate drives in the selected drive class, and the new array capacity after the expansion. The drive class and the number of drives to add can be modified as required and the projected new array capacity is updated. To add more rebuild areas to the array, click **Advanced Settings** and modify the number of extra spares.

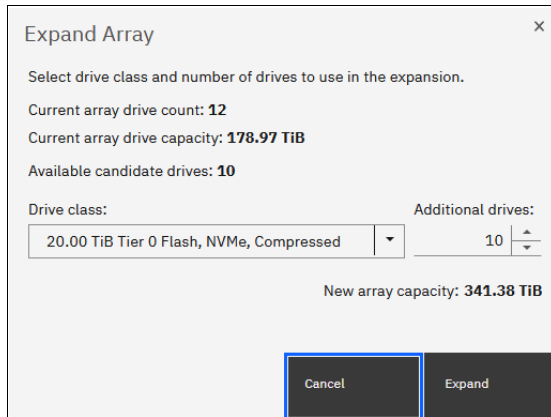


Figure 5-59 Expanding a DRAID

Clicking **Expand** starts a background process that adds the selected number of drives to the array. As part of the expansion, the system automatically migrates data for optimal performance for the new expanded configuration.

You can monitor the progress of the expansion by clicking the **Running Tasks** icon in the upper right of the GUI or by selecting **Monitoring** → **Background tasks** as shown in Figure 5-60.



Figure 5-60 Array expansion progress

Note: When you expand a thin-provisioned NVMe array, the physical capacity is not immediately available, and the availability of new physical capacity is not tracked with logical expansion progress.

On the CLI, this task is performed by running the `expandarray` command. To get a list of compatible drive classes, run the `lscompatibledriveclasses` command, as shown in Example 5-22.

Example 5-22 Expanding an array by using the CLI

```
IBM_IBM FlashSystem 7200:superuser>lsarray 0
<..>
capacity 3.2TB
<..>
drive_class_id 0
drive_count 6
<..>
rebuild_areas_total 1
IBM_IBM FlashSystem 7200:superuser>lscompatibledriveclasses 0
id
0
IBM_IBM FlashSystem 7200:superuser>expandarray -driveclass 0 -totaldrivecount 10
-totalrebuildareas 2 0
IBM_IBM FlashSystem 7200:superuser>lsarrayexpansionprogress
progress estimated_completion_time target_capacity additional_capacity_remaining
29          191018233758             5.17TB          1.38TB
```

Note: The `expandarray` command uses the total drive count *after* the expansion as a parameter, including both the number of new drives and the number of drives in the array before the expansion. The same is true for the number of rebuild areas.

Delete

Select **Delete** to remove the array from the storage pool and delete it. An array MDisk does not exist outside of a storage pool. Therefore, an array cannot be removed from the pool without being deleted. All drives that belong to the deleted array take on the Candidate role.

If there are no volumes that use extents from this array, the command runs immediately without extra confirmation. If there are volumes that use extents from this array, you are prompted to confirm the action, as shown in Figure 5-61.

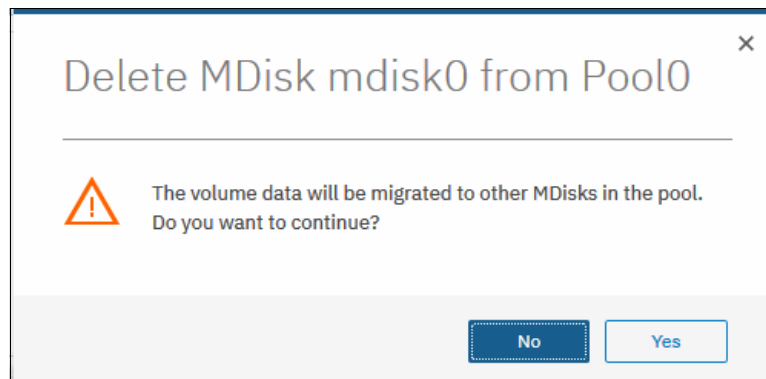


Figure 5-61 Deleting an array from a non-empty storage pool

Confirming the deletion starts a background process that migrates used extents on the MDisk to other MDisks in the same storage pool. After that process completes, the array is removed from the storage pool and deleted.

Note: The command fails if you do not have enough available capacity remaining in the storage pool to allocate the capacity that is being migrated away from the removed array.

To delete the array with the CLI, run the `rmarray` command. The `-force` parameter is required if volume extents must be migrated to other MDisks in a storage pool.

To monitor the progress of the migration, use the **Running Tasks** section in the GUI or the `ismigrate` command on the CLI. The MDisk continues to exist until the migration completes.

Dependent Volumes

A volume depends on an MDisk if the MDisk becoming unavailable results in a loss of access or a loss of data for that volume. Use this option before you do maintenance operations to confirm which volumes (if any) will be affected.

If an MDisk in a storage pool goes offline, the entire storage pool goes offline, which means all volumes in a storage pool depend on each MDisk in the same pool, even if the MDisk does not have extents for each of the volumes. Clicking the **Dependent Volumes Action** menu of an MDisk lists the volumes that depend on that MDisk, as shown in Figure 5-62.

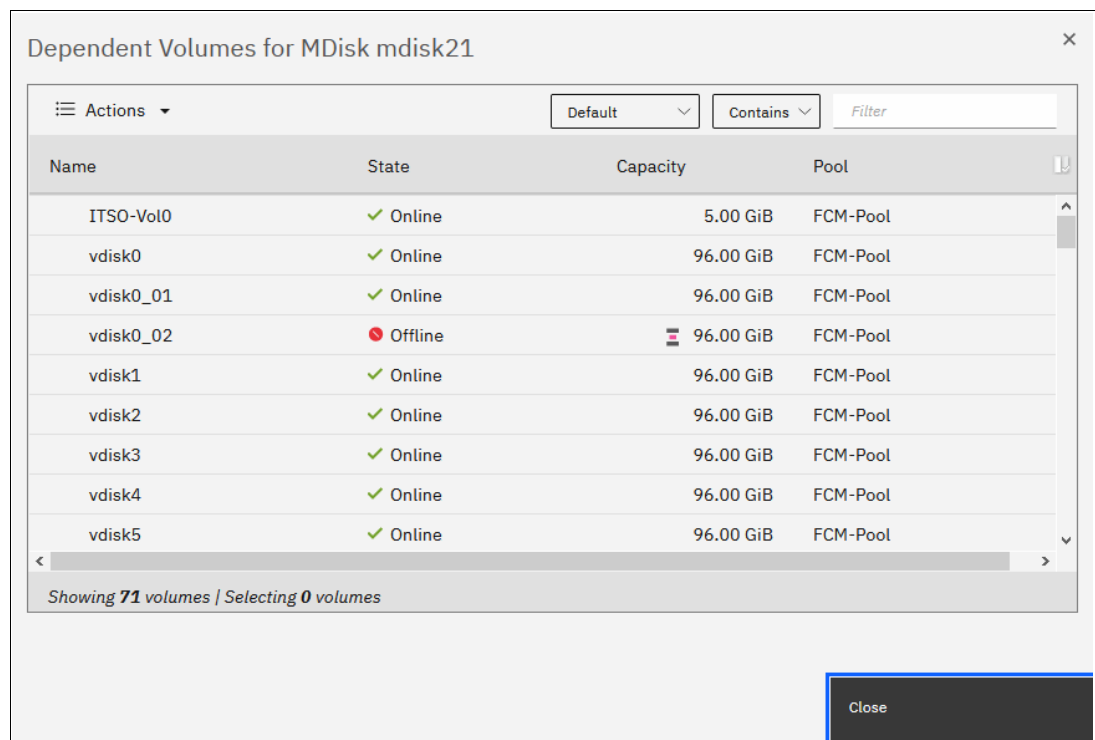


Figure 5-62 Dependent volumes for MDisk mdisk21

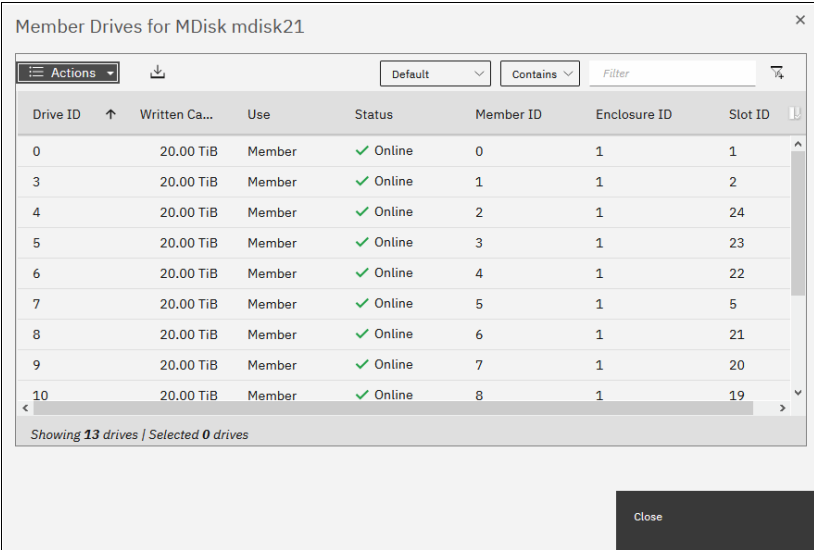
You can get the same information by running the `lsdependentvdisks` command, as shown in Example 5-23.

Example 5-23 Listing virtual disks that depend on a MDisk by using the CLI

```
IBM_IBM FlashSystem 7200:superuser>lsdependentvdisks -mdisk mdisk21
vdisk_id vdisk_name
48 vdisk0
49 vdisk1
50 vdisk2
<...>
```

Drives

To see information about the member drives that are included in the array, select **Drives**, as shown in Figure 5-63.



The screenshot shows a window titled "Member Drives for MDisk mdisk21". It contains a table with the following columns: Drive ID, Written Ca..., Use, Status, Member ID, Enclosure ID, and Slot ID. The table lists 13 drives, all with a status of "Online". A "Close" button is visible at the bottom right of the window.

Drive ID	Written Ca...	Use	Status	Member ID	Enclosure ID	Slot ID
0	20.00 TiB	Member	✓ Online	0	1	1
3	20.00 TiB	Member	✓ Online	1	1	2
4	20.00 TiB	Member	✓ Online	2	1	24
5	20.00 TiB	Member	✓ Online	3	1	23
6	20.00 TiB	Member	✓ Online	4	1	22
7	20.00 TiB	Member	✓ Online	5	1	5
8	20.00 TiB	Member	✓ Online	6	1	21
9	20.00 TiB	Member	✓ Online	7	1	20
10	20.00 TiB	Member	✓ Online	8	1	19

Showing 13 drives / Selected 0 drives

Figure 5-63 List of drives in an array

You can get the same information by running the `lsarraymember` command. Provide an array name or ID as the parameter to filter the output from the array. If you run the command without arguments, the command lists all members of all configured arrays.

Properties

This section shows all the available array MDisk parameters: its state, capacity, RAID level, and others.

To get a list of all configured arrays, run the `lsarray` command with the array name or ID as the parameter to get more information about the array, as shown in Example 5-24.

Example 5-24 The `lsarray` command output (truncated)

```
IBM_IBM FlashSystem 7200:superuser>lsarray
mdisk_id mdisk_name      status mdisk_grp_id mdisk_grp_name capacity
0         mdisk0           online 0             NVMe-Poo10 744.2GB
32        mdisk21         online 2             FCM-Poo1 194.2TB
IBM_IBM FlashSystem 7200:superuser>lsarray 32
mdisk_id 32
mdisk_name mdisk21
status online
mode array
mdisk_grp_id 2
mdisk_grp_name FCM-Poo1
capacity 194.3TB
<...>
```

5.3 Working with external controllers and MDisks

Controllers are external storage systems that provide storage resources that are used as MDisks. The system supports external storage controllers that are attached through internet Small Computer Systems Interface (iSCSI) and through Fibre Channel (FC).

A key feature of the system is its ability to consolidate disk controllers from various vendors into storage pools. The storage administrator can manage and provision storage to applications from a single user interface and use a common set of advanced functions across all of the storage systems under the control of the system.

This concept is called *External Virtualization*, which makes your storage environment more flexible, cost-effective, and easy to manage. External Virtualization is a licensed function.

For more information about how to configure external storage systems, see 2.9, “Back-end storage configuration” on page 88.

5.3.1 External storage controllers

External storage controllers can be attached by using FC and iSCSI. The following sections describe how to attach external storage controllers to the system and how to manage them by using the GUI.

System layers

A *system layer* affects how the system interacts with other external IBM Storwize or IBM FlashSystem family systems. A system is in either the *storage* layer (default) or the *replication* layer.

In the storage layer, the system can provide external storage for a replication-layer system, but it cannot use another Storwize or IBM FlashSystem family system that is configured with the storage-layer external storage.

In the replication layer, the system cannot provide external storage for a replication-layer system, but the system can use another Storwize or IBM FlashSystem family system that is configured with storage-layer external storage.

You get a warning that your system is in the *storage layer* if you try to add an external iSCSI storage controller by using the GUI. You are prompted to convert the system to the *replication layer* automatically.

Note: Before you change the system layer, the following conditions must be met:

- ▶ No host object can be configured with worldwide port names (WWPNs) from a Storwize or IBM FlashSystem family system.
- ▶ No system partnerships can be defined.
- ▶ No Storwize or IBM FlashSystem family system can be visible on the storage area network (SAN) fabric.

To switch the system layer, you can also run the **chsystem** CLI command, as shown in Example 5-25 on page 287. If the command runs successfully, it returns no output.

Example 5-25 Changing the system layer

```
IBM_IBM FlashSystem 7200:superuser>lssystem | grep layer
layer storage
IBM_IBM FlashSystem 7200:superuser>chsystem -layer replication
IBM_IBM FlashSystem 7200:superuser>
```

For more information about layers and how to change them, go to [IBM FlashSystem 9200 documentation](#) and select **Product overview** → **Technical overview** → **System layers**.

Attachment by using Fibre Channel

A controller that is connected through FC is detected automatically by the system if the cabling, zoning, and system layer are configured correctly. For more information about how to attach and zone back-end storage controllers to the system, see 2.6, “Fibre Channel SAN configuration planning” on page 76.

If the external controller is not detected, ensure that the system is cabled and zoned into the same SAN as the external storage system. Check that layers are set correctly on both virtualizing and virtualized systems if they belong to the IBM Storwize or IBM FlashSystem family.

After the problem is corrected, rescan the FC network immediately by selecting **Pools** → **External Storage**, and then selecting **Actions** → **Discover Storage**, as shown in Figure 5-64.

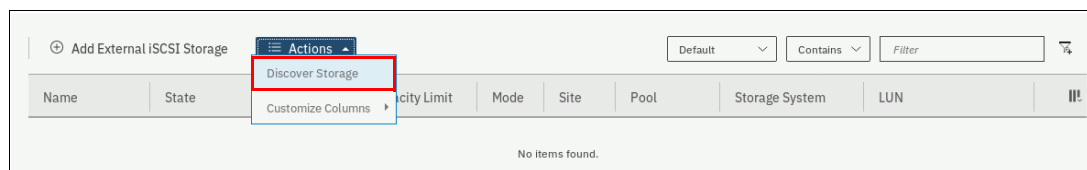


Figure 5-64 Discover Storage menu

This action runs the `detectmdisk` command. It returns no output. Although it might appear that the command completed, some extra time might be required for it to run. The command is asynchronous and returns a prompt while the command continues to run in the background.

Attachment by using iSCSI

You must manually configure iSCSI connections between the system and the external storage controller. Until you do this task, the controller is not listed in the External Storage window. For more information about how to attach back-end storage controllers to the system, see Chapter 2, “Planning” on page 71.

To start virtualizing an iSCSI back-end controller, you must follow the steps in [IBM FlashSystem 9200 documentation](#) to perform configuration steps that are specific to your back-end storage controller. You can see find the steps by selecting **Configuring** → **Configuring and servicing storage systems** → **External storage system configuration with iSCSI connections**.

For more information about configuring the system to virtualize a back-end storage controller with iSCSI, see *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327.

Managing external storage controllers

You can manage both FC and iSCSI storage controllers through the External Storage window. To access the External Storage window, select **Pools** → **External Storage**, as shown in Figure 5-65.

Name	State	Written Capacity Limit		Mode	Site
controller1	Online	IBM 2145	Serial Number: 2076	Site: Unassigned	WWNN: 5005076800018F8C
controller3	Online	IBM 2145	Serial Number: 2076	Site: Unassigned	WWNN: 50050768000054F0
controller0	Online	IBM 2145	Serial Number: 2076	Site: Unassigned	WWNN: 5005076800018F8D
mdisk10	Online			512.00 GiB	Managed

Figure 5-65 External Storage window

Note: A controller that is connected through FC is detected automatically by the system. The cabling, the zoning, and the system layer must be configured correctly. A controller that is connected through iSCSI must be added to the system manually.

Depending on the type of back-end system, it might be detected as one or more controller objects.

If the External Storage window does not appear in the Pools windows, the virtualization licenses are not configured. To use the system's virtualization functions, you must order the correct External Virtualization licenses. You can configure the licenses by selecting **Settings** → **System** → **Licensed Functions**. For assistance with licensing questions or to purchase any of these licenses, contact your IBM account team or IBM Business Partner.

The External Storage window lists the external controllers that are connected to the system and all the external MDisks that are detected by the system. The MDisks are organized by the external storage system that presents them. Toggle the sign to the left of the controller icon to show or hide the MDisks that are associated with the controller.

If you configured logical unit names on your external storage systems, it is not possible for the system to determine these names because they are local to the external storage system. However, you can use the LU unique identifiers (UIDs), or external storage system worldwide node names (WWNNs) and LU number to identify each device.

To list all visible external storage controllers with CLI, run the `lscontroller` command, as shown in Example 5-26.

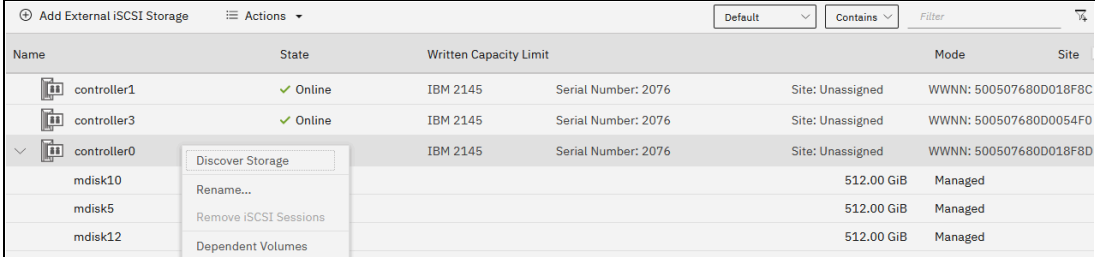
Example 5-26 Listing controllers by using the CLI (some columns are not shown)

```
IBM_IBM FlashSystem 7200:superuser>lscontroller
id controller_name ctrl_s/n          vendor_id          product_id_low
0 controller0      2076              IBM                2145
1 controller1      2076              IBM                2145
2 controller2      2076              IBM                2145
<...>
```


5.3.2 Actions for external storage controllers

You can perform many actions on external storage controllers. Some actions are available for external iSCSI controllers only.

To select any action, select **Pools** → **External Storage** and right-click the controller, as shown in Figure 5-66. Alternatively, select the controller and click **Actions**.



Name	State	Written Capacity	Limit	Mode	Site
controller1	Online	IBM 2145	Serial Number: 2076	Site: Unassigned	WWNN: 500507680D018F8C
controller3	Online	IBM 2145	Serial Number: 2076	Site: Unassigned	WWNN: 500507680D0054F0
controller0		IBM 2145	Serial Number: 2076	Site: Unassigned	WWNN: 500507680D018F8D
mdisk10				512.00 GiB	Managed
mdisk5				512.00 GiB	Managed
mdisk12				512.00 GiB	Managed

Figure 5-66 Actions for external storage controllers

Discover Storage

When you create or remove LUs on an external storage system, the change might not be detected immediately. In this case, click **Discover Storage** so that the system can rescan the FC or iSCSI network. In general, the system automatically detects disks when they appear on the network. However, some FC controllers do not send the required SCSI primitives that are necessary to automatically discover the new disks.

The rescan process discovers any new MDisks that were added to the system and rebalances MDisk access across the available ports. It also detects any loss of availability of the controller ports.

This action runs the `detectmdisk` command.

Rename

To modify the name of an external controller to simplify administration tasks, click **Rename**. The naming rules are the same as for storage pools, and they can be found in 5.1.1, “Creating storage pools” on page 240.

To rename a storage controller by using the CLI, run the `chcontroller` command.

Removing iSCSI sessions

This action is available only for external controllers that are attached with iSCSI. To remove the iSCSI session that is established between the source and target port, right-click the session and select **Remove**.

For more information about the CLI commands and detailed instructions, see *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327.

Modifying a site

This action is available only for systems that use IBM HyperSwap or are in a topology. To change the site with which the external controller is associated, select **Modify Site**, as shown in Figure 5-67.

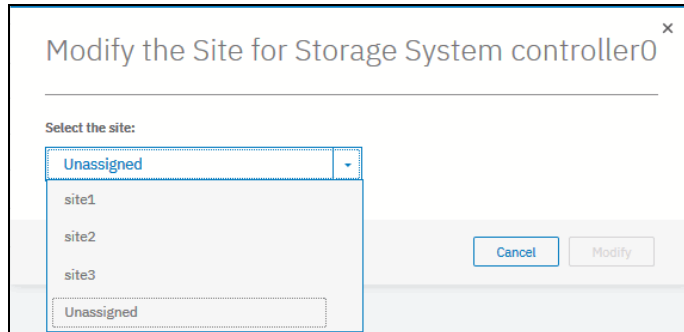


Figure 5-67 Modifying the site of an external controller

To change the controller site assignment by using the CLI, run the **chcontroller** command. Example 5-27 shows that controller0 was renamed and reassigned to a different site.

Example 5-27 Changing a controller's name and site

```
IBM_IBM FlashSystem 7200:superuser>chcontroller -name site3_controller -site site3
controller0
IBM_IBM FlashSystem 7200:superuser>
```

5.3.3 Working with external MDisks

After an external back-end storage controller is configured, attached to the system, and detected as a controller, you can work with LUs that are provisioned from it. Each LU is represented by an MDisk object.

External MDisks can have one of the following modes:

► *Unmanaged*

External MDisks are initially discovered by the system as unmanaged MDisks. An unmanaged MDisk is not a member of any storage pool. It is not associated with any volumes, and has no metadata that is stored on it. The system does not write to an MDisk that is in unmanaged mode except when it attempts to change the mode of the MDisk to one of the other modes. Removing an external MDisk from a pool returns it to unmanaged mode.

► *Managed*

When unmanaged MDisks are added to storage pools, they become managed. Managed mode MDisks are always members of a storage pool, and their extents contribute to the storage pool. This mode is the most common and normal mode for an MDisk.

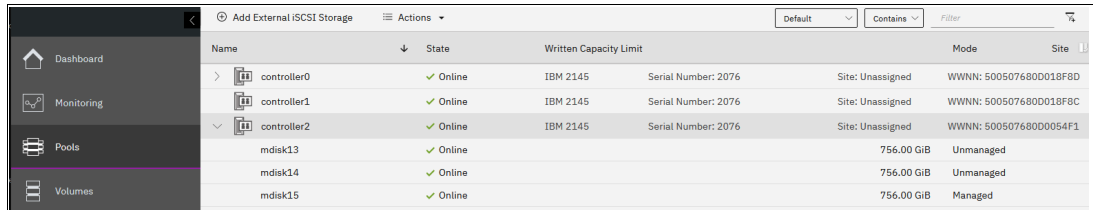
► *Image*

Image mode provides a direct block-for-block conversion from the MDisk to a volume. This mode is provided to satisfy the following major usage scenarios:

- Presenting existing data on an MDisk through the system to an attached host
- Importing existing data on an MDisk into the system
- Exporting data on a volume by performing a migration to an image mode MDisk

Listing external MDisks

You can manage external MDisks by using the External Storage window, which is accessed by selecting **Pools** → **External Storage**, as shown in Figure 5-68.



Name	State	Written Capacity Limit	Mode	Site
controller0	Online	IBM 2145 Serial Number: 2076	Site: Unassigned	WWNN: 500507680D018F8D
controller1	Online	IBM 2145 Serial Number: 2076	Site: Unassigned	WWNN: 500507680D018F8C
controller2	Online	IBM 2145 Serial Number: 2076	Site: Unassigned	WWNN: 500507680D0054F1
mdisk13	Online	756.00 GiB	Unmanaged	
mdisk14	Online	756.00 GiB	Unmanaged	
mdisk15	Online	756.00 GiB	Managed	

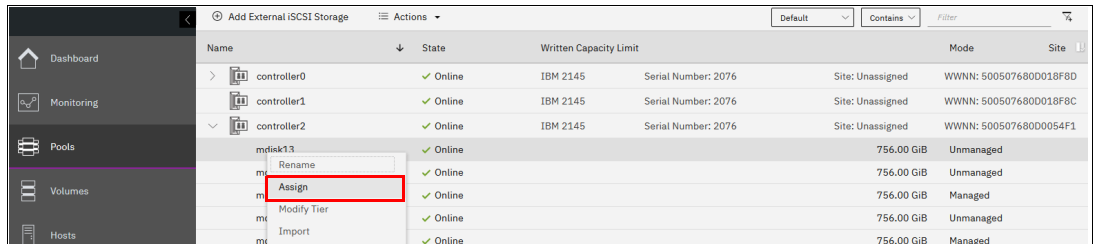
Figure 5-68 External Storage window

To list all MDisks that are visible by the system by using the CLI, run the `lsmdisk` command without any parameters. If required, you can filter output to include only external or only array type MDisks.

Assigning MDisks to pools

You can add *Unmanaged* MDisks to an existing pool or create a pool to include them. If no storage pool exists yet, follow the procedure that is outlined in 5.1.1, “Creating storage pools” on page 240.

Figure 5-69 shows how to add selected MDisk to an existing storage pool. Click **Assign** under the **Actions** menu or right-click the MDisk and select **Assign**.



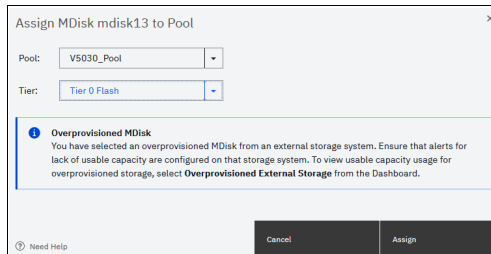
Name	State	Written Capacity Limit	Mode	Site
controller0	Online	IBM 2145 Serial Number: 2076	Site: Unassigned	WWNN: 500507680D018F8D
controller1	Online	IBM 2145 Serial Number: 2076	Site: Unassigned	WWNN: 500507680D018F8C
controller2	Online	IBM 2145 Serial Number: 2076	Site: Unassigned	WWNN: 500507680D0054F1
mdisk13	Online	756.00 GiB	Unmanaged	
mdisk14	Online	756.00 GiB	Unmanaged	
mdisk15	Online	756.00 GiB	Managed	

Context menu for mdisk13:

- Rename
- Assign
- Modify Tier
- Import

Figure 5-69 Assigning an unmanaged MDisk

After you click **Assign**, a dialog box opens, as shown in Figure 5-70. Select the target pool, MDisk storage tier, and external encryption setting.



Assign MDisk mdisk13 to Pool

Pool: V5030_Pool

Tier: Tier 0 Flash

Overprovisioned MDisk
You have selected an overprovisioned MDisk from an external storage system. Ensure that alerts for lack of usable capacity are configured on that storage system. To view usable capacity usage for overprovisioned storage, select **Overprovisioned External Storage** from the Dashboard.

Cancel Assign

Figure 5-70 Assign MDisk dialog box

When you add MDisks to pools, you must assign them to the correct storage tiers. It is important to set the tiers correctly if you plan to use the Easy Tier feature. Using an incorrect tier can mean that the Easy Tier algorithm might make wrong decisions and thus affect system performance. For more information about storage tiers, see Chapter 9, “Advanced features for storage efficiency” on page 509.

The storage tier setting can also be changed after the MDisk is assigned to the pool.

Select the **Externally encrypted** checkbox if your back-end storage performs data encryption. For more information about encryption, see Chapter 12, “Encryption” on page 735.

After the task completes, click **Close**.

Note: If the external storage LUs that are presented to the system contain data that must be retained, do not use the **Assign** option to add the MDisks to a pool. This option destroys the data on the LU. Instead, use the **Import** option to create an image mode MDisk. For more information, see Chapter 8, “Storage migration” on page 485.

To see the external MDisks that are assigned to a pool within the system, select **Pools** → **MDisks by Pools**.

When a new MDisk is added to a pool that already contains MDisks and volumes, the Easy Tier feature automatically balances volume extents between the MDisks in the pool as a background process. The goal of this process is to distribute extents in a way that provides the best performance to the volumes. It does *not* attempt to balance the amount of data evenly between all MDisks.

The data migration decisions that Easy Tier makes between tiers of storage (inter-tier) or within a single tier (intra-tier) are based on the I/O activity that is measured. Therefore, when you add an MDisk to a pool, extent migrations are not necessarily performed immediately. No migration of extents occurs until there is sufficient I/O activity to trigger it.

If Easy Tier is turned off, no extent migration is performed. Only newly allocated extents are written to a new MDisk.

For more information about the Easy Tier feature, see Chapter 9, “Advanced features for storage efficiency” on page 509.

To assign an external MDisk to a storage pool by using the CLI, run the `addmdisk` command. You must specify the MDisk name or ID, MDisk tier, and target storage pool, as shown in Example 5-28. The command returns no feedback.

Example 5-28 The addmdisk command

```
IBM_IBM FlashSystem 7200:superuser>addmdisk -mdisk mdisk3 -tier enterprise Pool0
IBM_IBM FlashSystem
7200:superuser>
```

5.3.4 Actions for external MDisks

External MDisks support specific actions that are not supported on RAID arrays that are made from internal storage. Some actions are supported only on unmanaged external MDisks, and some are supported only on managed external MDisks.

To choose an action, select **Pools** → **External Storage** or **Pools** → **MDisks by Pools**, select the **external MDisk**, and click **Actions**, as shown in Figure 5-71 on page 293. Alternatively, right-click the **external MDisk**.

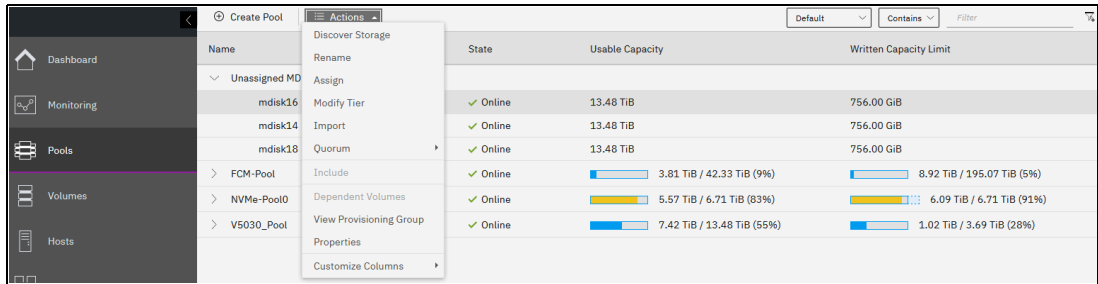


Figure 5-71 Actions for MDisks

Discover Storage

This option is available even if no MDisks are selected. By running it, you cause the system to rescan the iSCSI and FC network for these purposes:

- ▶ Find any new MDisks that might have been added.
- ▶ Rebalance MDisk access across all available controller device ports.

This action runs the `detectmdisk` command.

Assign

This action is available only for unmanaged MDisks. Select **Assign** to open the dialog box that is explained in “Assigning MDisks to pools” on page 291.

Modify Tier

To modify the tier to which the external MDisk is assigned, select **Modify Tier**, as shown in Figure 5-72. This setting is adjustable because the system cannot always detect the tiers that are associated with external storage automatically, unlike with internal arrays.

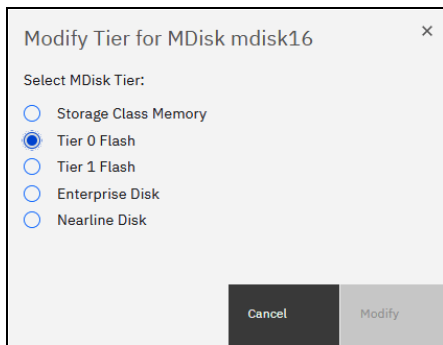


Figure 5-72 Modifying an external MDisk tier

For more information about storage tiers and their importance, see Chapter 9, “Advanced features for storage efficiency” on page 509.

To change the external MDisk storage tier, run the `chmdisk` command. Example 5-29 shows setting the new tier to `mdisk2`. No feedback is returned.

Example 5-29 Changing the tier setting by using the CLI

```
IBM_IBM FlashSystem 7200:superuser>chmdisk -tier tier1_flash mdisk16
IBM_IBM FlashSystem 7200:superuser>
```

Modify Encryption

To modify the encryption setting for the MDisk, select **Modify Encryption**. This option is available only when encryption is enabled.

If the external MDisk is already encrypted by the external storage system, change the encryption state of the MDisk to **Externally encrypted**. This setting stops the system from encrypting the MDisk again if the MDisk is part of an encrypted storage pool.

For more information about encryption, encrypted storage pools, and self-encrypting MDisks, see Chapter 12, “Encryption” on page 735.

To perform this task by using the CLI, run the `chmdisk` command, as shown in Example 5-30.

Example 5-30 Using chmdisk to modify the encryption

```
IBM_IBM FlashSystem 7200:superuser>chmdisk -encrypt yes mdisk5
IBM_IBM FlashSystem 7200:superuser>
```

Import

This action is available only for unmanaged MDisks. Importing an unmanaged MDisk enables you to preserve the existing data on the MDisk. You can migrate the data to a new volume or keep the data on the external system.

MDisks are imported for storage migration. The system provides a migration wizard to help with this process, which is described in Chapter 8, “Storage migration” on page 485.

Note: The wizard is the preferred method to migrate data from legacy storage to the system. When an MDisk is imported, the data on the original LU is not modified. The system acts as a pass-through, and the extents of the imported MDisk do not contribute to storage pools.

To choose one of the following migration methods, select **Import**:

► Import to temporary pool as image mode volume

This method does not migrate data from the source MDisk. It creates an *image mode volume* that has a direct block-for-block conversion of the MDisk. The existing data is preserved on the external storage controller, but it is also accessible from the system.

In this method, the image mode volume is created in a temporary migration pool and is presented through the system. Choose the extent size of the temporary pool and click **Import**, as shown in Figure 5-73 on page 295.

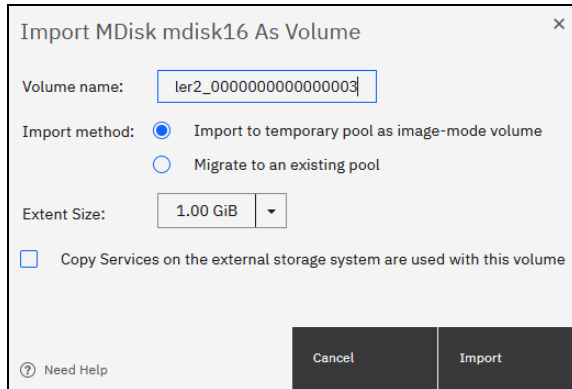


Figure 5-73 Importing an unmanaged MDisk

The MDisk is imported and listed as an image mode MDisk in the temporary migration pool, as shown in Figure 5-74.

Name	State	Usable Capacity	Written Capacity Limit
Unassigned MDisks (2)			
FCM-Pool	Online	3.87 TiB / 42.33 TiB (9%)	8.92 TiB / 195.07 TiB (5%)
MigrationPool_1024	Online		756.00 GiB / 756.00 GiB (100%)
mdisk16	Online	13.48 TiB	756.00 GiB

Figure 5-74 Image mode imported MDisk

A corresponding image mode volume is now available in the same migration pool, as shown in Figure 5-75.

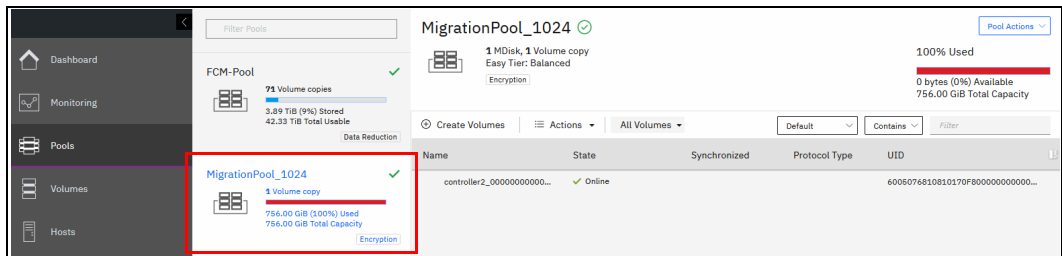


Figure 5-75 Image mode volume

The image mode volume can then be mapped to the original host. The data is still physically present on the MDisk of the original external storage controller and no automatic migration process is running. The original host sees no difference and its applications can continue to run. The image mode volume is now under the control of the system and it can optionally be migrated to another storage pool or be converted from image mode to a striped virtualized volume. You can use the **Volume Migration** wizard or perform the tasks manually.

► **Migrate to an existing pool**

This method starts by creating an image mode volume like the first method. However, it then automatically migrates the image mode volume to a virtualized volume in the selected storage pool. Free extents must be available in the selected target pool so that the data can be copied there.

If this method is selected, choose the storage pool to hold the new volume and click **Import**, as shown in Figure 5-76.

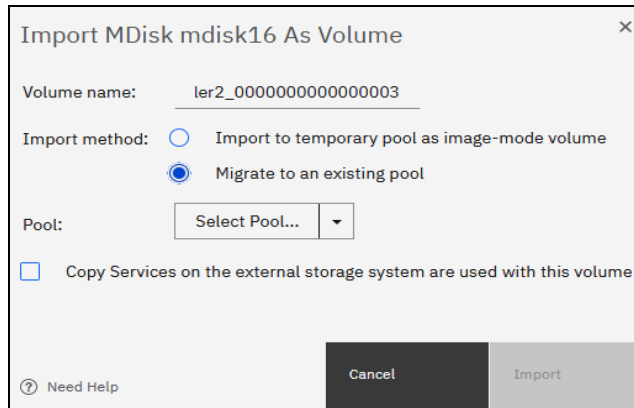


Figure 5-76 Migrating an MDisk to an existing pool

The data migration begins automatically after the MDisk is imported successfully as an image mode volume. You can check the migration progress by clicking the task under **Running Tasks**, as shown in Figure 5-77.

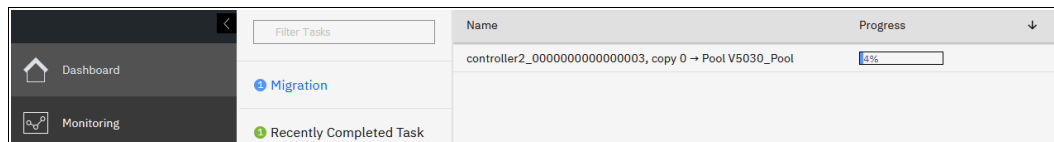


Figure 5-77 MDisk migration in the Running Tasks window

After the migration completes, the volume is available in the chosen destination pool. This volume is no longer an image mode volume. It is now virtualized by the system.

All data is migrated off the source MDisk, and the MDisk has switched its mode, as shown in Figure 5-78.

MDisk Name	Mode	Pool	Status	Volume ID	Controller
mdisk13	Managed	V5030_Pool	Online	0000000000000000	controller2
mdisk14	Unmanaged		Online	0000000000000001	controller2
mdisk15	Managed	V5030_Pool	Online	0000000000000002	controller2
mdisk16	Managed	MigrationPool_4096	Online	0000000000000003	controller2
mdisk17	Managed	V5030_Pool	Online	0000000000000004	controller2

Figure 5-78 Imported MDisk appears as Managed

The MDisk can be removed from the migration pool. It returns to the list of external MDisk as Unmanaged. The MDisk can now be used as a regular managed MDisk in a storage pool, or it can be decommissioned.

Alternatively, importing and migrating external MDisk to another pool can be done by selecting **Pools** → **System Migration** to start the system migration wizard. For more information, see Chapter 8, “Storage migration” on page 485.

Include

The system can exclude an MDisk from its storage pool if it has multiple I/O failures or has persistent connection errors. Exclusion ensures that there is no excessive error recovery that might impact other parts of the system. If an MDisk is automatically excluded, run the DMP to resolve any connection and I/O failure errors.

If no error event is associated with the MDisk in the log and the external problem is corrected, click **Include** to add the excluded MDisk back to the storage pool.

The `includemdisk` command performs the same task. The command needs the MDisk name or ID to be provided as a parameter, as shown in Example 5-31.

Example 5-31 Including a degraded MDisk by using the CLI

```
IBM_IBM FlashSystem 7200:superuser>includemdisk mdisk3
IBM_IBM FlashSystem 7200:superuser>
```

Remove

In some cases, you might want to remove external MDisks from their storage pool. To remove the MDisk from the storage pool, click **Remove**. After the MDisk is removed, it goes back to the Unmanaged state. If there are no volumes in the storage pool to which this MDisk is allocated, the command runs immediately without extra confirmation. If there are volumes in the pool, you are prompted to confirm the action, as shown in Figure 5-79.

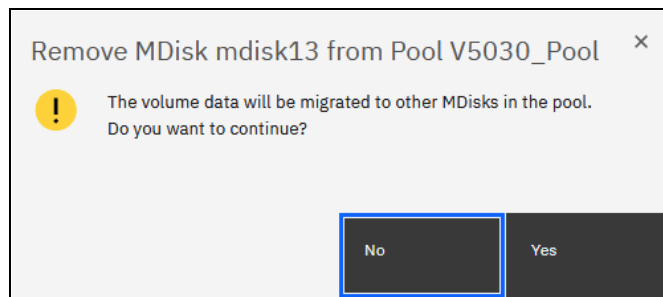


Figure 5-79 Removing an MDisk from a pool

Confirming the action starts a migration of volumes to extents on that MDisk to other MDisks in the pool. During this background process, the MDisk remains a part of the storage pool. Only when the migration completes is the MDisk removed from the storage pool and returns to Unmanaged mode.

Ensure that you have enough available capacity remaining in the storage pool to allocate the data being migrated from the removed MDisk, or this command fails.

Important: The MDisk that you are removing must remain accessible to the system while all data is copied to other MDisks in the same storage pool. If the MDisk is unmapped before the migration finishes, all volumes in the storage pool go offline and remain in this state until the removed MDisk is connected again.

To remove an MDisk from a storage pool by using the CLI, run the `rmmdisk` command. You must use the `-force` parameter if you must migrate volume extents to other MDisks in a storage pool.

The command fails if you do not have enough available capacity remaining in the storage pool to allocate the data that you are migrating from the removed array.

Dependent Volumes

A volume depends on an MDisk if the MDisk becoming unavailable results in a loss of access or a loss of data for that volume. Use this option before you do maintenance operations to confirm which volumes (if any) are affected. Selecting an MDisk and clicking **Dependent Volumes** lists the volumes that depend on that MDisk.

You can get the same information by running the `lsdependentvdisks` command.

View Provisioning Groups

Provisioning groups are used for capacity reporting and monitoring of overprovisioned external storage controllers. Each overprovisioned MDisk is part of a provisioning group that defines the physical storage resources that are available to a set of MDisks. Storage controllers report the usable capacity of an overprovisioned MDisk based on its provisioning group. If multiple MDisks are part of the same provisioning group, then these MDisks share the physical storage resources and report the same usable capacity. However, this usable capacity is not available to each MDisk individually because it is shared among all these MDisks.

To know the usable capacity that is available to the system or to a pool when overprovisioned storage is used, you must account for the usable capacity of each provisioning group. To show a summary of overprovisioned external storage, including controllers, MDisks, and provisioning groups, click **View Provisioning Groups**, as shown in Figure 5-80.

For more information, see 9.6, “Overprovisioning and data reduction on external storage” on page 548.

View Provisioning Groups

Select an overprovisioned external storage system to view its related MDisks and pools.

Configure capacity alerts
Ensure that you configure alerts for usable capacity consumption on all external storage systems. When an external storage system uses 100% of usable capacity, volumes that are dependent on that storage go offline.

controller2
Usable capacity: 7.42 TiB / 13.48 TiB (55%)
Provisioning Group ID: 0
MDisks: 8

Usable capacity: 4.03 TiB / 42.33 TiB (10%)
Provisioning Group ID: 1
MDisks: 1

MDisk Name	State	Written Capacity Limit	Pool	Storage System - LU
mdisk13	Online	756.00 GiB	V5030_Pool	controller2 - 000000
mdisk14	Online	756.00 GiB	V5030_Pool	controller2 - 000000
mdisk15	Online	756.00 GiB	V5030_Pool	controller2 - 000000

Need Help

Close

Figure 5-80 View Provisioning Groups



Volumes

In IBM Spectrum Virtualize, a *volume* is storage space that is provisioned out of a storage pool and presented to a host as a Small Computer System Interface (SCSI) logical unit (LU), that is, a logical disk.

This chapter describes how to create and provision volumes on IBM Spectrum Virtualize systems. The first part of this chapter provides a brief overview of IBM Spectrum Virtualize volumes, the classes of volumes that are available, and the available volume customization options.

The second part of this chapter describes how to create, modify, and map volumes by using the GUI.

The third part of this chapter provides an introduction to volume manipulation from the command-line interface (CLI).

This chapter includes the following topics:

- ▶ 6.1, “Introduction to volumes” on page 300
- ▶ 6.2, “Volume characteristics” on page 300
- ▶ 6.3, “Virtual volumes” on page 318
- ▶ 6.4, “Volumes in multi-site topologies” on page 319
- ▶ 6.5, “Operations on volumes” on page 321
- ▶ 6.6, “Volume operations by using the CLI” on page 376

6.1 Introduction to volumes

For an IBM Spectrum Virtualize system cluster, the volume that is presented to a host is internally represented as a virtual disk (VDisk). A *VDisk* is an area of usable storage that was allocated out of a pool of storage that is managed by an IBM Spectrum Virtualize cluster. The term *virtual* is used because the volume that is presented does not necessarily exist on a single physical entity.

Note: Volumes are composed of extents that are allocated from a storage pool. Storage pools group managed disks (MDisks), which are redundant arrays of independent disks (RAIDs) that are configured by using internal storage, or LUs that are presented to and virtualized by an IBM Spectrum Virtualize system. Each MDisk is divided into sequentially numbered extents (zero-based indexing). The extent size is a property of a storage pool, and is used for all MDisks that make up the storage pool.

MDisks are internal objects that are used for storage management. They are not directly visible to or used by host systems.

Every volume is presented to hosts by an I/O group. One of nodes within that group is defined as a preferred node, that is, a node that by default serves I/O requests to that volume. When a host requests an I/O operation to a volume, the multipath driver on the host identifies the preferred node for the volume and by default uses only paths to this node for I/O requests.

6.2 Volume characteristics

There are several parameters that characterize each volume. They should be set correctly to match the requirements of the storage user (an application running on a host). These characteristics are:

- ▶ Size
- ▶ Performance (input/output operations per second (IOPS), response time, and bandwidth)
- ▶ Resiliency
- ▶ Storage efficiency
- ▶ Security (data-at-rest encryption)
- ▶ Extent allocation policy
- ▶ Management mode

Additionally, volumes can be configured as VMware vSphere Virtual Volumes (VVOLs).

VVOLs change the approach to VMware virtual machines (VMs) disk configuration from “The VM disk is a file on a VMware Virtual Machine File System (VMFS) volume” to one-to-one mapping between VM disks and storage volumes. VVOLs can be managed by the VMware infrastructure so that the storage system administrator can delegate VM disk management to VMware infrastructure specialists, which greatly simplifies storage allocation for virtual infrastructure and reduces the storage management team’s effort that is required to support VMware infrastructure administrators.

The downside of using VVOLs is multiplication of the number of volumes that are presented by a storage system because typically there are multiple VM disks that are configured on every VMFS volume. As excessive proliferation of volumes that is presented to Elastic Sky X Integrated (ESXi) clusters can have a negative impact on performance. Therefore, it is a best practice to carefully plan a storage system configuration before production deployment and include in the assessment the projected system growth.

Note: If there are too many logical unit numbers (LUNs) that are presented to a sufficiently large ESXi cluster, I/O requests that are simultaneously generated by ESXi hosts might exceed the storage system command queue. Such overflow leads to I/O request retries, which reduce storage system performance as perceived by the connected hosts.

To provide storage users adequate service, all parameters must be correctly set. Importantly, the various parameters might be *interdependent*, that is, setting one of them might affect other properties of the volume.

The volume parameters and their interdependencies are covered in the following sections.

6.2.1 Volume type

The *type* attribute of a volume defines the method of allocation of extents that make up the volume copy:

- ▶ A *striped* volume contains a volume copy that has extents that are allocated from multiple MDisks from the storage pool. By default, extents are allocated from all MDisks in the storage pool by using a round-robin algorithm. However, it is possible to supply a list of MDisks to use for volume creation.

Attention: By default, striped volume copies are striped across all MDisks in the storage pool. If some of the MDisks are smaller than others, the extents on the smaller MDisks are used up before the larger MDisks run out of extents. Manually specifying the stripe set in this case might result in the volume copy not being created.

If you are unsure whether sufficient free space is available to create a striped volume copy, use one of the following approaches:

- ▶ Check the free space on each MDisk in the storage pool by running the `lsfreeextents` command, and ensure that each MDisk that is included in the manually specified stripe set has enough free extents.
- ▶ Allow the system to automatically create the volume copy by not supplying a specific stripe set.

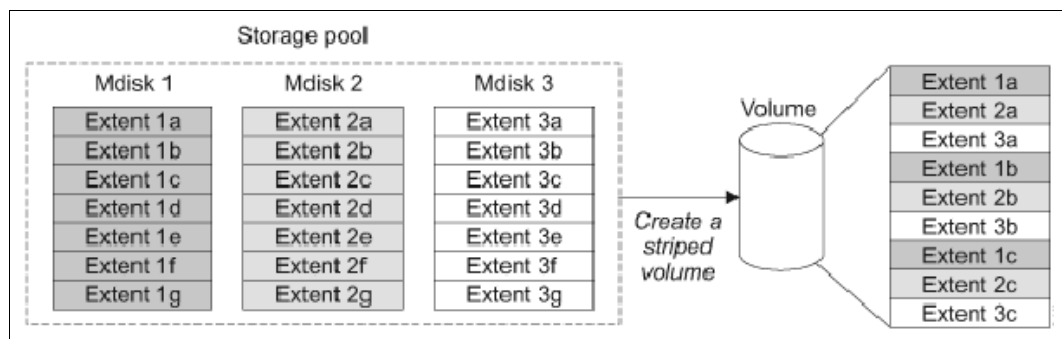


Figure 6-1 Striped extent allocation

- ▶ A *sequential* volume contains a volume copy with extents that are allocated sequentially on one MDisk.
- ▶ An *image mode* volume is a special type of volume that has a direct one-to-one mapping to one (image mode) MDisk.

For striped volumes, the extents are allocated from the set of MDisks (by default, all MDisks in the storage pool):

- ▶ An MDisk is picked by using a pseudo-random algorithm and an extent is allocated from this MDisk. This approach minimizes the probability of triggering the *striping effect*, which might lead to poor performance for workloads that generate many metadata I/Os, or that create multiple sequential streams.
- ▶ All subsequent extents (if required) are allocated from the MDisk set by using a round-robin algorithm.
- ▶ If an MDisk has no free extents when its turn arrives, the algorithm moves to the next MDisk in the set that has a free extent.

Note: The *striping effect* occurs when multiple logical volumes that are defined on a set of physical storage devices (MDisks) store their metadata or file system transaction log on the same physical device (MDisk).

Because of the way the file systems work, system metadata disk regions are typically busy. For example, in a journaling file system, a write to a file might require two or more writes to the file system journal: At minimum, one to make a note of the intended file system update, and one marking the successful completion of the file write.

If multiple volumes (each with their own file system) are defined on the same set of MDisks, and all (or most) of them store their metadata on the same MDisk, a disproportionately large I/O load is generated on this MDisk, which can result in suboptimal performance of the storage system. Pseudo-randomly allocating the first MDisk for new volume extent allocation minimizes the probability that multiple file systems that are created on these volumes place their metadata regions on the same physical MDisk.

Note: Some file systems allow specifying different logical disks for data and metadata storage. When taking advantage of this file system feature, you may allocate differently configured volumes that are dedicated to data and metadata storage.

6.2.2 Managed mode and image mode

Volumes are configured within IBM Spectrum Virtualize by allocating a set of extents off one or more managed mode MDisks in the storage pool. *Extents* are the smallest allocation unit at the time of volume creation, so each MDisk extent maps to exactly one volume extent.

Note: An MDisk extent maps to exactly one volume extent. For volumes with two copies, one volume extent maps to two MDisk extents (one for each volume copy).

Figure 6-2 on page 303 shows this mapping. It also shows a volume that consists of several extents that are shown as V0 - V7. Each of these extents is mapped to an extent on one of the MDisks: A, B, or C. The mapping table stores the details of this indirection.

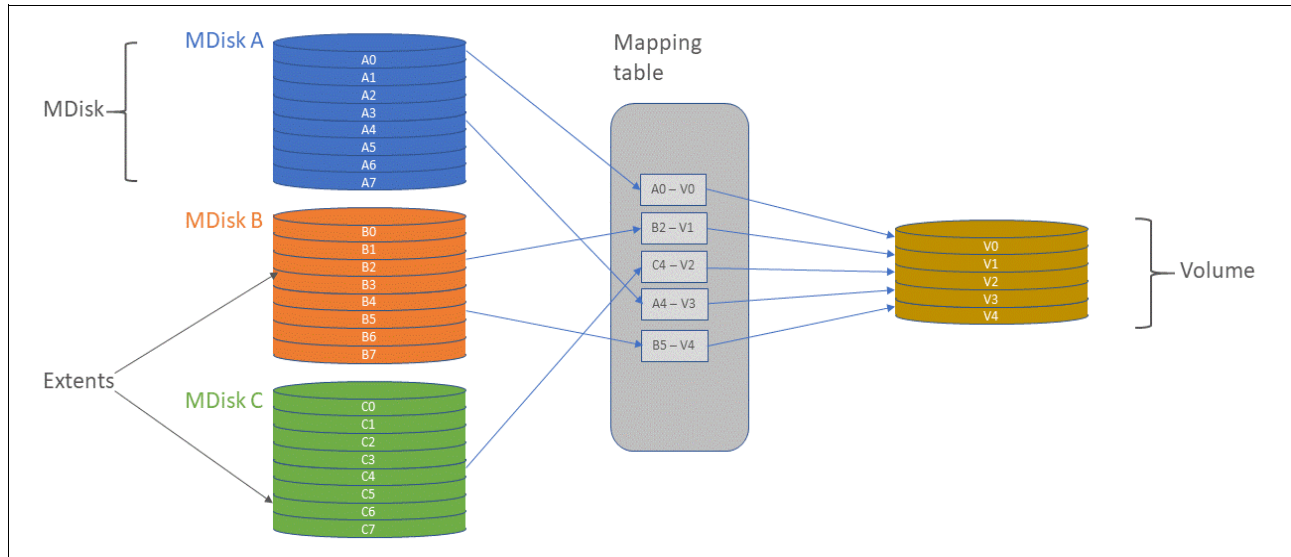


Figure 6-2 Simple view of block virtualization

Several of the MDisk extents are unused, that is, no volume extent maps to them. These unused extents are available for volume creation, migration, and expansion.

The default and most common type of volumes in IBM Spectrum Virtualize are managed mode volumes. *Managed mode volumes* are allocated from a set of MDisk belonging to a storage pool, and they can be subjected to the full set of virtualization functions. In particular, they offer full flexibility in mapping between logical volume representation (a continuous set of logical blocks) and the physical storage that is used to store these blocks. This function requires that physical storage (MDisks) is fully managed by IBM Spectrum Virtualize, which means that the LUs that are presented to IBM Spectrum Virtualize by the back-end storage systems do not contain any data when they are added to the storage pool.

Image mode volumes enable IBM Spectrum Virtualize to work with LUs that were previously directly mapped to hosts, which are often required when IBM Spectrum Virtualize is introduced into a storage environment. In such scenario, image mode volumes are used to enable seamless migration of data and a smooth transition to virtualized storage.

The image mode creates one-to-one mapping of logical block addresses (LBAs) between a volume and a single MDisk (a LU that is presented by the virtualized storage). Image mode volumes have a minimum size of one block (512 bytes) and always occupy at least one extent. An image mode MDisk cannot be used as a quorum disk and no IBM Spectrum Virtualize system metadata extents are allocated from it. All the IBM Spectrum Virtualize copy services functions can be applied to image mode disks.

The difference between a managed mode volume (with striped extent allocation) and an image mode volume is shown in Figure 6-3.

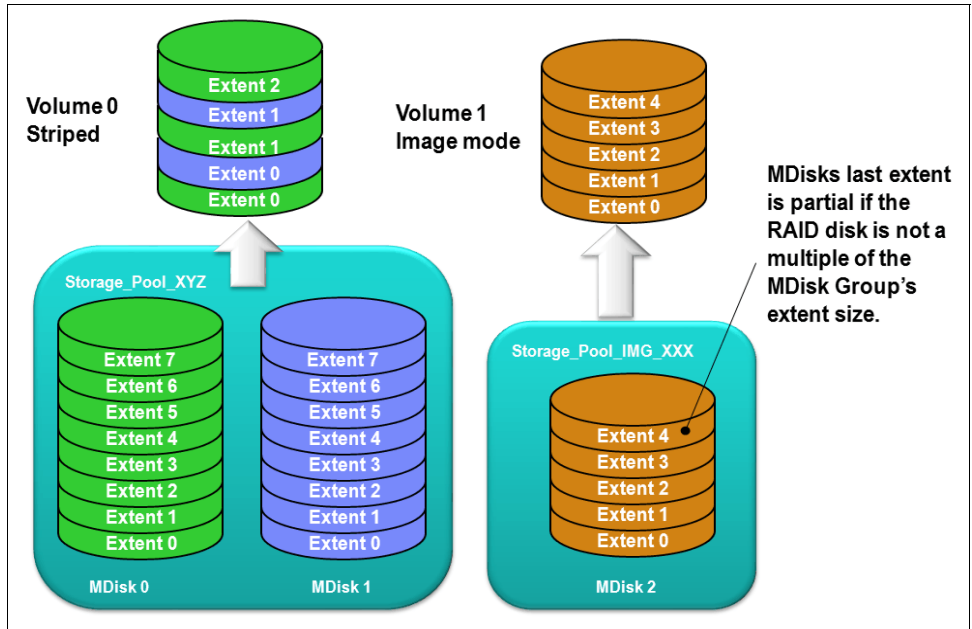


Figure 6-3 An image mode volume versus a striped volume

An image mode volume is mapped to only one image mode MDisk, and it is mapped to the entirety of this MDisk. Therefore, the image mode volume capacity is equal to the size of the corresponding image mode MDisk. If the size of the (image mode) MDisk is not a multiple of the MDisk group's extent size, the last extent is marked as partial (not filled).

When you create an image mode volume, you map it to an MDisk that must be in unmanaged mode and must not be a member of a storage pool. As the image mode volume is configured, the MDisk is made a member of the specified storage pool. It is a best practice to use a dedicated pool for image mode MDisks with a name indicating its role, such as Storage Pool_IMG_XXX.

An image mode volume can be migrated to a managed mode volume, which is a standard procedure that is used to perform non-disruptive migration of the organization's SAN to an environment managed by or based on IBM Spectrum Virtualize systems. After the data is migrated off the managed image volume, the space it used on the source storage system can be reclaimed. After all data is migrated off the storage system, it can be decommissioned or used as a back-end storage system that is managed by the IBM Spectrum Virtualize system (see 2.9, "Back-end storage configuration" on page 88).

IBM Spectrum Virtualize also supports the reverse process in which a managed mode volume can be migrated to an image mode volume. During the migration, the volume is identified in the system as being in managed mode. Its mode changes to "image" only after the process completes.

6.2.3 VSize

Each volume has two associated values that describe its size: real capacity and virtual capacity.

- ▶ The real (physical) capacity is the size of storage space that is allocated to the volume from the storage pool. It determines how many MDisk extents are allocated to form the volume. The real capacity is used to store the user data, and in the case of thin-provisioned volumes, the metadata of the volume.
- ▶ The virtual capacity is capacity that is reported to the host, but also any other IBM Spectrum Virtualize components or functions (for example, IBM FlashCopy, cache, and Remote Copy (RC)) that operate based on a volume size.

In a standard-provisioned volume, the real and virtual capacities are the same. In a thin-provisioned volume, the real capacity can be as little as a few percent of virtual capacity. The volume size can be specified in units down to 512-byte blocks (see Figure 6-4). The real capacity can be specified as an absolute value or as a percentage of the virtual capacity.

The screenshot displays the 'Create Volumes' configuration window. It features three tabs: 'Basic' (selected), 'Mirrored', and 'Custom'. Below the tabs, a message states: 'Create a preset volume with all the basic features.' The 'Pool:' field is set to 'Click to select.' The 'Capacity Details:' section includes a progress bar and the text 'Total 0 bytes'. The 'Volume Details' section contains a table with columns for 'Quantity', 'Capacity', and 'Name'. The 'Quantity' is '1', 'Capacity' is '512', and the unit dropdown is open, showing options: 'bytes', 'KiB', 'MiB', 'GiB', and 'TiB'. Below this, 'Capacity savings:' is set to 'None'. There is a '+ Define another volume' link. The 'I/O group:' is set to 'Automatic'. At the bottom, there is a 'Summary' section with 'Fields Incomplete' and three empty boxes. The footer contains a 'Need Help' link, a 'Cancel' button, and 'Create and Map' and 'Create' buttons.

Figure 6-4 Smallest possible volume size

A volume is composed of storage pool extents, so it is not possible to allocate less than one extent to create a volume. Effectively, the internal unit of volume size is the extent size of the pool (or pools) in which the volume is created.

For example, a basic volume of 512 bytes that is created in a pool with the default extent size (1024 mebibytes (MiB)) uses 1024 MiB of the pool space because a whole extent must be allocated to provide the space for the volume.

In practice, this rounding up of volume size to the whole number of extents has little impact on storage use efficiency unless the storage system serves many small volumes. For more information about storage pools and extents, see Chapter 5, “Storage pools” on page 237.

6.2.4 Performance

The basic metrics of volume performance are the number of IOPS the volume can provide, the time to service an I/O request (average, median, and first percentile), and the bandwidth of the data that is served to a host.

Volume performance is defined by the pool or pools that are used to create the volume. The pool determines the media bus (Non-Volatile Memory Express (NVMe) or serial-attached SCSI (SAS)); media type (IBM FlashCore Module (FCM) drives, solid-state drives (SSDs), or hard disk drives (HDDs)); redundant array of independent disks (RAID) level and number of drives per RAID array; and the possibility for the Easy Tier function to optimize the performance of a volume. However, volumes that are configured in the same storage pool or pools might still have different performance characteristics, depending on the storage resiliency, efficiency, security, and allocation policy configuration settings of a volume.

6.2.5 Volume copies

A volume can have one or two physical copies. Each copy of the volume has the same virtual capacity, but the two copies can have different characteristics, including different real capacity. However, each volume copy is not a separate object and can be manipulated only in the context of the volume. A mirrored volume behaves in the same way as any other volume, such as:

- ▶ All its copies are expanded or shrunk when the volume is resized.
- ▶ It can participate in FlashCopy and RC relationships.
- ▶ It is serviced by an I/O group.
- ▶ It has a preferred node.

Volume copies are identified in the GUI by a copy ID, which can have value 0 or 1. Copies of the volume can be split, which provides a point-in-time (PIT) copy of a volume. An overview of volume mirroring is shown in Figure 6-5 on page 307.

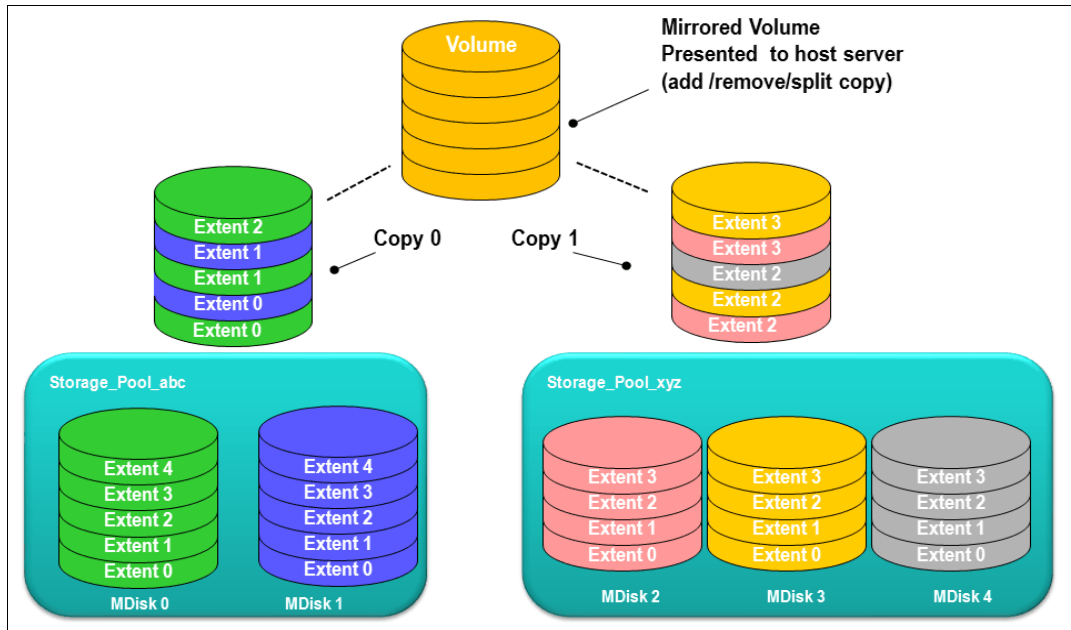


Figure 6-5 Volume mirroring overview

A copy can be added to a volume with a single copy or removed from a volume with two copies. Internal safety mechanisms prevent accidental removal of the only remaining copy of a volume.

A newly created, unformatted volume with two copies initially has the two copies in an out-of-synchronization state. The primary copy is defined as “fresh” and the secondary copy is defined as “stale”, and the volume is immediately available for use.

The synchronization process updates the secondary copy until it is fully synchronized, that is, data that is stored on the secondary copy matches the data that is on the primary copy. This update is done at the *synchronization rate* that is defined when the volume is created, but can be modified after volume creation. The synchronization status for mirrored volumes is recorded on the storage system quorum disk.

If a mirrored volume is created by using the **format** parameter, both copies are formatted in parallel. The volume comes online when both operations are complete with the copies in sync.

If it is known that MDisk space (which is used for creating volume copies) is formatted or if the user does not require read stability, a **no synchronization** option can be used that declares the copies as synchronized even when they are not.

Creating volume with more than one copy is beneficial in multiple scenarios. For example:

- ▶ Improving volume resilience by protecting it from a single back-end storage system failure (requires each volume copy to be configured on a different back-end storage system).
- ▶ Providing concurrent maintenance of a storage system that does not natively support concurrent maintenance (for volumes on external virtualized storage).
- ▶ Providing an alternative method of data migration with improved availability characteristics. While a volume is being migrated by using the data migration feature, it is vulnerable to failures on both the source and target storage pool. Volume mirroring provides an alternative migration method that is not affected by the destination volume pool availability.

For more information about this volume migration method, see “Volume migration by adding a volume copy” on page 372.

Note: When migrating volumes to a Data Reduction Pool (DRP), volume mirroring is the only migration method because DRPs do not support `migrate` commands.

- ▶ Converting between standard-provisioned volumes and thin-provisioned volumes (in either direction).

Typically, each volume copy is allocated from a different storage pool. Although not required, using different pools that are backed by different back-end storage for each volume copy is the typical configuration because it markedly increases volume resiliency.

If one of the mirrored volume copies becomes temporarily unavailable (for example, because the storage system that provides its pool is unavailable), the volume remains accessible to hosts. The storage system remembers which areas of the volume were modified after the loss of access to a volume copy and resynchronizes only these areas when both copies are available.

Note: Volume mirroring is not a disaster recovery (DR) solution because both copies are accessed by the same node pair and addressable by only a single cluster. However, if correctly planned, it can improve availability.

The storage system tracks the synchronization status of volume copies by dividing the volume into 256 kibibyte (KiB) grains and maintaining a bitmap of stale grains (on the quorum disk), mapping 1 bit to one grain of the volume space. If the mirrored volume needs resynchronization, the system copies to the out-of-sync volume copy only these grains that were written to (changed) since the synchronization was lost. This approach is known as an *incremental synchronization*, and it minimizes the time that is required to synchronize the volume copies.

Important: Mirrored volumes can be taken offline if no quorum disk is available. This behavior occurs because the synchronization status of mirrored volumes is recorded on the quorum disk.

A volume with more than one copy can be checked to see whether all of the copies are identical or consistent. If a medium error is encountered while it is reading from one copy, a check is repaired by using data from the other copy. This consistency check is performed asynchronously with host I/O.

Because mirrored volumes use bitmap space at a rate of 1 bit per 256 KiB grain, 1 MiB of bitmap space supports up to 2 TiB of mirrored volumes. The default size of the bitmap space is 20 MiB, which allows a configuration of up to 40 TiB of mirrored volumes. If all 512 MiB of variable bitmap space is allocated to mirrored volumes, 1 PiB of mirrored volumes can be supported.

Table 6-1 on page 309 lists the bitmap space configuration options.

Table 6-1 Bitmap space default configuration

Copy service	Minimum allocated bitmap space	Default allocated bitmap space	Maximum allocated bitmap space	Minimum ^a capacity when using the default values
RC ^b	0	20 MiB	512 MiB	40 TiB of remote mirroring volume capacity
FlashCopy ^c	0	20 MiB	2 GiB	<ul style="list-style-type: none"> ▶ 10 TiB of FlashCopy source volume capacity ▶ 5 TiB of incremental FlashCopy source volume capacity
Volume mirroring	0	20 MiB	512 MiB	40 TiB of mirrored volumes
RAID	0	40 MiB	512 MiB	<ul style="list-style-type: none"> ▶ 80 TiB array capacity by using RAID 0, 1, or 10 ▶ 80 TiB array capacity in three-disk RAID 5 array ▶ Slightly less than 120 TiB array capacity in five-disk RAID 6 array

- a. The actual amount of available capacity might increase based on the settings, such as grain size and strip size. RAID is subject to a 15% margin of error.
- b. RC includes Metro Mirror (MM), Global Mirror (GM), and active-active relationships.
- c. FlashCopy includes the FlashCopy function, Global Mirror with Change Volumes (GMCV), and active-active relationships.

The sum of all bitmap memory allocation for all functions except FlashCopy must not exceed 552 MiB.

6.2.6 I/O operations data flow

Although a mirrored volume looks to its users the same as a volume with a single copy, some differences exist in how I/O operations are performed internally for volumes with single or two copies.

Read I/O operations data flow

If the volume is mirrored (that is, two copies of the volume exist), one copy is known as the *primary copy*. If the primary copy is available and synchronized, host read requests are directed to that copy. The choice of the primary copy is part of initial configuration of a mirrored volume, but this setting can be changed at any time. In the management GUI, an asterisk indicates the primary copy of the mirrored volume. Placing the primary copy on a high-performance controller maximizes the read performance of the volume.

For non-mirrored volumes, only one volume copy exists, so no choice exists for the read source, and all reads are directed to the single volume copy.

Write I/O operations data flow

The host sends all write I/O operation requests to any volume to the preferred node for this volume. The preferred node is responsible for destaging the data from cache to persistent storage. Figure 6-6 shows the data flow for this scenario.

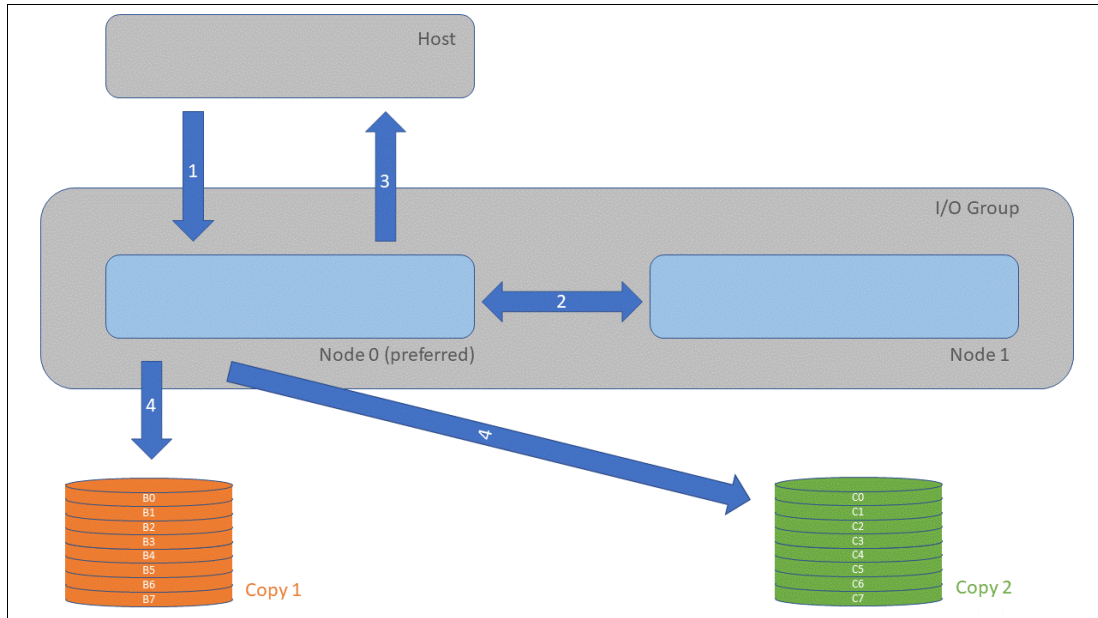


Figure 6-6 Data flow for write I/O processing in a mirrored volume

As shown in Figure 6-6, the writes are sent by the host to the preferred node for the volume (1). Then, the data is mirrored to the cache of the partner node in the I/O group (2), and acknowledgment of the write operation is sent to the host (3). The preferred node then destages the written data to all volume copies (4). The example that is shown in Figure 6-7 on page 311 shows a case with destaging to a mirrored volume, that is, one with two physical data copies.

With Version 7.3, the cache architecture changed from an upper-cache design to a two-layer cache design. With this change, the data is written once, and then it is directly destaged from the controller to the disk system.

Figure 6-7 on page 311 shows the data flow in a stretched environment.

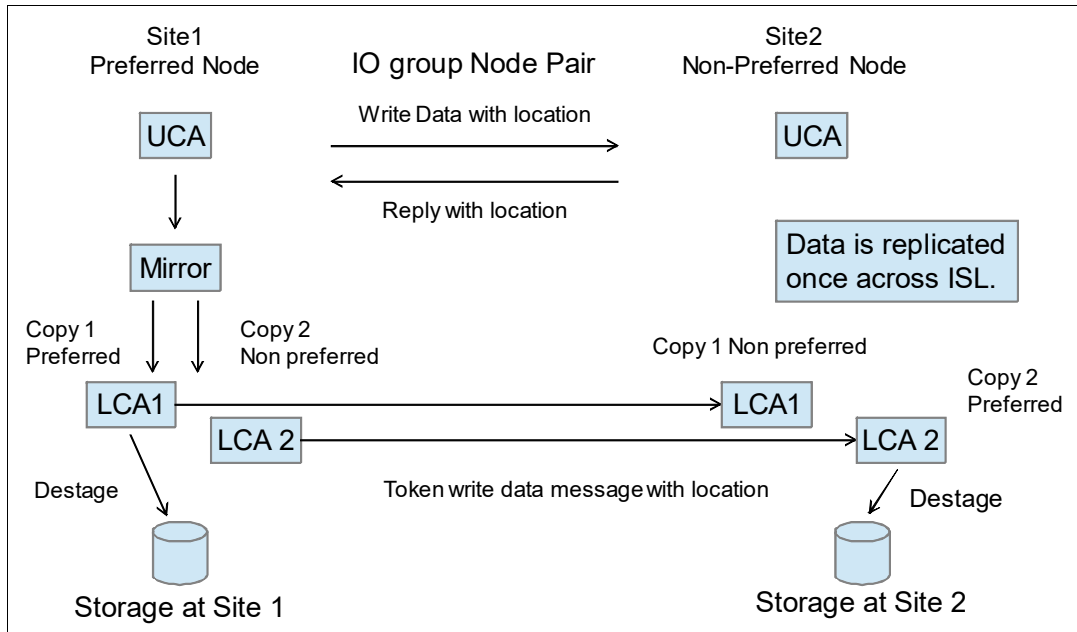


Figure 6-7 Design of an enhanced stretched cluster

6.2.7 Storage efficiency

When aiming for the maximum efficiency of data storage with IBM Spectrum Virtualize, you can configure DRPs that provide several technologies that increase the efficiency of physical storage use:

- ▶ Thin provisioning
- ▶ Deduplication (block-level and pattern-matching)
- ▶ Compression
- ▶ SCSI UNMAP support

Note: Storage efficiency options might require more licenses and hardware components depending on the model and configuration of your storage system.

Implementation of DRPs requires careful planning and sizing. Before configuring the first space-efficient volume on a storage system, see the relevant sections in Chapter 2, “Planning” on page 71 and Chapter 9, “Advanced features for storage efficiency” on page 509.

DRPs use multithreading and hardware acceleration (where available) to provide storage efficiency functions on IBM Spectrum Virtualize storage systems. When you consider using storage efficiency options, remember that they increase the number of I/O operations that the storage system must realize compared to accessing a basic volume. Space-efficient volumes require the storage system to both to write the data that is sent by the host and the metadata that is required to maintain a space-efficient volume.

Note: FCM drives include compression hardware, so it provides data set size reduction with no performance penalty.

For more information about the storage efficiency functions of IBM Spectrum Virtualize, see Chapter 5, “Storage pools” on page 237 and *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430.

It is possible to benefit from both compression and data-at-rest encryption because encryption is done after compression. However, the size of data that is encrypted at the host level is unlikely to be reduced by either compression or deduplication at the storage system.

Standard and thin-provisioned volumes

A standard-provisioned volume directly maps logical blocks on the virtual volume to physical blocks on storage media. Therefore, its virtual and physical capacities are identical.

A thin-provisioned volume has virtual capacity larger than physical capacity. Thin-provisioning is the base technology for all space-efficient volumes. When a thin-provisioned volume is created, a small amount of the real capacity is used for initial metadata. This metadata holds a mapping of a set of continuous LBAs in the volume to a *grain* on a physically allocated extent.

Note: If you use of thin-provisioned volumes, then it is recommended to monitor closely the available space in the pool that contains these volumes. If a thin-provisioned volume does not have enough real capacity for a write operation, the volume is taken offline and an error is logged. There is limited ability to recover with UNMAP. Also, consider creating a fully allocated sacrificial emergency space volume.

The grain size is defined when the volume is created and cannot be changed afterward. The grain size can be 32 KiB, 64 KiB, 128 KiB, or 256 KiB. The default grain size is 256 KiB, which is the preferred option. However, the following factors must be considered when deciding on the grain size:

- ▶ A smaller grain size helps to save space. If a 16 KiB write I/O requires a new physical grain to be allocated, the used space is 50% of a 32 KiB grain, but just over 6% of 256 KiB grain. If no subsequent writes to other blocks of the grain occur, the volume provisioning is less efficient for volumes with larger grain.
- ▶ A smaller grain size requires more metadata I/O to be performed, which increases the load on the physical back-end storage systems.
- ▶ When a thin-provisioned volume is a FlashCopy source or target volume, specify the same grain size for FlashCopy and the thin-provisioned volume configuration. Use 256 KiB grain to maximize performance.
- ▶ The grain size affects the maximum size of the thin-provisioned volume. For 32 KiB size, the volume size cannot exceed 260 TiB.

Figure 6-8 on page 313 shows the thin-provisioning concept.

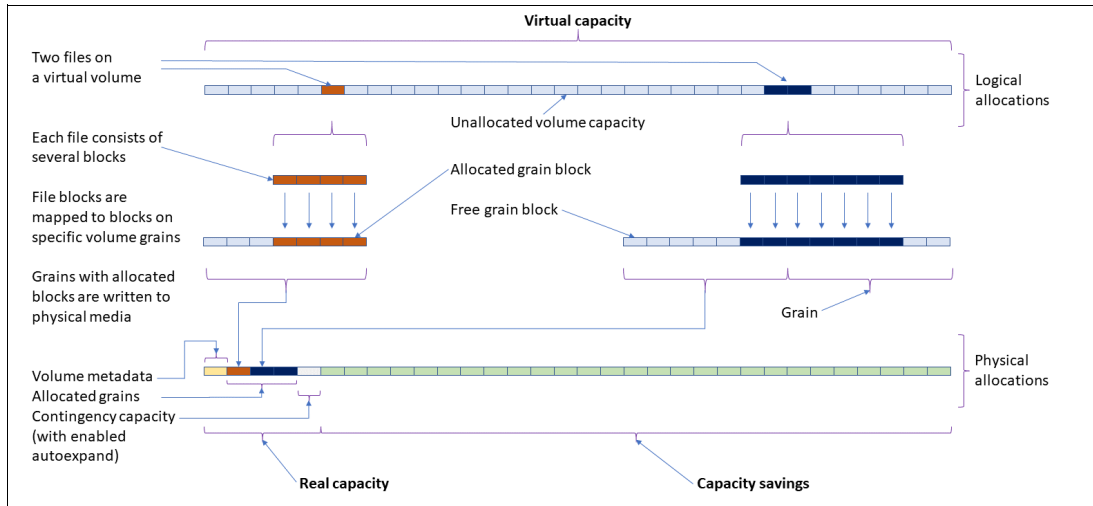


Figure 6-8 Conceptual diagram of a thin-provisioned volume

Thin-provisioned volumes use metadata to enable capacity savings, and each grain of user data requires metadata to be stored. Therefore, the I/O rates that are obtained from thin-provisioned volumes are lower than the I/O rates that are obtained from standard-provisioned volumes.

When a write request comes from a host, the block address for which the write is requested is checked against the mapping table. If the write is directed to a block that maps to a grain with physical storage that is allocated by a previous write, then physical storage was allocated for this LBA and can be used to service the request. Otherwise, a new physical grain is allocated to store the data, and the mapping table is updated to record that allocation.

The metadata storage that is used is never greater than 0.1% of the user data. The resource usage is independent of the virtual capacity of the volume.

Thin-provisioned volume format: Thin-provisioned volumes do not need formatting. A read I/O, which requests data from unallocated data space, returns zeros. When a write I/O causes space to be allocated, the grain is “zeroed” before use. Additionally, when a full-grain write consists of “all zeros”, no space is physically allocated on disk.

The real capacity of a thin-provisioned volume can be changed if the volume is not in image mode. Thin-provisioned volumes use the grains of real capacity that is provided in ascending order as new data is written to the volume. If the user initially assigns too much real capacity to the volume, the real capacity can be reduced to free storage for other uses.

A thin-provisioned volume can be configured to *autoexpand*. This feature causes IBM Spectrum Virtualize to automatically add a fixed amount of extra real capacity to the thin-provisioned volume as required. Autoexpand does not cause the real capacity to grow much beyond the used capacity. Instead, it attempts to maintain a fixed amount of unused real capacity for the volume, which is known as the *contingency capacity*.

The contingency capacity is initially set to the real capacity that is assigned when the volume is created. If the user modifies the real capacity, the contingency capacity is reset to be the difference between the used capacity and real capacity.

A volume that is created without the autoexpand feature and has zero contingency capacity goes offline when the real capacity is used, and it receives a write request that requires real capacity allocation.

To facilitate management of the auto expansion of thin-provisioned volumes, a capacity warning should be set for the storage pools from which they are allocated. When the used capacity of the pool exceeds the warning capacity, a warning event is logged. For example, if a warning of 80% is specified, an event is logged when 20% of the pool capacity remains free.

A thin-provisioned volume can be converted nondisruptively to a standard-provisioned volume (or vice versa) by using the volume mirroring function. You can create a thin-provisioned copy to a standard-provisioned primary volume and then remove the standard-provisioned copy from the volume after they are synchronized.

The standard-provisioned to thin-provisioned migration procedure uses a zero-detection algorithm so that grains that contain all zeros do not use up any real capacity.

Thin-provisioned volumes can be used as volumes that are assigned to the host by FlashCopy to implement thin-provisioned FlashCopy targets. When creating a mirrored volume, a thin-provisioned volume can be created as a second volume copy, whether the primary copy is a standard or thin-provisioned volume.

Deduplicated volumes

Deduplication is a specialized data set reduction technique. However, in contrast to the standard file-compression tools that work on single files or sets of files, deduplication is a technique that is applied on a block level to larger scale data sets, such as a file system or volume. In IBM Spectrum Virtualize, deduplication can be enabled for thin-provisioned and compressed volumes that are created in DRPs.

IBM Spectrum Virtualize uses two techniques to detect duplicate data:

- ▶ Pattern matching
- ▶ Data signature (hash)

Deduplication works by identifying repeating chunks in the data that is written to the storage system. Pattern matching looks for a known data patterns (for example, “all ones”), and the data signature-based algorithm calculates a signature for each data chunk (by using a hash function) and checks whether the calculated signature is present in the deduplication database.

If a known pattern or a signature match is found, the data chunk is replaced by a reference to a stored chunk, which reduces storage space that is required for storing the data. Conversely, if no match is found, the data chunk is stored without modification, and its signature is added to the deduplication database.

To maximize the space that is available for the deduplication database, the system distributes it between all nodes in the I/O groups that contain deduplicated volumes. Each node holds a distinct portion of the records that are stored in the database. If nodes are removed or added to the system, the database is redistributed between the anodes to ensure optimal use of available resources.

Depending on the data type that is stored on the volume, the capacity savings can be significant. Examples of use cases that typically benefit from deduplication are virtual environments with multiple VMs running the same operating system (OS), and backup servers. In both cases, it is expected, that multiple copies of identical files exist, such as components of the standard OS or applications that are used in the organization.

Note: If data is encrypted by the host, you should expect no benefit from deduplication because the same cleartext (for example, a standard OS library file) encrypted with different keys results in different output, making deduplication impossible.

Although deduplication (and other features of IBM Spectrum Virtualize) is transparent to users and applications, it must be planned for and understood before implementation because it might reduce the redundancy of a solution. For example, if an application stores two copies of a file to reduce the chances of data corruption because of a random event, the copies are deduplicated and the intended redundancy is removed from the system if these copies are on the same volume.

When planning the use of deduplicated volumes, be aware of update and performance considerations and the following software and hardware requirements:

- ▶ Code level V8.1.2 or higher is needed for DRPs.
- ▶ Code level V8.1.3 or higher is needed for deduplication.

Tip: Code level 8.3.1 is needed for the best performance in DRP pools.

- ▶ Nodes must have at least 32 GB to support deduplication. Nodes that have more than 64 GB can use a bigger deduplication fingerprint database, which might lead to better deduplication.
- ▶ You must run supported hardware. For more information about the valid hardware and features combinations, go to [IBM FlashSystem 9200 documentation](#), select your system, and read the “Planning for deduplicated volumes” section by expanding **Planning** → **Storage configuration planning**.

Compressed volumes

A volume that is created in a DRP can be compressed. Data that is written to the volume is compressed before committing it to back-end storage, which reduces the physical capacity that is required to store the data. Because enabling compression does not incur an extra metadata handling penalty, in most cases it is a best practice to enable compression on thin-provisioned volumes.

Notes:

- ▶ When a volume is backed by FCM drives that compress data at line speed, the volume should be configured with compression that is turned on. IBM Spectrum Virtualize is tightly integrated with the storage controller and uses knowledge of both the logical and physical space.
- ▶ You can use the management GUI or the CLI to run the built-in compression estimation tool. This tool can be used to determine the capacity savings that are possible for existing data on the system by using compression.
- ▶ Another benefit of data compression for volumes that are backed by flash-based storage is the reduction of write amplification, which has a beneficial effect on media longevity.

Capacity reclamation

File deletion in modern file systems is realized by updating file system metadata and marking the physical storage space that is used by the removed file as unused. The data of the removed file is not overwritten, which improves file system performance by reducing the number of I/O operations on physical storage that is required to perform file deletion.

However, this approach affects the management of the real capacity of volumes with enabled capacity savings. File system deletion frees space at the file system level, but physical data blocks that are allocated by the storage for the file still take up the real capacity of a volume.

To address this issue, file systems added support for the SCSI **UNMAP** command, which can be run after file deletion. It informs the storage system that physical blocks that are used by the removed file should be marked as no longer in use so that they can be freed. Modern OSs run SCSI **UNMAP** commands only to storage that advertises support for this feature.

Version 8.1.0 and later releases support the SCSI **UNMAP** command on IBM Spectrum Virtualize systems, which enables hosts to notify the storage controller of capacity that is no longer required and may be reused or deallocated, which might improve capacity savings.

Note: For volumes that are outside DRPs, the complete stack from the OS down to back-end storage controller must support UNMAP to enable the capacity reclamation. SCSI UNMAP is passed only to specific back-end storage controllers.

Consider the following points:

- ▶ Version 8.1.2 can also reclaim capacity in DRPs when a host runs SCSI **UNMAP** commands.
- ▶ By default, Version 8.2.1 does not advertise support for SCSI UNMAP to hosts.
- ▶ In Version 8.3.1, support for the host SCSI **UNMAP** command is enabled by default.

Before enabling SCSI UNMAP, see [SCSI Unmap support in IBM Spectrum Virtualize systems](#).

Analyze your storage stack to optimally balance the advantages and costs of data reclamation.

6.2.8 Encryption

IBM Spectrum Virtualize systems can be configured to enable data-at-rest encryption. This function is realized in hardware (self-encrypting drives or in SAS controller for drives that do not support self-encryption and are connected through the SAS bus) or in software (external virtualized storage).

For more information about creating and managing encrypted volumes, see Chapter 12, “Encryption” on page 735.

6.2.9 Cache mode

Another volume parameter is its cache characteristics. Under normal conditions, a volume’s read and write data is held in the cache of its preferred node with a mirrored copy of write data that is held in the partner node of the same I/O group. However, it is possible to create a volume with different cache characteristics if this configuration is required.

The cache setting of a volume can have the following values:

readwrite	All read and write I/O operations that are performed by the volume are stored in cache. This mode is the default cache mode for all volumes.
readonly	Read only I/O operations that are performed on the volume are stored in cache. Writes to the volume are not cached.
disabled	No I/O operations on the volume are stored in cache. I/Os are passed directly to the back-end storage controller rather than being held in the node's cache.

Having cache-disabled volumes makes it possible to use the native copy services in the underlying RAID array controller for MDisks (LUNs) that are used as IBM Spectrum Virtualize image mode volumes. However, using IBM Spectrum Virtualize Copy Services rather than the underlying disk controller copy services provides better results.

Note: Disabling the volume cache is a prerequisite for using native copy services on image mode volumes that are defined on storage systems that are virtualized by IBM Spectrum Virtualize. Contact IBM Support before turning off the cache for volumes in your production environment to avoid performance degradation.

6.2.10 I/O throttling

You can set a limit on the rate of I/O operations that are realized by a volume. This limitation is called *I/O throttling* or *governing*.

The limit can be set in terms of number of IOPS or bandwidth (megabytes per second (MBps), gigabytes per second (GBps), or terabytes per second (TBps)). By default, I/O throttling is disabled, but each volume can have up to two throttles that are defined: one for bandwidth and one for IOPS.

When deciding between using IOPS or bandwidth as the I/O governing throttle, consider the disk access profile of the application that is the primary volume user. Database applications generally issue large amounts of I/O operations, but transfer a relatively small amount of data. In this case, setting an I/O governing throttle that is based on bandwidth might not achieve much. A throttle that is based on IOPS is better suited for this use case.

Conversely, a video streaming or editing application issues a small amount of I/O but transfers large amounts of data. Therefore, it is better to use a bandwidth throttle for the volume in this case.

An I/O governing rate of 0 does not mean that zero IOPS or bandwidth can be achieved for this volume; rather, it means that no throttle is set for this volume.

Note: Consider the following points:

- ▶ I/O governing does not affect FlashCopy and data migration I/O rates.
- ▶ I/O governing on MM or GM secondary volumes does not affect the rate of data copy from the primary volume.

For more information about how to configure I/O throttle on a volume, see 6.5.4, “I/O throttling” on page 339.

6.2.11 Volume protection

Volume protection prevents volumes or host mappings from being deleted if the system detects recent I/O activity. This global setting is enabled by default on new systems. You can either set this value to apply to all volumes that are configured on your system or control whether the system-level volume protection is enabled or disabled on specific pools.

There are two levels at which the volume protection must be enabled to be effective: system level and pool level. Both levels must be enabled for protection to be active on a pool. The pool-level protection depends on the system-level setting to ensure that protection is applied consistently for volumes within that pool. If system-level protection is enabled, but pool-level protection is not enabled, any volumes in the pool can be deleted.

When you enable volume protection at the system level, you specify a period in minutes that the volume must be idle before it can be deleted. If volume protection is enabled and the period is not expired, the volume deletion fails even if the **-force** parameter is used. The following CLI commands and the corresponding GUI activities are affected by the volume protection setting:

- ▶ **rmvdisk**
- ▶ **rmvdiskcopy**
- ▶ **rmvvolume**
- ▶ **rmvdiskhostmap**
- ▶ **rmvolumehostclustermap**
- ▶ **rmmdiskgrp**
- ▶ **rmhostiogr**
- ▶ **rmhost**
- ▶ **rmhostcluster**
- ▶ **rmhostport**
- ▶ **mkrcrelationship**

Volume protection can be set from the GUI (new in V.8.3.1, see 6.5.5, “Volume protection” on page 344) and CLI (see 6.6.9, “Volume protection” on page 390).

6.2.12 Secure data deletion

The system provides methods to securely erase data from a drive or from a boot drive when a control enclosure is decommissioned.

Secure data deletion effectively erases or overwrites all traces of existing data from a data storage device. The original data on that device becomes inaccessible and cannot be reconstructed. You can securely delete data on individual drives and on a boot drive of a control enclosure. The methods and commands that are used to securely delete data enable the system to be used in compliance with European Regulation EU2019/424.

The procedure is described in [IBM Documentation](#).

6.3 Virtual volumes

IBM Spectrum Virtualize V7.6 introduced support for *virtual volumes*. These volumes enable support for VVOLs, which allow VMware vCenter to manage system objects, such as volumes and pools. The IBM Spectrum Virtualize system administrators can create volume objects of this class, and assign ownership to VMware administrators to simplify management.

For more information about configuring VVOLs with IBM Spectrum Virtualize, see *Configuring VMware Virtual Volumes for Systems Powered by IBM Spectrum Virtualize*, SG24-8328.

6.4 Volumes in multi-site topologies

IBM Spectrum Virtualize can be set up in a multi-site configuration, which makes the system aware of which system components (I/O groups, nodes, and back-end storage) are at which site. For the storage system topology description, a site is defined as an independent failure domain, which means that if one site fails, the other site can continue to operate without disruption.

Depending on the type and scale of the failure that the solution must survive, the sites can be two places in the same data center room (one end of the IBM Spectrum system) or buildings in different cities on different tectonic plates and powered from independent grids (the other end of the IBM Spectrum system).

The following storage system topologies are available:

- ▶ Standard topology, which is intended for single-site configurations that do not allow site definition and assume that all components of the solution are at a single site. You can use GM or MM to maintain a copy of a volume on a different system at a remote site, which can be used for DR.
- ▶ IBM HyperSwap topology, which is a three-site high availability (HA) configuration in which each I/O group is at a different site. A volume can be active on two I/O groups so that if one site is not available, it can immediately be accessed through the other site.

Note: Multi-site topologies of IBM Spectrum Virtualize use two sites as storage component locations (nodes and back-end storage). The third site is used as a location for a tie-breaker component that prevents split-brain scenarios if the storage system components lose communication with each other.

The **Create Volumes** menu provides the following options, depending on the configured system topology:

- ▶ With standard topology, the available options are Basic, Mirrored, and Custom.
- ▶ With HyperSwap topology, the options are Basic, HyperSwap, and Custom.

The HyperSwap function provides HA volumes that are accessible through two sites up to 300 km (186.4 miles) apart. A fully independent copy of the data is maintained at each site.

Note: The determining factor for HyperSwap configuration validity is the time that it takes to send the data between the sites. Therefore, while estimating the distance, consider the fact that the distance between the sites that is measured along the data path is longer than the geographic distance. Additionally, each device on the data path that adds latency increases the effective distance between the sites.

When data is written by hosts at either site, both copies are synchronously updated before the write operation completion is reported to the host. The HyperSwap function automatically optimizes itself to minimize data that is transmitted between sites and to minimize host read and write latency.

If the nodes or storage at either site go offline, the HyperSwap function automatically fails over access to the other copy. The HyperSwap function also automatically resynchronizes the two copies when possible.

The HyperSwap function is built on a foundation of two earlier technologies: The Non-disruptive Volume Move (NDVM) function that was introduced in IBM Spectrum Virtualize V6.4, and the RC features that include MM, GM, and GMCV.

The HyperSwap volume configuration is possible only after the IBM Spectrum Virtualize system is configured in the HyperSwap topology. After this topology change, the GUI presents an option to create a HyperSwap volume and creates them by running the `mkvolume` command instead of the `mkvdisk` command. The GUI continues to use the `mkvdisk` command when all other classes of volumes are created.

Note: It is still possible to create HyperSwap volumes in Version 7.5.

For more information, see *IBM Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317.

For more information about HyperSwap topology, see [IBM Documentation](#).

From the perspective of a host or a storage administrator, a HyperSwap volume is a single entity, but it is realized by using four volumes, a set of FlashCopy maps, and an RC relationship (see Figure 6-9).

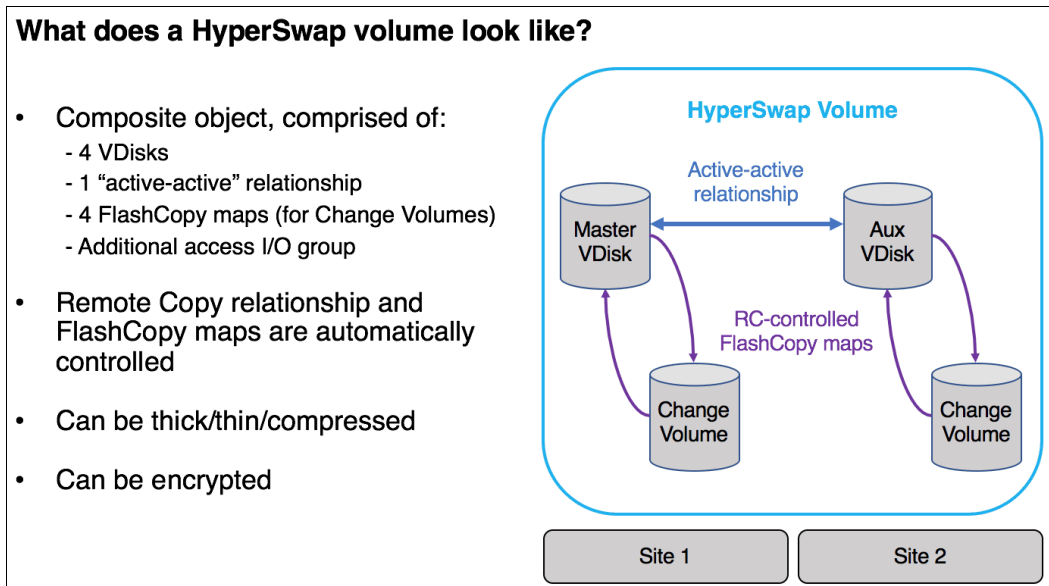


Figure 6-9 What makes up a HyperSwap volume

The GUI simplifies the HyperSwap volume creation process by asking about required volume parameters only and automatically configuring all the underlying volumes, FlashCopy maps, and volume replications relationships.

An example of a HyperSwap volume configuration is shown in Figure 6-29 on page 337.

6.5 Operations on volumes

This section describes how to perform operations on volumes by using the GUI. The following operations can be performed on a volume:

- ▶ Volumes can be created and deleted.
- ▶ Volumes can have their characteristics modified, including:
 - Size (expanding or shrinking)
 - Number of copies (adding or removing a copy)
 - I/O throttling
 - Protection
- ▶ Volumes can be migrated at run time to another MDisk or storage pool.
- ▶ A PiT volume snapshot can be created by using FlashCopy. Multiple snapshots and quick restore from snapshots (reverse FlashCopy) are supported.
- ▶ Volumes can be mapped to (and unmapped from) hosts.

Note: With Version 7.4 and later, it is possible to prevent accidental deletion of volumes if they recently performed any I/O operations. This feature is called *volume protection*, and it prevents active volumes or host mappings from being deleted inadvertently. This process is done by using a global system setting. For more information, see 6.6.9, “Volume protection” on page 390 and the “Changing volume protection settings” topic in [IBM Documentation](#).

6.5.1 Creating volumes

This section focuses on using the **Create Volumes** menu to create Basic and Mirrored volumes in a system with a standard topology. Volume creation is available with the following volume classes:

- ▶ Basic
- ▶ Mirrored
- ▶ Custom

To create a volume, complete the following steps:

1. Click the **Volumes** menu and click the **Volumes** option of the IBM Spectrum Virtualize GUI, as shown in Figure 6-10.

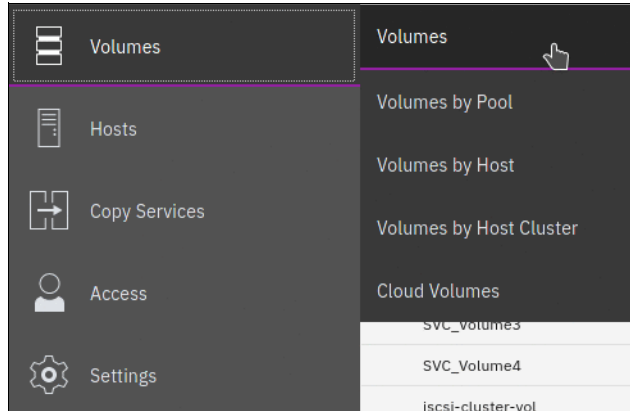


Figure 6-10 Volumes menu

A list of volumes, their state, capacity, and associated storage pools are displayed.

2. To create a volume, click **Create Volumes**, as shown in Figure 6-11.

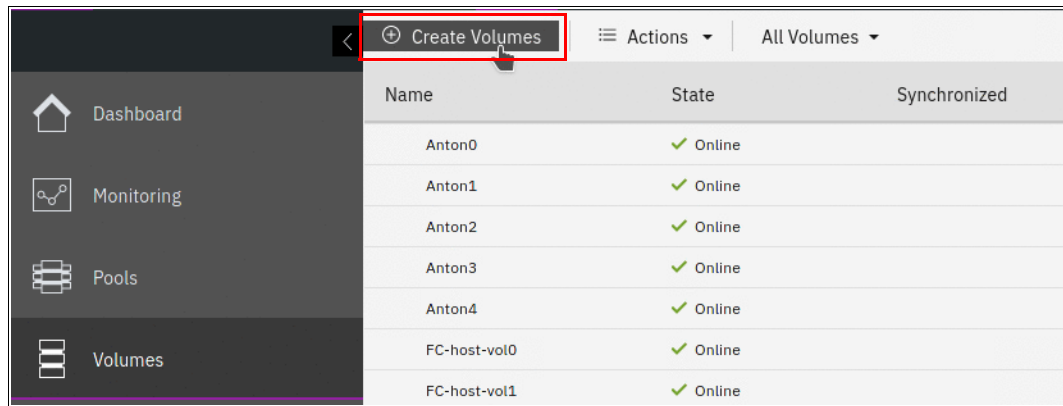


Figure 6-11 Create Volumes

The Create Volumes tab opens the Create Volumes window, which shows the available creation methods.

Note: The volume classes that are displayed in the Create Volumes window depend on the topology of the system.

The Create Volumes window for standard topology is shown in Figure 6-12.

The screenshot shows the 'Create Volumes' window with three tabs: 'Basic', 'Mirrored', and 'Custom'. The 'Basic' tab is selected and highlighted with a red box. Below the tabs, there is a description: 'Create a preset volume with all the basic features.' The 'Pool' field has a dropdown menu with 'Click to select.' and a small arrow. The 'Capacity Details' section shows a progress bar and 'Total 0 bytes'. The 'Volume Details' section has three columns: 'Quantity' with a spinner set to '1', 'Capacity' with a dropdown set to 'GiB' and a red error icon, and 'Name' with an empty text field. Below this is the 'Capacity savings' section with a dropdown set to 'None' and a checkbox for 'Deduplicated'. A blue link '+ Define another volume' is present. The 'I/O group' dropdown is set to 'Automatic'. At the bottom, there is a 'Summary' section with three stacked boxes and the text 'Fields Incomplete'. The footer contains a 'Need Help' link, a 'Cancel' button, and 'Create and Map' and 'Create' buttons.

Figure 6-12 Basic, mirrored, and custom volume creation options

Note: Consider the following points:

- ▶ A *basic volume* is a volume that has only one physical copy, uses storage that is allocated from a single pool on one site, and uses read/write cache mode.
- ▶ A *mirrored volume* is a volume with two physical copies, where each volume copy can belong to a different storage pool.
- ▶ A *custom volume* (in the context of this menu) is a basic or mirrored volume with values of some of its parameters that are changed from the defaults.
- ▶ The Create Volumes window also provides (by using the Capacity Savings parameter) the ability to change the default provisioning of a basic or mirrored volume to thin-provisioned or compressed. For more information, see “Capacity savings option” on page 329.

Creating basic volumes

A *basic volume* is a volume that has only one physical copy. Basic volumes are supported in any system topology and are common to all configurations. Basic volumes can be of any type of virtualization: striped, sequential, or image. They can also use any type of capacity savings: thin-provisioning, compressed, or none. Deduplication can be configured with thin-provisioned and compressed volumes in DRPs for added capacity savings.

To create a basic volume, click **Basic**, as shown in Figure 6-13 on page 325. This action opens the **Basic volume** menu, where you can define the following parameters:

- ▶ Pool: The pool in which the volume is created (drop-down menu).
- ▶ Quantity: Number of volumes to be created (numeric up or down).
- ▶ Capacity: Size of the volume in specified units (drop-down menu).
- ▶ Capacity Savings (drop-down menu):
 - None
 - Thin-provisioned
 - Compressed
- ▶ Name: Name of the volume (cannot start with a number).
- ▶ I/O group.

The Basic Volume creation window is shown in Figure 6-13 on page 325.

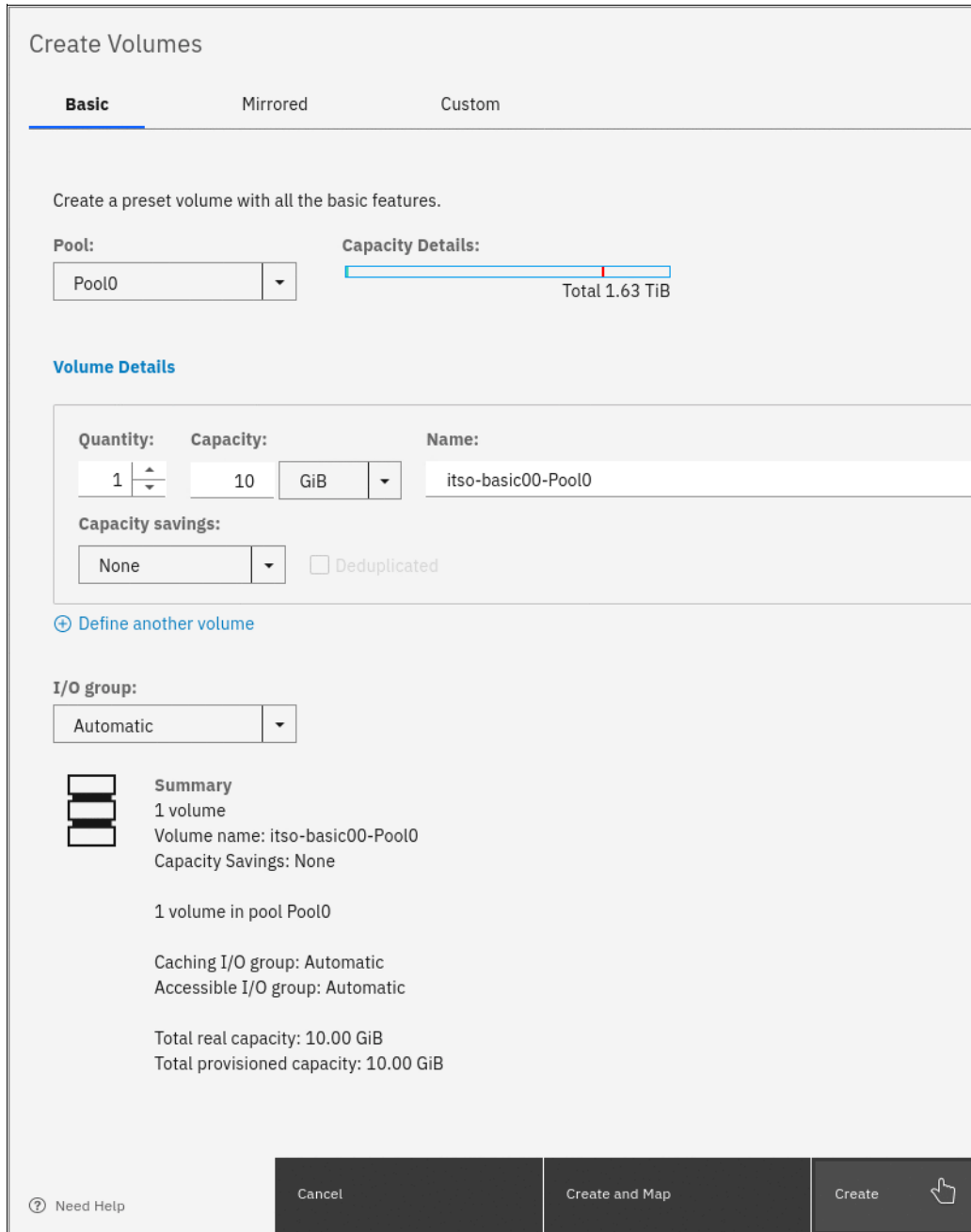


Figure 6-13 Create Volumes window

Define and consistently use a suitable volume naming convention to facilitate easy identification. For example, a volume name can contain the name of the pool or some tag that identifies the underlying storage subsystem, the host or cluster name that the volume is mapped to, and the content of this volume, such as the name of the applications that use the volume.

When all of the characteristics of the basic volume are defined, it can be created by selecting one of the following options:

- ▶ **Create**
- ▶ **Create and Map**

Note: The plus sign (+) icon can be used to create more volumes in the same instance of the volume creation wizard.

In the example, the **Create** option was selected. The volume-to-host mapping can be performed later, as described in 6.5.8, “Mapping a volume to a host” on page 358.

When the operation completes, the volume is seen in the Volumes window in the state “Online (formatting)”, as shown in Figure 6-14.

Name	State	Synchronized	Pool
iscsi-host-vol1	✓ Online		SVC Pool
iscsi-host-vol2	✓ Online		SVC Pool
iscsi-host-vol3	✓ Online		SVC Pool
itso-basic00-Pool0	✓ Online (formatting)		Pool0

Figure 6-14 Basic volume formatting

By default, the GUI does not show any details about the commands it runs to complete a task. However, while a command runs you can click **View more details** to see the underlying CLI commands that are run to create the volume and a report of completion of the operation, as shown in Figure 6-15.

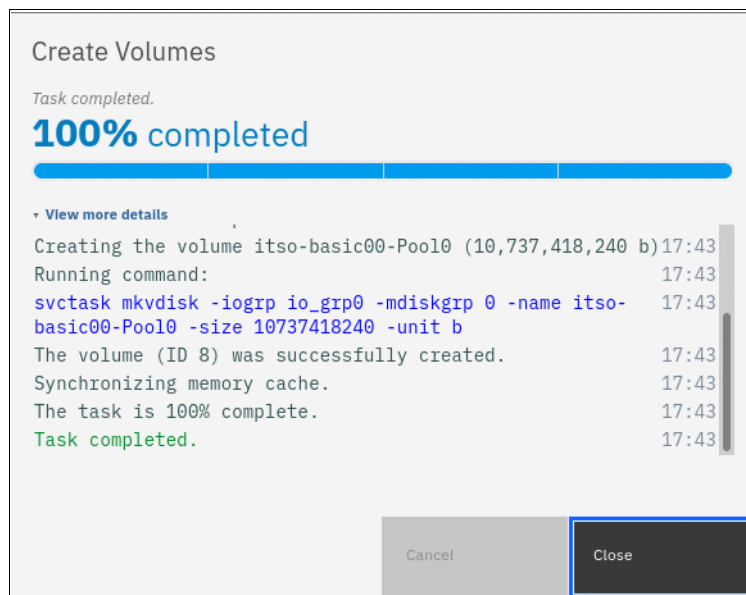


Figure 6-15 Viewing details of volume creation operation

Note: Consider the following points:

- ▶ Standard-provisioned volumes are automatically formatted through the quick initialization process after the volume is created. This process makes standard-provisioned volumes available for use immediately.
- ▶ Quick initialization requires a small amount of I/O to complete, and limits the number of volumes that can be initialized at the same time. Some volume actions, such as moving, expanding, shrinking, or adding a volume copy, are disabled when the specified volume is initializing. Those actions become available after the initialization process completes.
- ▶ The quick initialization process can be disabled in circumstances where it is not necessary. For example, if the volume is the target of a Copy Services function, the Copy Services operation formats the volume. The quick initialization process can also be disabled for performance testing so that the measurements of the raw system capabilities can take place without waiting for the process to complete.

For more information, see [IBM FlashSystem 9200 documentation](#) and expand **Product overview** → **Technical overview** → **Volumes** → **Standard-provisioned volumes**.

Creating mirrored volumes

To create a mirrored volume, complete the following steps:

1. In the Create Volumes window, click **Mirrored**, and choose the **Pool** for **Copy1** and **Copy2** by using the drop-down menus. Although the mirrored volume can be created in the same pool, this setup is not typical. Generally, keep volumes copies on separate set of physical disks (pools).
2. Enter the following Volume Details:
 - Quantity
 - Capacity
 - Capacity savings
 - Name

Leave the I/O group option at its default setting of **Automatic** (see Figure 6-16).

Figure 6-16 Mirrored volume creation

3. Click **Create** (or **Create and Map**).

When the operation completes, the volume is seen in the Volumes window in the state “Online (formatting)”, as shown in Figure 6-17.

Name	State	Synchronized	Pool
itso-basic00-Pool1	✓ Online (formatting)		Pool1
∨ itso-mirrored00-Pool0-Po...	✓ Online (formatting)		Pool0
Copy 0*	✓ Online (formatting)	Yes	Pool0
Copy 1	✓ Online (formatting)	Yes	Pool1

Figure 6-17 Mirrored volume formatting

A mirrored volume is displayed in the GUI as configured in the pool in which it has its primary. In this example, volume its0-mirrored00-Pool0-Pool1 is displayed as configured in Pool0 because it has its primary copy in Pool0.

Note: When creating a mirrored volume by using this menu, you are not required to specify the Mirrored Sync rate (it defaults to 2 MBps). The synchronization rate can be customized by using the **Custom** menu.

Capacity savings option

When the basic or mirrored method of volume creation is used, the GUI provides a Capacity Savings option, which enables altering the volume provisioning parameters without using the Custom volume provisioning method. You can select **Thin-provisioned** or **Compressed** from the drop-down menu, as shown in Figure 6-18, to create thin-provisioned or compressed volumes.

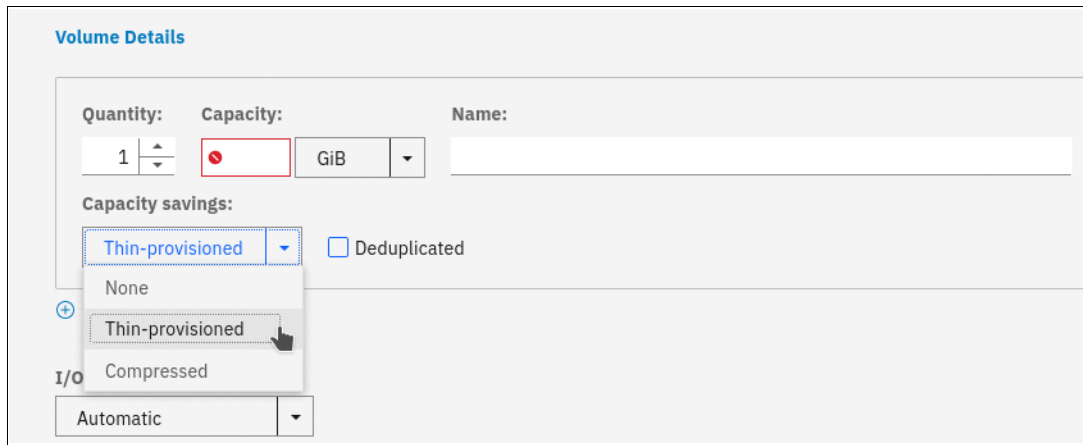


Figure 6-18 Volume creation with capacity saving option

Note: Consider the compression guidelines in Chapter 9, “Advanced features for storage efficiency” on page 509 before creating the first compressed volume copy on a system.

When a thin-provisioned or compressed volume is defined in a DRP, the Deduplicated option becomes available. Select this option to enable deduplication of the volume.

Thin-provisioned and compressed volumes have a special icon in the Capacity column of the **Volumes** menu that makes it easy to distinguish them, as shown in Figure 6-19.



Name	State	Synchronized	Capacity
itso-basic00-Pool0	✓ Online (formatting)		10.00 GiB
itso-basic00-Pool1	✓ Online (formatting)		10.00 GiB
> itso-mirrored00-Pool0-Po...	✓ Online (formatting)		10.00 GiB
itso-thin00-Pool0	✓ Online		10.00 GiB 
itso-thin01-Pool1	✓ Online		10.00 GiB 
Showing 24 volumes Selecting 1 volume (10.00 GiB)			

Figure 6-19 Space-efficient volumes icons

Volume iso-thin00-Pool0 is thin-provisioned. Volume iso-thin01-Pool1 is compressed. There is no icon indicating whether a volume is deduplicated or not.

6.5.2 Creating custom volumes

The Create Volumes window also enables Custom volume creation that expands the set of options for volume creation that are available to the administrator.

The **Custom** menu consists of several windows:

- ▶ Volume Location: Mandatory. It defines the number of volume copies, pools to be used, and I/O group preferences.
- ▶ Volume Details: Mandatory. It defines the Capacity savings option.
- ▶ Thin Provisioning: Enables configuration of thin-provisioning settings if this capacity saving option is selected.
- ▶ Compressed: Enables configuration of compression settings if this capacity saving option is selected.
- ▶ General: Configures cache mode and formatting.
- ▶ Summary.

Use these windows to customize your Custom volume as wanted, and then commit these changes by clicking **Create**.

You can mix and match settings on different windows to achieve the final volume configuration that meets your requirements.

Volume Location window

The Volume Location window is shown in Figure 6-20 on page 331.

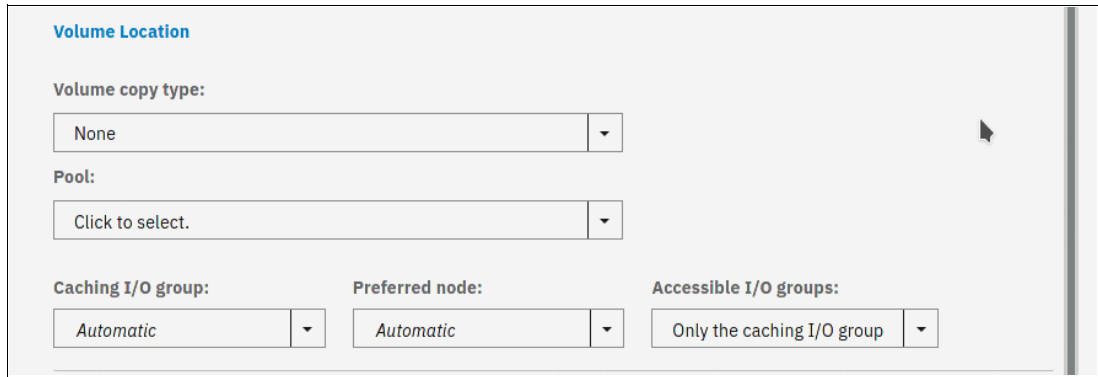


Figure 6-20 Volume Location window

This window has the following options:

- ▶ Volume copy type: You can choose between **None** (single volume copy) and **Mirrored** (two volume copies).
- ▶ Pool: Specifies storage pool to use for each of volume copies.
- ▶ Mirror sync rate: You can set the mirror sync rate for the volume copies. This option is displayed only for the Mirrored volume copy type, and you can set the volume copy synchronization rate to a value 128 KiBps - 64 MiBps.
- ▶ Caching I/O group: You can choose between **Automatic** (allocated by the system) and manually specifying the I/O group.
- ▶ Preferred node: You can choose between **Automatic** (allocated by the system) and manually specifying the preferred node for the volume.
- ▶ Accessible I/O groups: You can choose between **Only the caching I/O group** and **All**.

Volume Details window

The Volume Details window is shown in Figure 6-21.



Figure 6-21 Volume Details window

This window has the following options:

- ▶ Quantity: You can specify how many volumes to create.
- ▶ Capacity: Capacity of the volume.
- ▶ Name: You can define the volume name.

- ▶ Capacity savings: You can choose between **None** (standard-provisioned volume), **Thin-provisioned**, and **Compressed**.
- ▶ Deduplicated: Thin-provisioned and compressed volumes that are created in a DRP can be deduplicated.

If you click **Define another volume**, the GUI displays a subpane in which you can define the configuration of another volume, as shown in Figure 6-22.

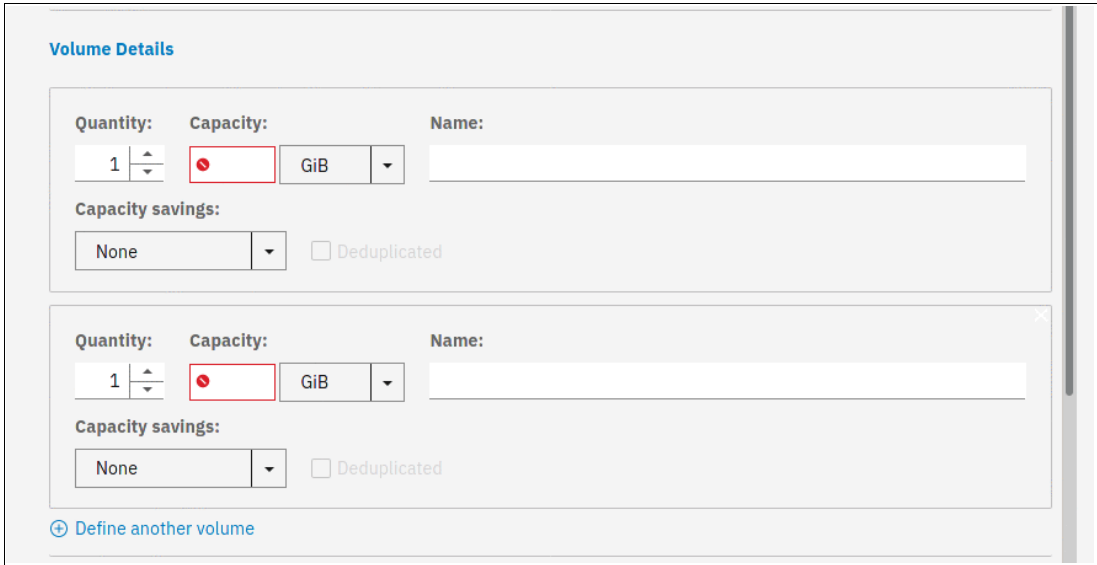


Figure 6-22 Volume Details window with two volume subpanes

This way, you can create volumes with different characteristics in a single invocation of the volume creation wizard.

Thin Provisioning window

If you choose to create a thin-provisioned volume, a Thin Provisioning window is displayed, as shown in Figure 6-23.

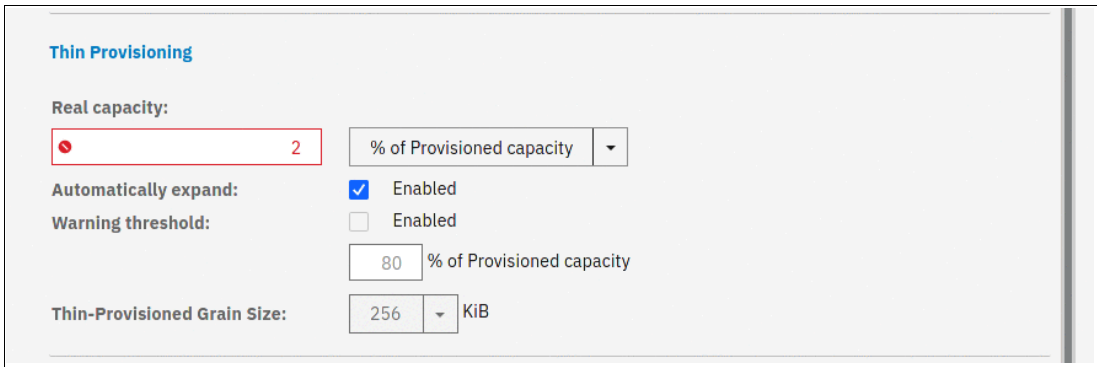


Figure 6-23 Thin Provisioning window

This window includes the following options:

- ▶ Real capacity: Real capacity of the volume, which is specified as a percentage of the virtual capacity or in bytes.
- ▶ Automatically expand: Whether to automatically expand the real capacity of the volume if needed. Defaults to Enabled.

- ▶ **Warning threshold:** Whether a warning message is sent and at what percentage of filled virtual capacity. Defaults to Enabled, with a warning threshold set at 80%.
- ▶ **Thin-Provisioned Grain Size:** You can define the grain size for the thin-provisioned volume. Defaults to 256 KiB.

Important: If you do not use the **autoexpand** feature, the volume goes offline if it receives a write request after all real capacity is allocated.

The default grain size is 256 KiB. The optimum choice of grain size depends on the volume use type. Consider the following points:

- ▶ If you are *not* going to use the thin-provisioned volume as a FlashCopy source or target volume, use 256 KiB to maximize performance.
- ▶ If you *are* going to use the thin-provisioned volume as a FlashCopy source or target volume, specify the same grain size for the volume and for the FlashCopy function.
- ▶ If you plan to use Easy Tier with thin-provisioned volumes, see the IBM Support article [Performance Problem When Using Easy Tier With Thin Provisioned Volumes](#).

Compressed window

If you choose to create a compressed volume, a Compressed window is displayed, as shown in Figure 6-24.

The screenshot shows a configuration window titled "Compressed". It contains the following settings:

- Real capacity:** A text input field containing the number "2" and a dropdown menu set to "% of Provisioned capacity".
- Automatically expand:** A checkbox that is checked, with the label "Enabled".
- Warning threshold:** A checkbox that is unchecked, with the label "Enabled". Below it is a text input field containing "80" and a dropdown menu set to "% of Provisioned capacity".

Figure 6-24 Compressed window

This window gives the following options:

- ▶ **Real capacity:** Real capacity of the volume, which is specified as percentage of the virtual capacity or in bytes.
- ▶ **Automatically expand:** Whether to automatically expand the real capacity of the volume if needed. Defaults to Enabled.
- ▶ **Warning threshold:** Whether a warning message is sent, and at what percentage of filled virtual capacity. Defaults to Enabled, with a warning threshold set at 80%.

You cannot specify the grain size for a compressed volume.

Note: Consider the compression guidelines in Chapter 9, “Advanced features for storage efficiency” on page 509 before creating the first compressed volume copy on a system.

General window

The General window is shown in Figure 6-25.



Figure 6-25 General window

This window provides the following options:

- ▶ Cache mode: Controls volume caching. Defaults to Enabled. Other available options are Read-only and Disabled.
- ▶ OpenVMS unit device identifier (UDID): Each OpenVMS Fibre Channel (FC)-attached volume requires a user-defined identifier or UDID. A *UDID* is a nonnegative integer that is used in the creation of the OpenVMS device name.

6.5.3 HyperSwap volumes

To create a HyperSwap volume, complete the following steps:

1. In the IBM Spectrum Virtualize GUI, select **Volumes** → **Volumes**, as shown in Figure 6-26.

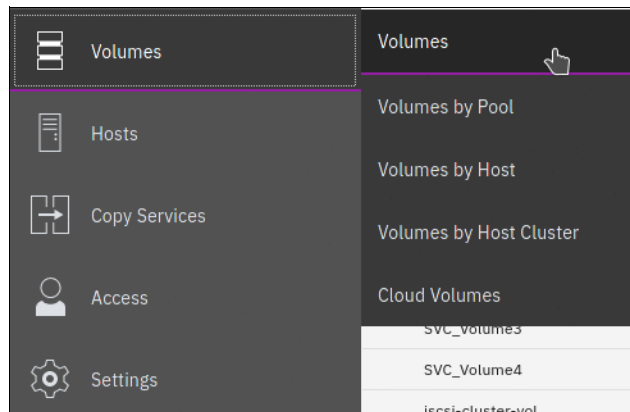


Figure 6-26 Volumes menu

A list of volumes, their state, capacity, and associated storage pools, is displayed.

2. Click **Create Volumes**, as shown in Figure 6-27 on page 335.

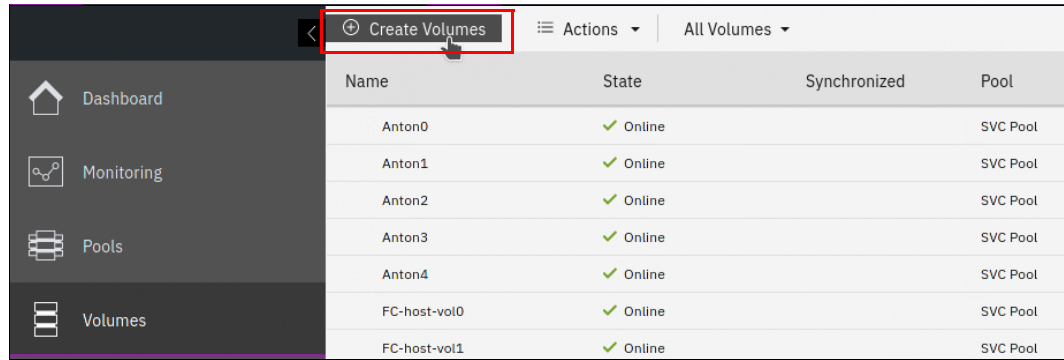


Figure 6-27 Create Volumes button

The Create Volumes tab opens the Create Volumes window, which displays available creation methods.

Note: The volume classes that are displayed in the Create Volumes window depend on the topology of the system.

The Create Volumes window for the HyperSwap topology is shown in Figure 6-28.

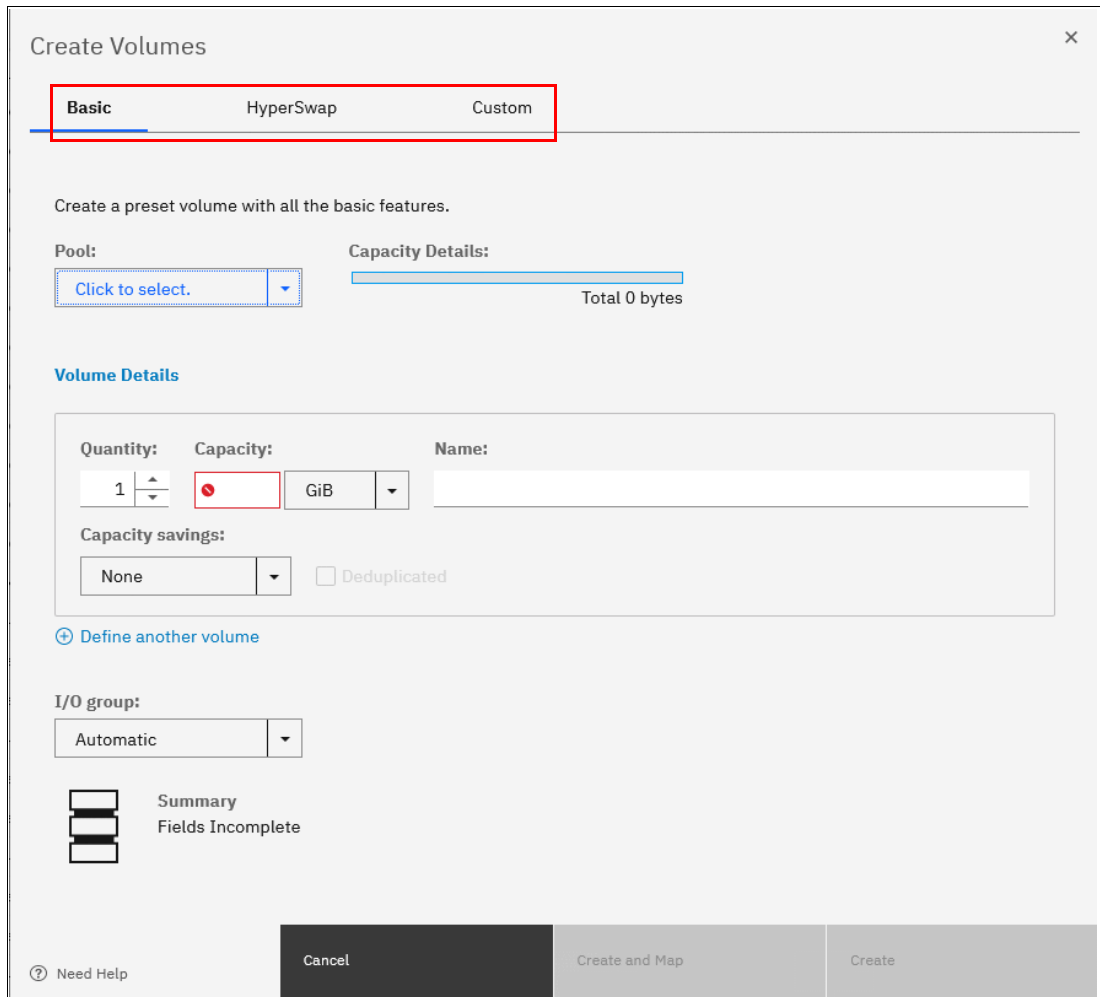


Figure 6-28 Basic, HyperSwap, and Custom volume creation options

Note: Consider the following points:

- ▶ A *basic volume* is a volume that has only one physical copy, uses storage that is allocated from a single pool on one site, and uses read/write cache mode.
- ▶ A *mirrored volume* is a volume with two physical copies, where each volume copy can belong to a different storage pool.
- ▶ A *HyperSwap volume* is a volume with two physical copies, where each volume copy is configured on storage in a different location.
- ▶ A *custom volume* (in the context of this menu) is a basic or mirrored volume with the values of some of its parameters changed from the defaults.
- ▶ The Create Volumes window also provides (by using the Capacity Savings parameter) the ability to change the default provisioning of a basic or mirrored volume to thin-provisioned or compressed. For more information, see “Capacity savings option” on page 329.

The notable difference between HyperSwap volume and basic volume creation is that HyperSwap volume creation includes specifying storage pool names at each site. The system uses its topology awareness to map storage pools to sites, which ensure that the data is correctly mirrored across locations.

As shown in Figure 6-29, a single volume is created with volume copies in sites site1 and site2. This volume is in an active-active (MM) relationship with extra resilience that is provided by two change volumes.

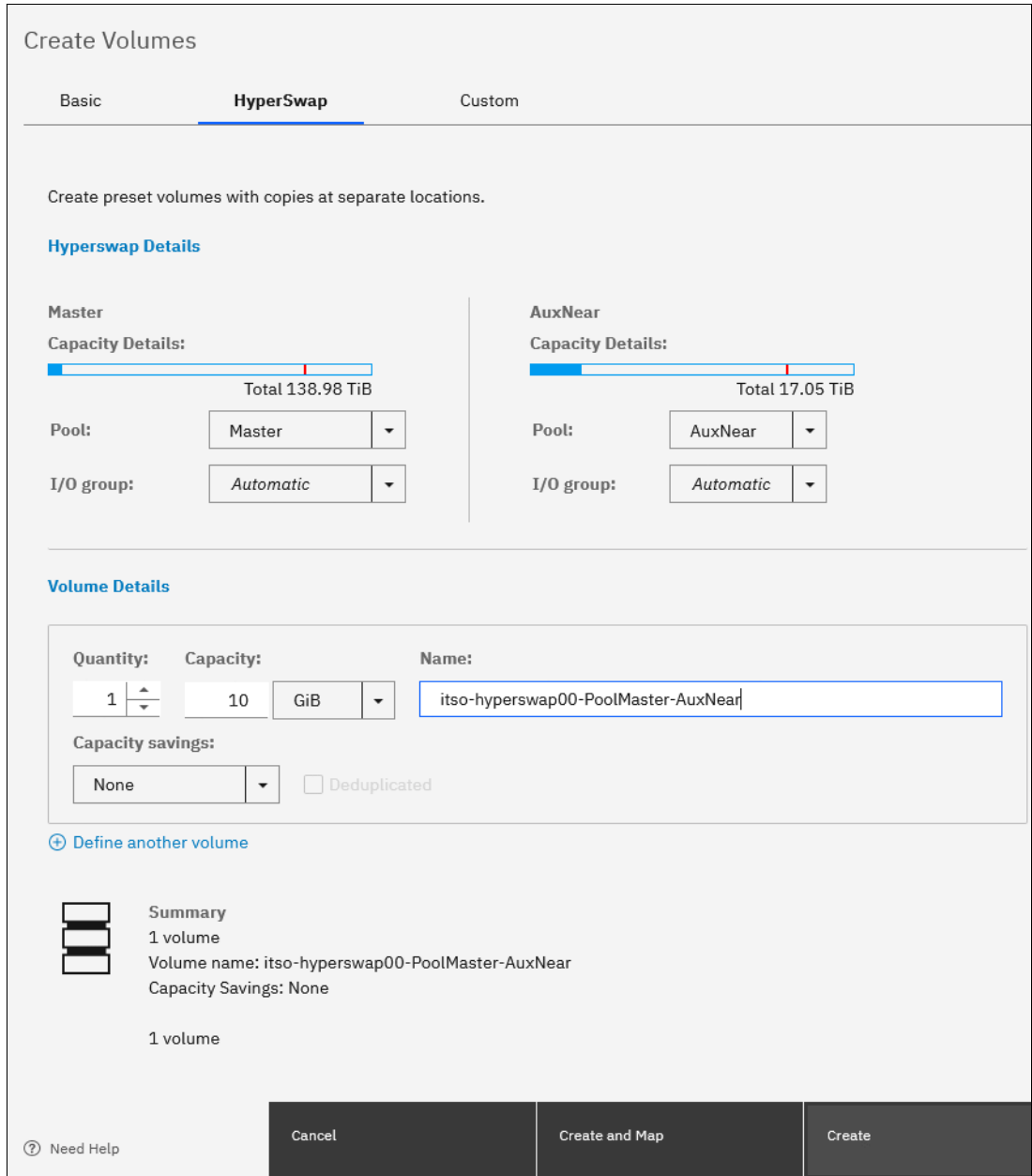


Figure 6-29 HyperSwap volume creation window

After the volume is created, it is visible in the volumes list, as shown in Figure 6-30.

Name	State	Synchronized	Pool
itso-hyperswap00-PoolMaster-AuxNear	✓ Online (formatting)		Multiple
itso-hyperswap00-PoolMaster-AuxNear (Master)	✓ Online (formatting)	Yes	Master
itso-hyperswap00-PoolMaster-AuxNear (AuxNear)	✓ Online (formatting)	Yes	AuxNear

Showing 25 volumes / Selecting 0 volumes

Latency 0 ms Read 0 ms Write 0 ms Bandwidth 0 MBps Read 0 MBps Write 0 MBps

Figure 6-30 A HyperSwap volume in the list of volumes

The Pool column shows the value “Multiple”, which indicates that a volume is a HyperSwap volume. A volume copy at each site is visible, and the change volumes that are used by the technology are not displayed in this GUI view.

Note: For volumes in multi-site topologies, the asterisk (*) does not indicate the primary copy, but the local volume copy that is used for data reads.

A single `mkvolume` command can create a HyperSwap volume. Up to IBM Spectrum Virtualize V7.5, this process required careful planning and running the following sequence of commands:

1. `mkvdisk master_vdisk`
2. `mkvdisk aux_vdisk`
3. `mkvdisk master_change_volume`
4. `mkvdisk aux_change_volume`
5. `mkrcrelationship -activeactive`
6. `chrcrelationship -masterchange`
7. `chrcrelationship -auxchange`
8. `addvdiskaccess`

Note: IBM Spectrum Virtualize Version 8.4 extends HyperSwap support to hosts that are attached through NVMe over Fabrics (NVMe-oF) through FC. The standard protocol mechanism Asymmetric Namespace Access (ANA), which is analogous to SCSI Asymmetric Logical Unit Access (ALUA), is used to provide this function to hosts that are attached through NVMe-oF.

6.5.4 I/O throttling

This section describes how to use I/O throttling on a volume.

Defining a volume throttle

To set a volume throttle, complete the following steps:

1. Select **Volumes** → **Volumes**, and then select the volume that you want to throttle. Select **Actions** → **Edit Throttle**, as shown in Figure 6-31.

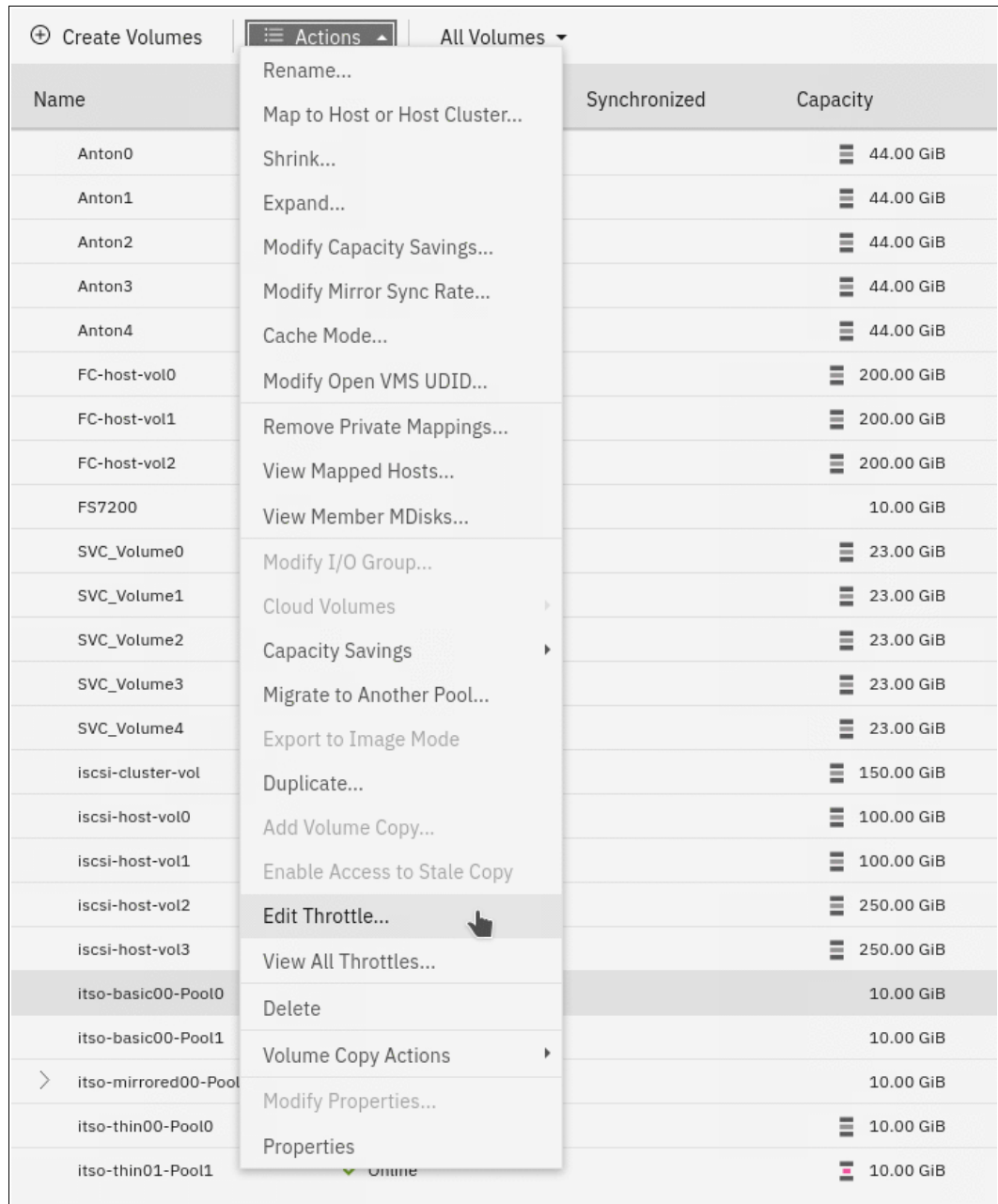


Figure 6-31 Edit Throttle menu item

2. In the Edit Throttle window, define the throttle in terms of number of IOPS or bandwidth. In our example, we set an IOPS throttle of 10,000, as shown in Figure 6-32. Click **Create**.

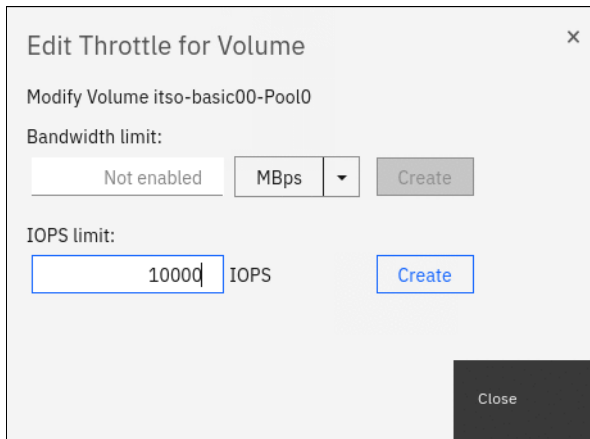


Figure 6-32 IOPS throttle on a volume

3. After the Edit Throttle task completes successfully, the Edit Throttle window opens again. You can now set the throttle based on the different metrics, modify the throttle, or close the window without performing further actions by clicking **Close**.

Listing volume throttles

To view volume throttles, select **Volumes** → **Volumes**, and then select **Actions** → **View All Throttles**, as shown in Figure 6-33.

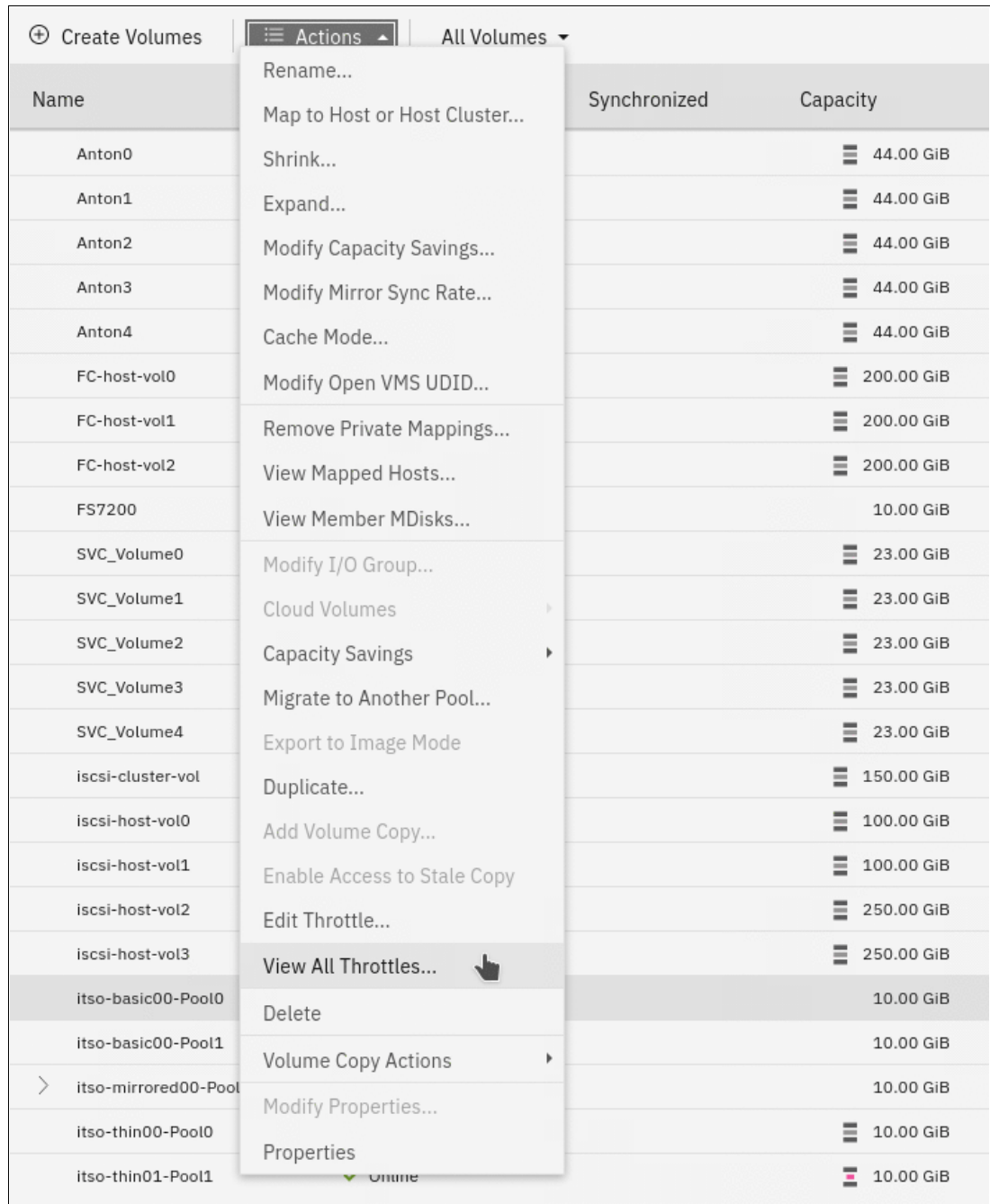


Figure 6-33 View All Throttles menu item

The **View All Throttles** menu shows all volume throttles that are defined in the system, as shown in Figure 6-34.

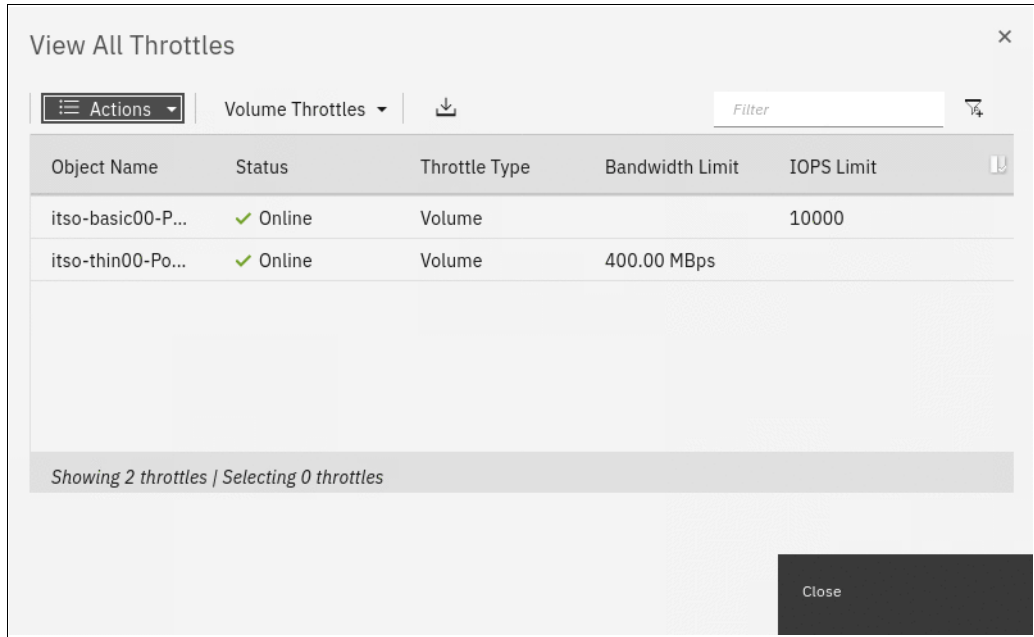


Figure 6-34 View All Throttles window

You can view other throttles by selecting a different throttle type in the drop-down menu, as shown in Figure 6-35.

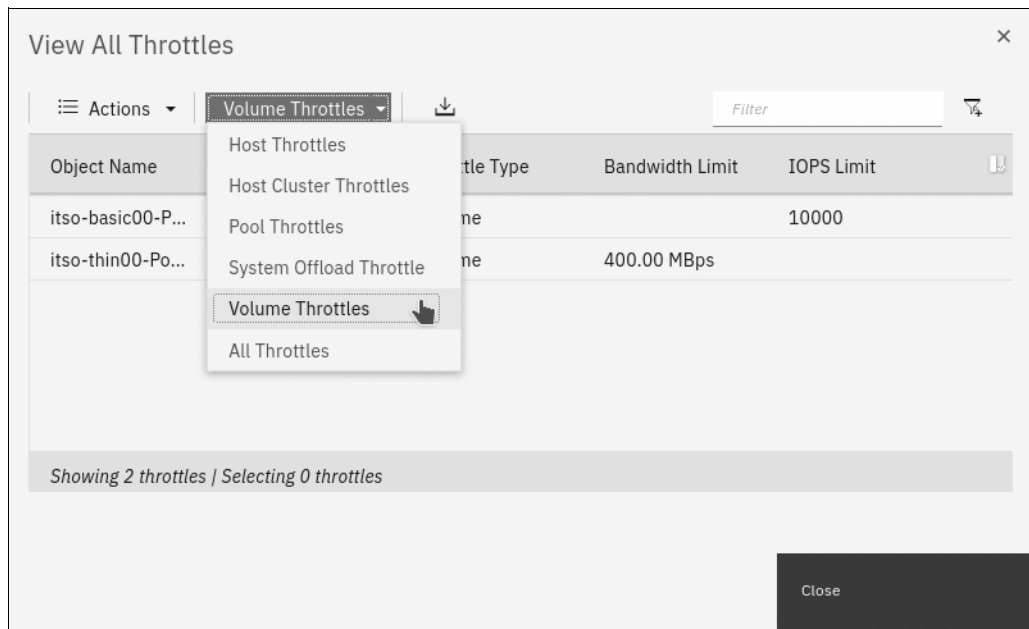


Figure 6-35 Filtering the throttle type

Modifying or removing a volume throttle

To remove a volume throttle, complete the following steps:

1. From the **Volumes** menu, select the volume that is attached the throttle that you want to remove. Select **Actions** → **Edit Throttle**, as shown in Figure 6-36.

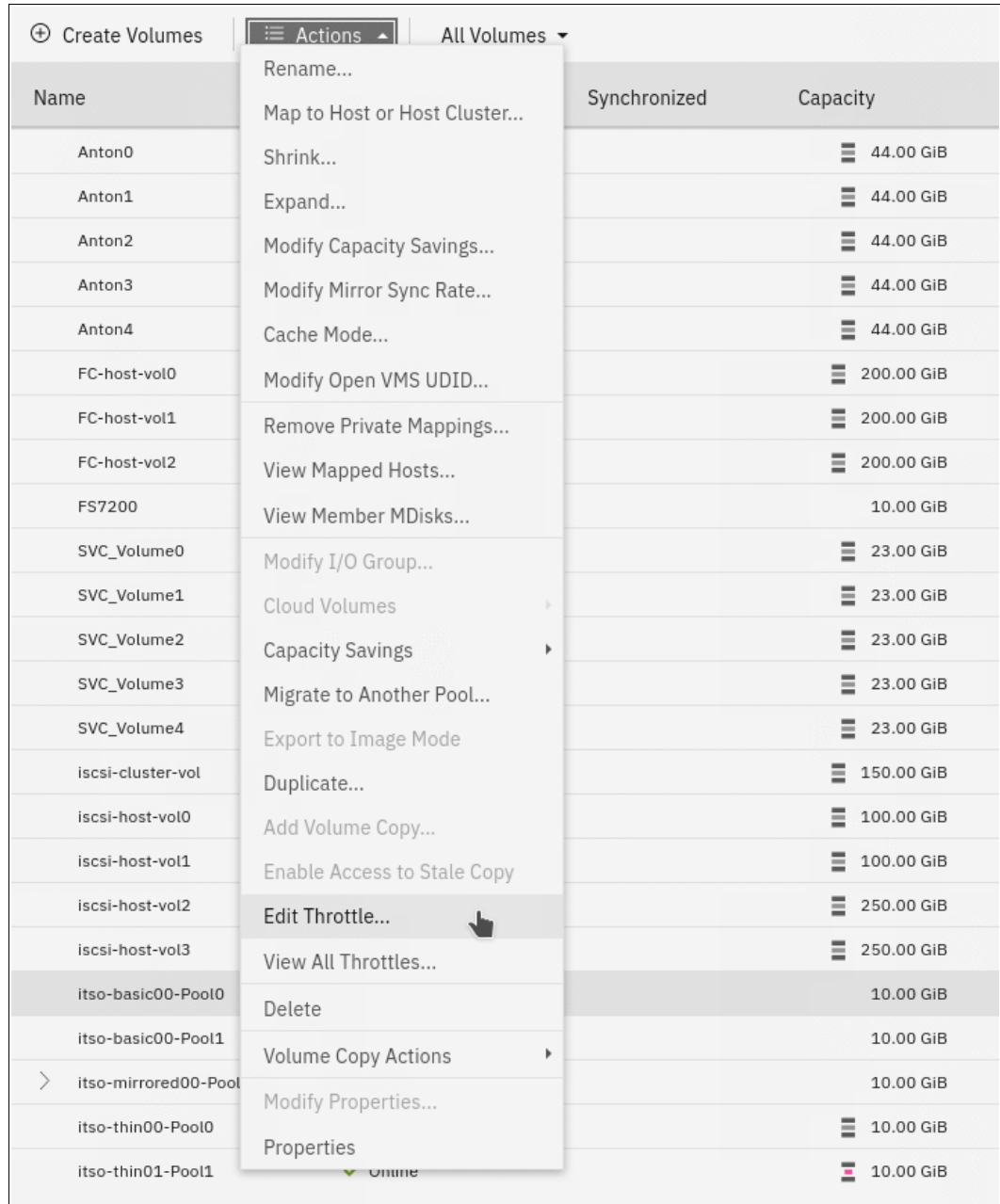


Figure 6-36 Edit Throttle menu

- To modify the volume throttle, enter new throttling parameters and click **Save**, as shown in Figure 6-37. In this example, the I/O throttling limit is increased to 15,000 IOPS.

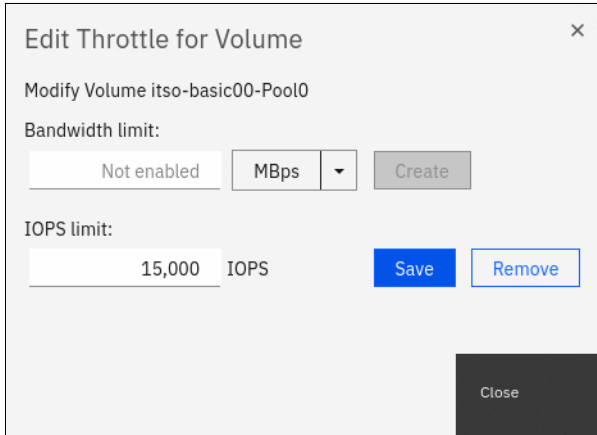


Figure 6-37 Modifying a volume throttle

- To remove the throttle completely, click **Remove** for the throttle that you want to remove, as shown in Figure 6-38.

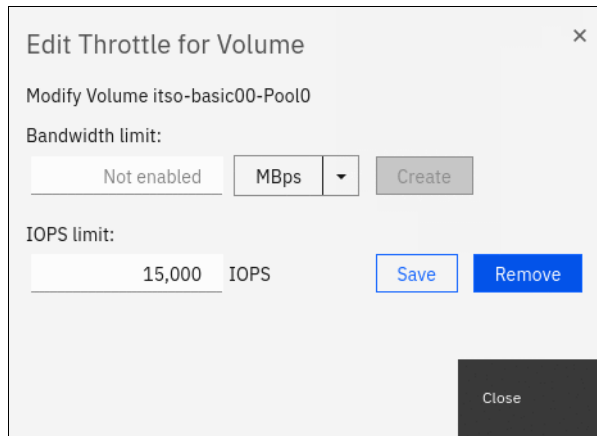


Figure 6-38 Removing a volume throttle

After the Edit Throttle task completes successfully, the Edit Throttle window opens again. You can now set the throttle based on the different metrics, modify the throttle, or close the window without performing any action by clicking **Close**.

6.5.5 Volume protection

To configure volume protection, select **Settings** → **System** → **Volume Protection**, as shown in Figure 6-39 on page 345.

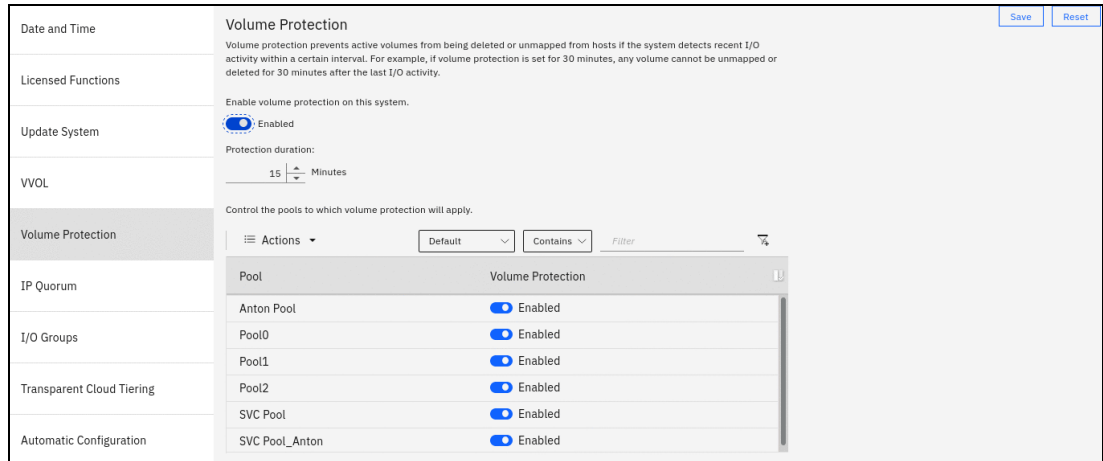


Figure 6-39 Volume Protection configuration

In this view, you can configure system-wide volume protection (enabled by default), set the minimum inactivity period that is required to allow volume deletion (protection duration), and configure volume protection for each configured pool (enabled by default). In the example, volume protection is enabled with the 15-minute minimum inactivity period and is turned on for all configured pools.

6.5.6 Modifying a volume

After a volume is created, it is possible to modify many of its characteristics. The following sections show how to perform those configuration changes.

Shrinking

To shrink a volume, complete the following steps:

1. Ensure that you have a current and verified backup of any in-use data that is stored on the volume that you intend to shrink.

- From the **Volumes** menu, select the volume that you want to shrink. Select **Actions** → **Shrink...**, as shown in Figure 6-40.

Name	Synchronized	Capacity
Anton0		44.00 GiB
Anton1		44.00 GiB
Anton2		44.00 GiB
Anton3		44.00 GiB
Anton4		44.00 GiB
FC-host-vol0		200.00 GiB
FC-host-vol1		200.00 GiB
FC-host-vol2		200.00 GiB
FS7200		10.00 GiB
SVC_Volume0		23.00 GiB
SVC_Volume1		23.00 GiB
SVC_Volume2		23.00 GiB
SVC_Volume3		23.00 GiB
SVC_Volume4		23.00 GiB
iscsi-cluster-vol		150.00 GiB
iscsi-host-vol0		100.00 GiB
iscsi-host-vol1		100.00 GiB
iscsi-host-vol2		250.00 GiB
iscsi-host-vol3		250.00 GiB
itso-basic00-Pool0		10.00 GiB
itso-basic00-Pool1		10.00 GiB
itso-mirrored00-Pool0		10.00 GiB
Copy 0*	✓ Online	Yes 10.00 GiB
Copy 1	✓ Online	Yes 10.00 GiB
itso-thin00-Pool0	✓ Online	10.00 GiB
itso-thin01-Pool1	✓ Online	10.00 GiB

Figure 6-40 Volume Shrink menu item

- Specify either **Shrink by** or **Final size** (the other choice is calculated automatically), as shown in Figure 6-41 on page 347.

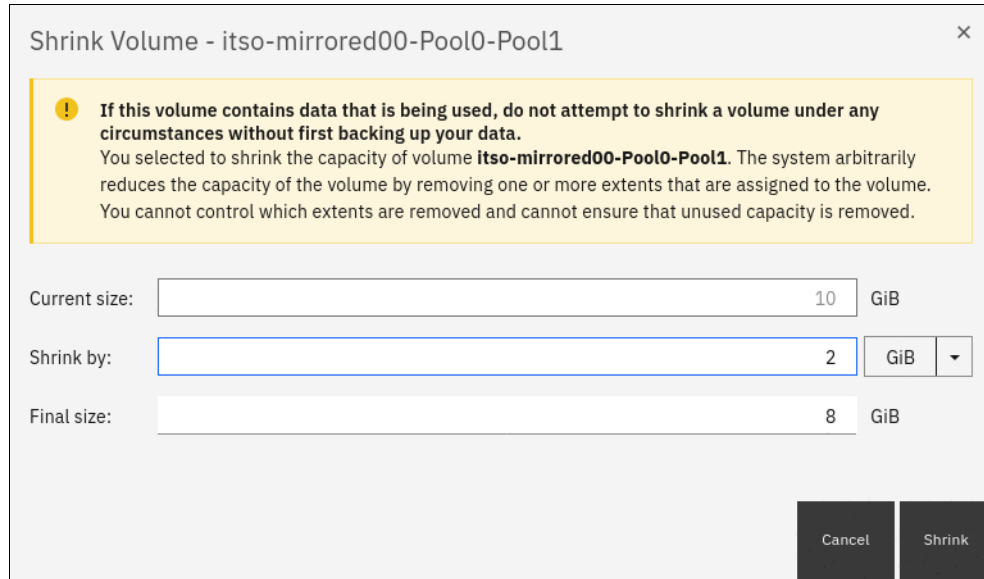


Figure 6-41 Specifying the size of the shrunk volume

Note: The storage system reduces the volume capacity by removing one or more arbitrarily selected extents. Do not shrink a volume that contains data that is being used unless you have a current and verified backup of the data.

4. Click **OK** to confirm the action, as shown in Figure 6-42.

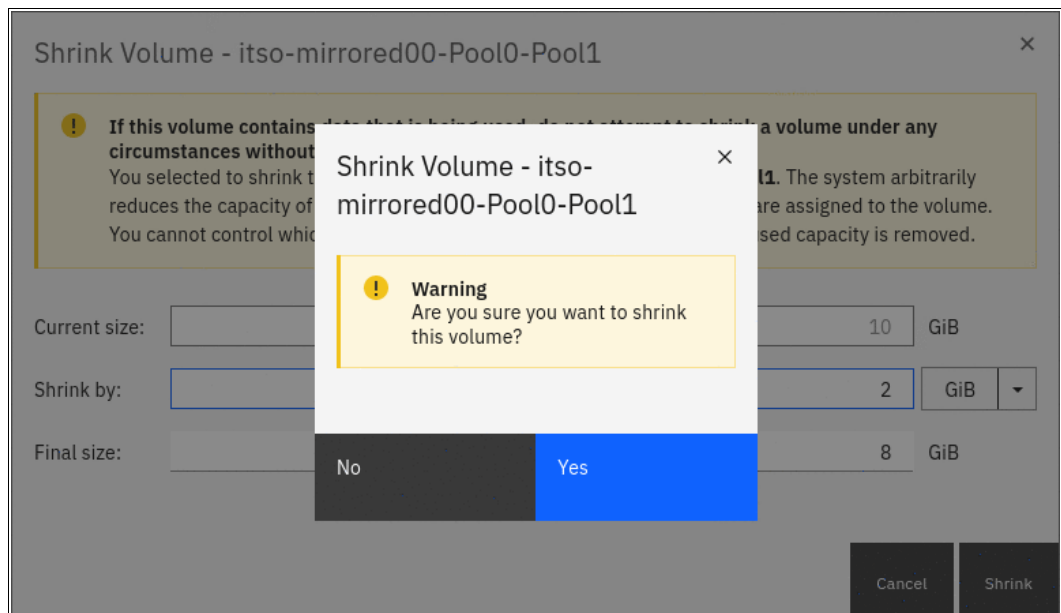
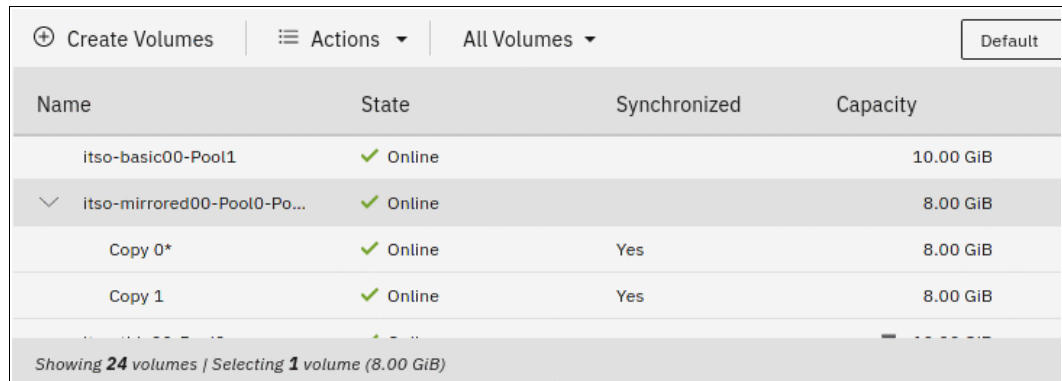


Figure 6-42 Confirming the volume shrinking operation

5. After the operation completes, you can see the volume with the new size by selecting **Volumes** → **Volumes**, as shown in Figure 6-43.



Name	State	Synchronized	Capacity
itso-basic00-Pool1	✓ Online		10.00 GiB
itso-mirrored00-Pool0-Po...	✓ Online		8.00 GiB
Copy 0*	✓ Online	Yes	8.00 GiB
Copy 1	✓ Online	Yes	8.00 GiB

Showing 24 volumes / Selecting 1 volume (8.00 GiB)

Figure 6-43 Shrunk volume size

Note: Version 8.4 introduces the ability to shrink a volume while it is formatting.

Expanding

To expand a volume, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to expand. Select **Actions** → **Expand...**, as shown in Figure 6-44.

Name	Synchronized	Capacity
Anton0		44.00 GiB
Anton1		44.00 GiB
Anton2		44.00 GiB
Anton3		44.00 GiB
Anton4		44.00 GiB
FC-host-vol0		200.00 GiB
FC-host-vol1		200.00 GiB
FC-host-vol2		200.00 GiB
FS7200		10.00 GiB
SVC_Volume0		23.00 GiB
SVC_Volume1		23.00 GiB
SVC_Volume2		23.00 GiB
SVC_Volume3		23.00 GiB
SVC_Volume4		23.00 GiB
iscsi-cluster-vol		150.00 GiB
iscsi-host-vol0		100.00 GiB
iscsi-host-vol1		100.00 GiB
iscsi-host-vol2		250.00 GiB
iscsi-host-vol3		250.00 GiB
itso-basic00-Pool0		10.00 GiB
itso-basic00-Pool1		10.00 GiB
itso-mirrored00-Pool		8.00 GiB
Copy 0*	✓ Online	Yes 8.00 GiB
Copy 1	✓ Online	Yes 8.00 GiB
itso-thin00-Pool0	✓ Online	10.00 GiB
itso-thin01-Pool1	✓ Online	10.00 GiB

Actions
Rename...
Map to Host or Host Cluster...
Shrink...
Expand...
Modify Capacity Savings...
Modify Mirror Sync Rate...
Cache Mode...
Modify Open VMS UDID...
Remove Private Mappings...
View Mapped Hosts...
View Member MDisks...
Modify I/O Group...
Cloud Volumes
Capacity Savings
Migrate to Another Pool...
Export to Image Mode
Duplicate...
Add Volume Copy...
Enable Access to Stale Copy
Edit Throttle...
View All Throttles...
Delete
Modify Properties...
Properties

Figure 6-44 Volume Expand menu item

- Specify either **Expand by:** or **Final size:** (the other value is calculated automatically) and click **Expand**, as shown in Figure 6-45.

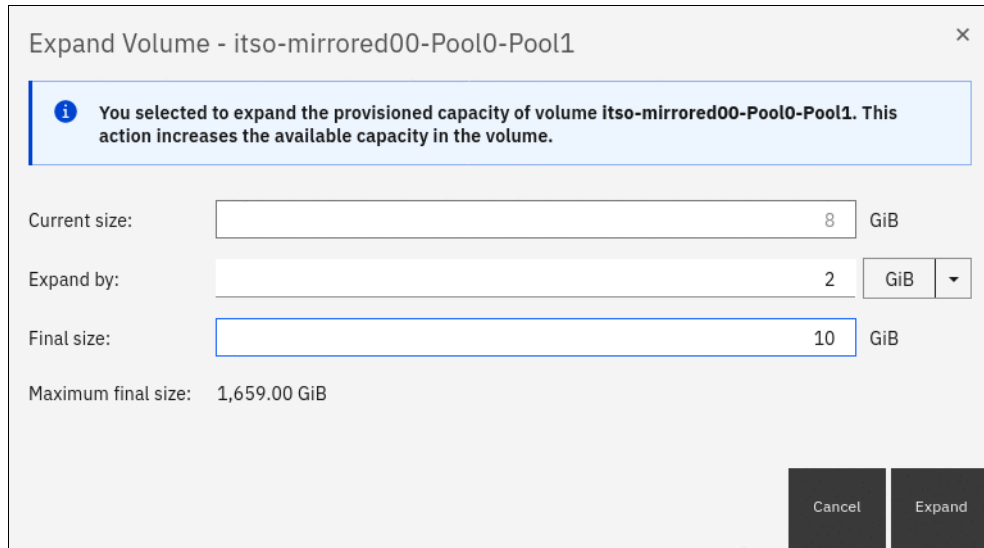


Figure 6-45 Specifying the expanded volume size

- After the operation completes (including the formatting of the extra space), you can see the volume with the new size by selecting **Volumes** → **Volumes**, as shown in Figure 6-46.

Name	State	Synchronized	Capacity
itso-basic00-Pool0	Online		10.00 GiB
itso-basic00-Pool1	Online		10.00 GiB
itso-mirrored00-Pool0-Po...	Online		10.00 GiB
Copy 0*	Online	Yes	10.00 GiB
Copy 1	Online	Yes	10.00 GiB
itso-thin00-Pool0	Online		10.00 GiB
itso-thin01-Pool1	Online		10.00 GiB

Showing 24 volumes | Selecting 0 volumes

Figure 6-46 Expanded volume size

Note: Expanding a volume is not sufficient to increase the available space that is visible to the host. The host must become aware of the changed volume size at the OS level, for example, through a bus rescan. More operations at the logical volume manager (LVM) or file system levels might be needed before more space is visible to applications running on the host.

Note: Version 8.4 introduces the ability to expand a volume while it is formatting.

Modifying capacity savings

This action is available only for space-efficient volumes. To modify capacity savings options for a volume, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to modify. Select **Actions** → **Modify Capacity Savings...**, as shown in Figure 6-47.



Figure 6-47 Modify Capacity Savings menu item

2. Select the capacity savings option that you want, as shown in Figure 6-48.

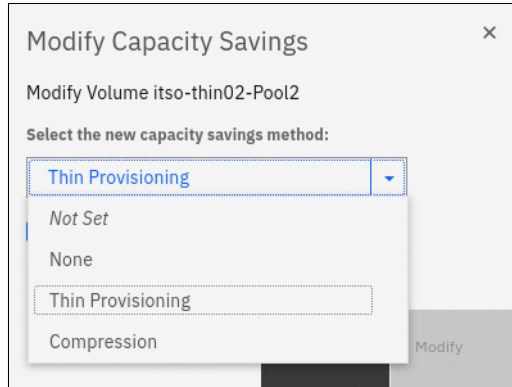


Figure 6-48 Capacity savings options for a volume

3. For volumes that are configured in a DRP, it is possible to enable deduplication, as shown in Figure 6-49.

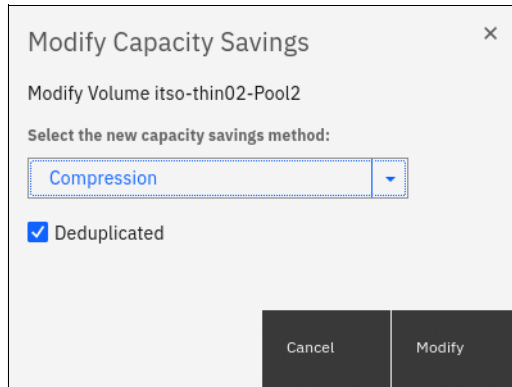


Figure 6-49 Enabling deduplication on a volume

After you configure the capacity savings options of a volume, click **Modify** to apply them. When the operation completes, you are returned to the Volumes view.

Modifying the mirror sync rate

This action is available only for mirrored volumes. To modify the mirror sync rate of a volume, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to modify. Select **Actions** → **Modify Mirror Sync Rate...**, as shown in Figure 6-50.

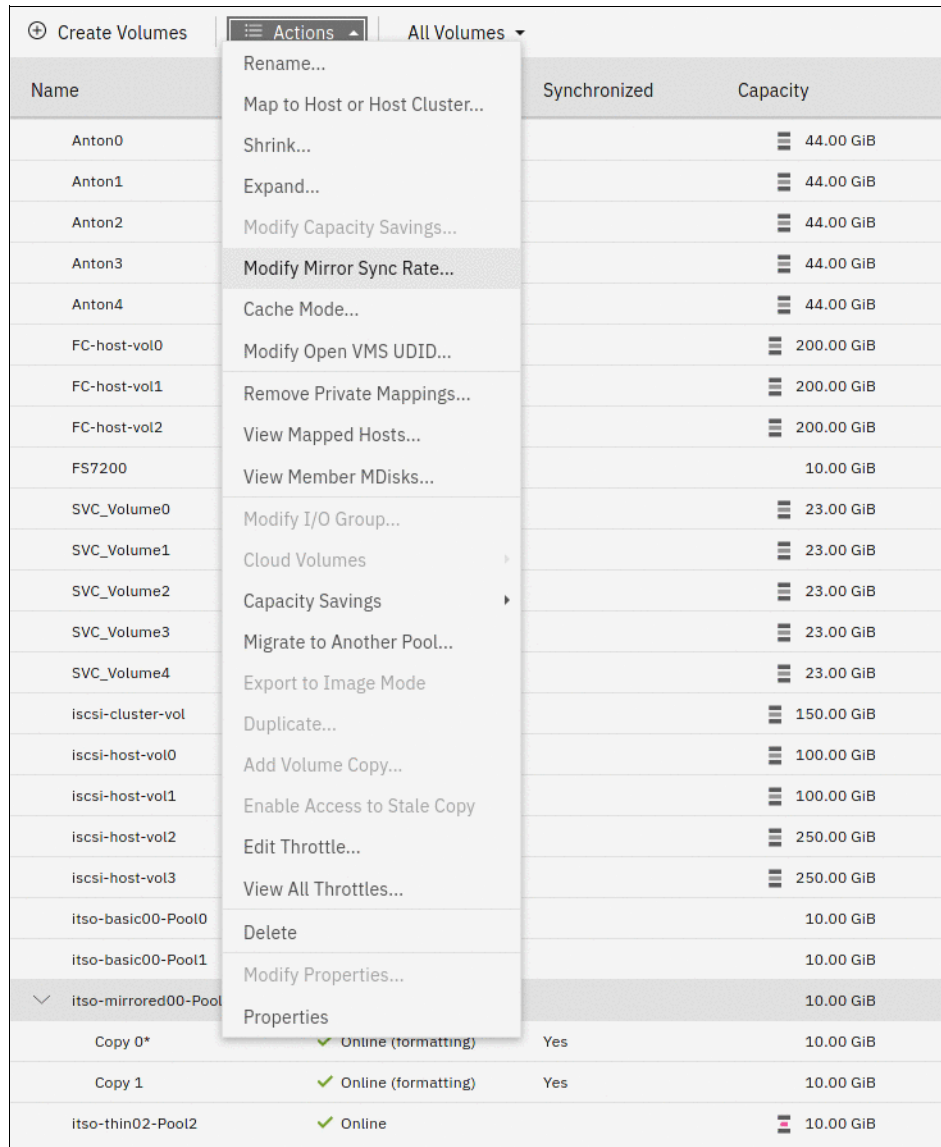


Figure 6-50 Modify Mirror Sync Rate menu item

2. Select the mirror sync rate from the list. Available values are 0 KBps - 64 MBps. Click **Modify** to set the mirror sync rate for the volume, as shown in Figure 6-51.

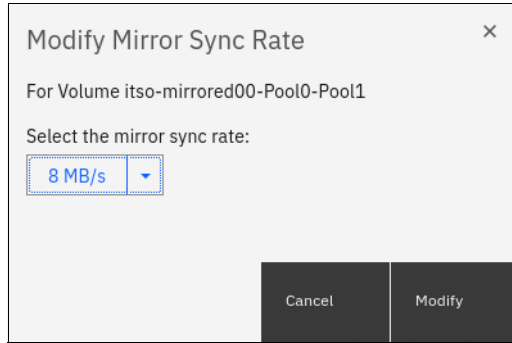


Figure 6-51 Setting the volume mirror sync rate

When the operation completes, you are returned to the Volumes view.

Changing the volume cache mode

To change the volume cache mode, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to modify. Select **Actions** → **Cache Mode...**, as shown in Figure 6-52.

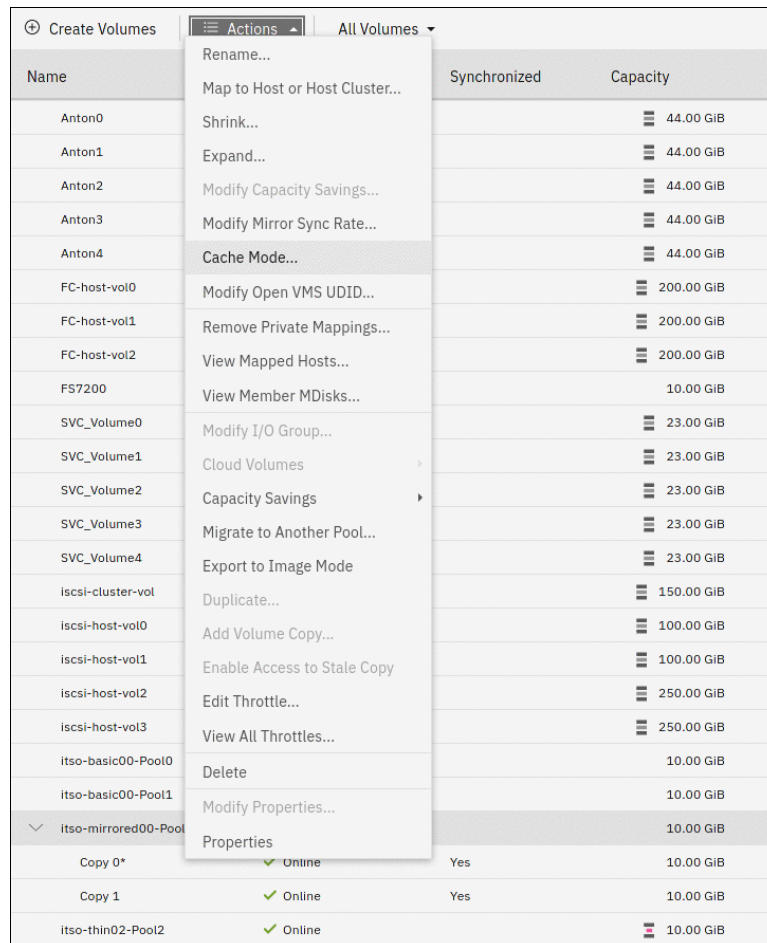


Figure 6-52 Modifying the volume cache mode

2. Select the cache mode that you want for the volume from the drop-down list and click **OK**, as shown in Figure 6-53.

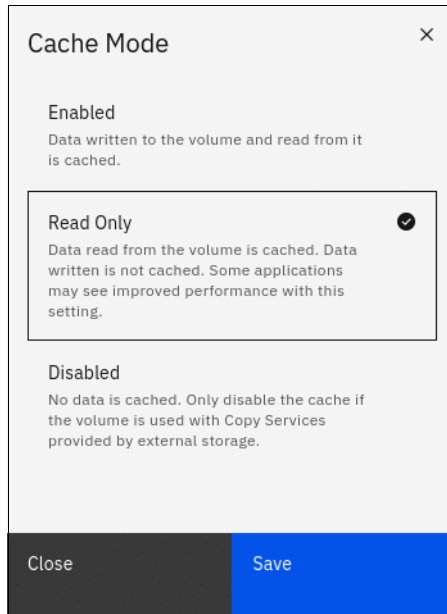


Figure 6-53 Setting the volume cache mode

When the operation completes, you are returned to the Volumes view.

Modify OpenVMS UDID menu

To change the OpenVMS UDID, use the **Modify OpenVMS UDID** menu.

A *UDID* is a nonnegative integer that is used in the creation of the OpenVMS device name. All fibre-attached volumes have an allocation class of \$1\$, followed by the letters DGA, and then followed by the UDID. All storage unit LUNs that you assign to an OpenVMS system need an UDID so that the OS can detect and name the device. LUN 0 must also have a UDID, but the system displays LUN 0 as \$1\$GGA<UDID>, not as \$1\$DGA<UDID>. For more information about fibre-attached storage devices, see [Guidelines for OpenVMS Cluster Configurations](#).

To change a volume OpenVMS UDID, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to modify. Select **Actions** → **Modify Open VMS UDID...**, as shown in Figure 6-54.

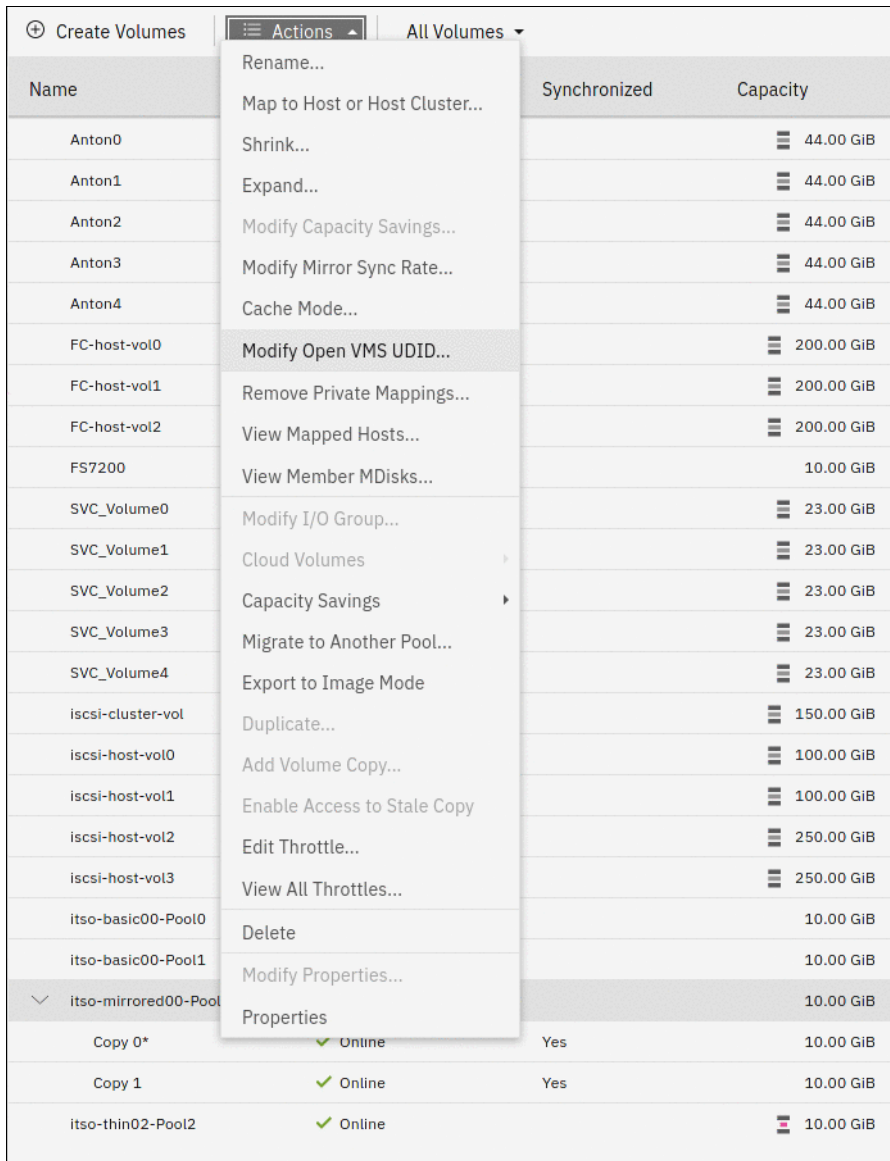


Figure 6-54 Modify Open VMS UDID menu item

2. Specify the UDID for the volume and click **Modify**, as shown in Figure 6-55 on page 357.

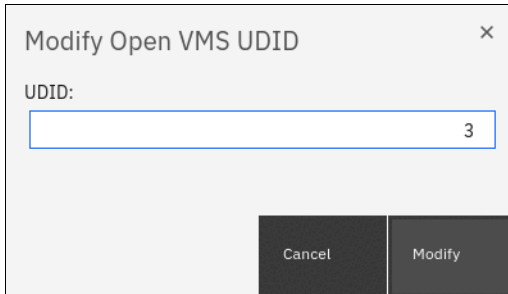


Figure 6-55 Setting the volume UDID

When the operation completes, you are returned to the Volumes view.

6.5.7 Deleting a volume

When you try to delete a volume, the system verifies whether it is a part of a host mapping, FlashCopy mapping, or RC relationship. If any of these mappings exists, the delete attempt fails unless the **-force** parameter is specified on the corresponding remove commands. If volume protection is enabled, a delete fails (even if the **-force** parameter is specified) if the system detects any I/O activity to the volume within the configured timeframe. The **-force** parameter overrides the volume dependencies, not the volume protection setting.

To delete a volume, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to modify. Select **Actions** → **Delete**, as shown in Figure 6-56.



Figure 6-56 Volume Delete menu item

2. Review the list of volumes that is selected for deletion and provide the number of volumes that you intend to delete, as shown in Figure 6-57. Click **Delete** to remove the volume from the system configuration.

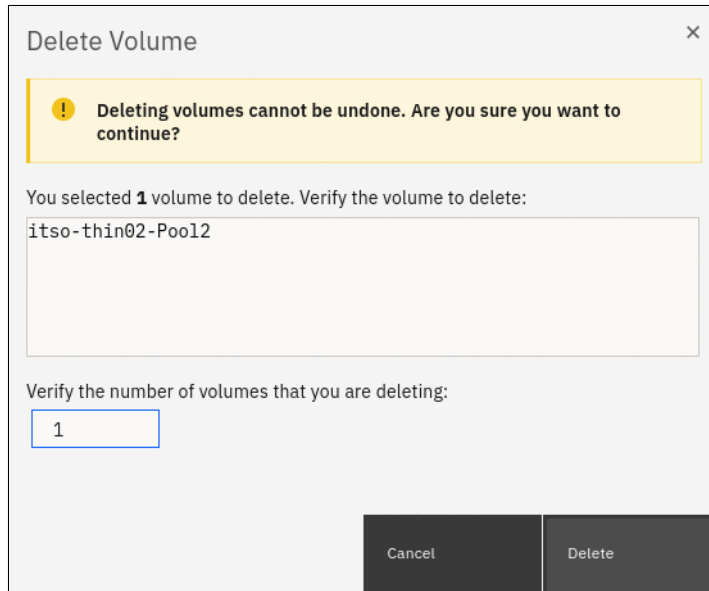


Figure 6-57 Confirming the volume deletion

When the operation completes, you are returned to the Volumes view.

6.5.8 Mapping a volume to a host

To make a volume available to a host or cluster of hosts, it must be mapped. A volume can be mapped to the host at creation time or later.

To map a volume to a host or cluster, complete the following steps:

1. From the **Volumes** menu, select the volume that you want to modify. Select **Actions** → **Map to Host or Host Cluster...**, as shown in Figure 6-58 on page 359.

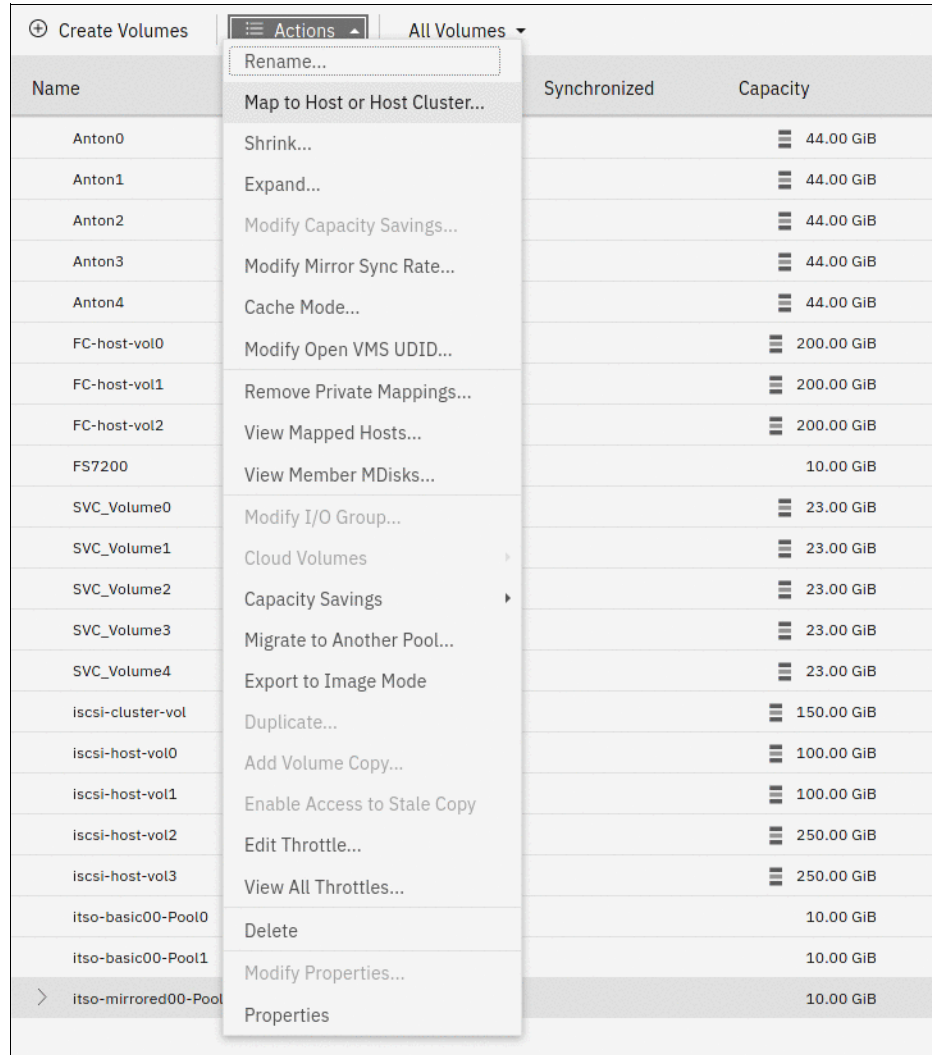


Figure 6-58 Volume mapping menu item

Tip: An alternative way of opening the **Actions** menu is to highlight (select) a volume and right-click.

- The Create Mapping window opens. In this window, select whether to create a mapping to a host or host cluster. The list of objects of the appropriate type is displayed. Select to which hosts or host clusters the volume should be mapped.

You can either allow the storage system to assign the SCIS LUN ID to the volume by selecting the **System Assign** option, or select **Self Assign** and provide the LUN ID yourself. Click **Next** to proceed to the next step.

In Figure 6-59, a single volume is mapped to a host and the storage system assigns the SCSI LUN IDs.

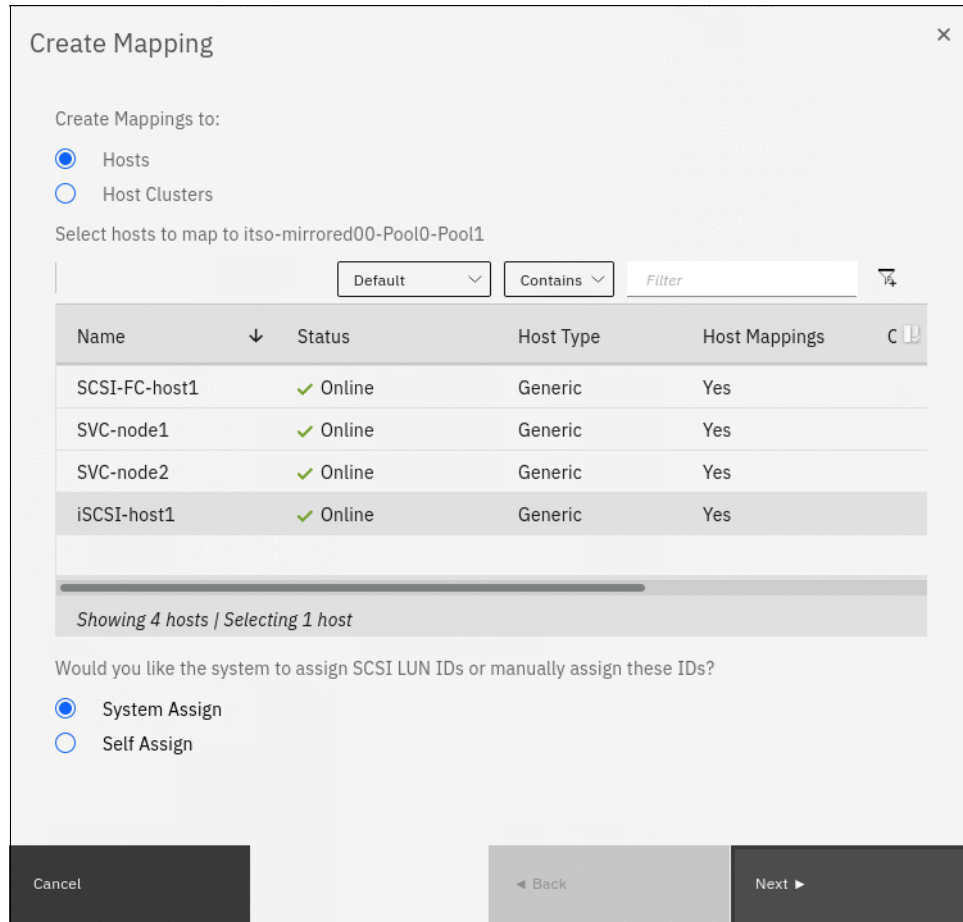


Figure 6-59 Mapping a volume to a host

3. A summary window opens and shows all the volume mappings for the selected host. The new mapping is highlighted, as shown in Figure 6-60. Review the future configuration state and click **Map Volumes** to map the volume.

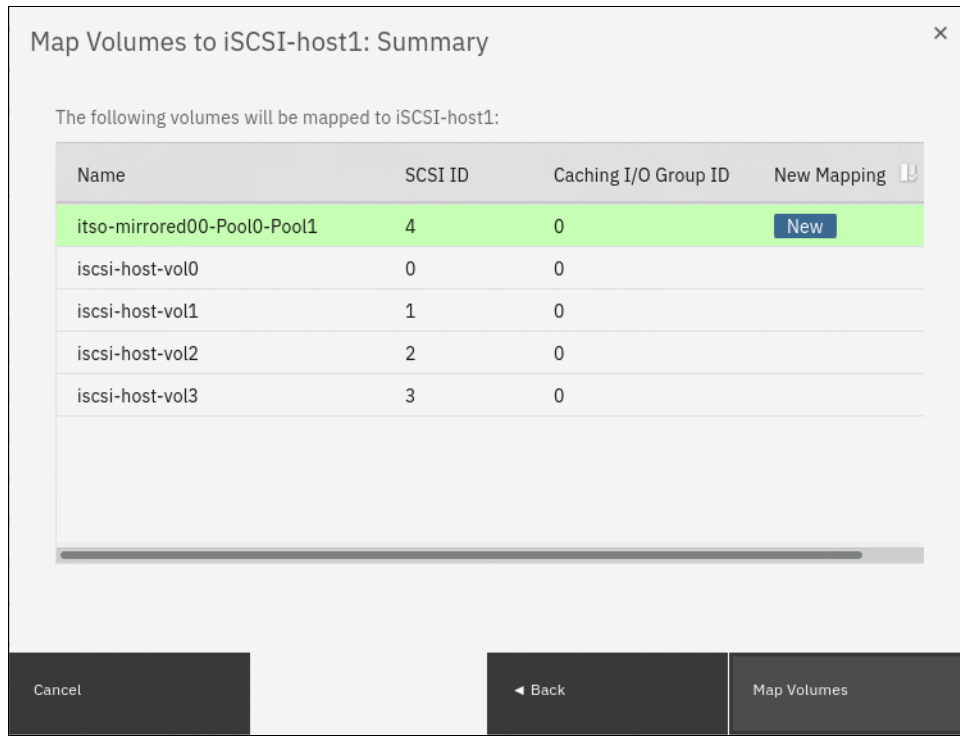


Figure 6-60 Mapping a volume to a host: Summary

4. After the task completes, the wizard returns to the Volumes window. You can list the volumes that are mapped to the host by selecting **Hosts** → **Mappings**, as shown in Figure 6-61.

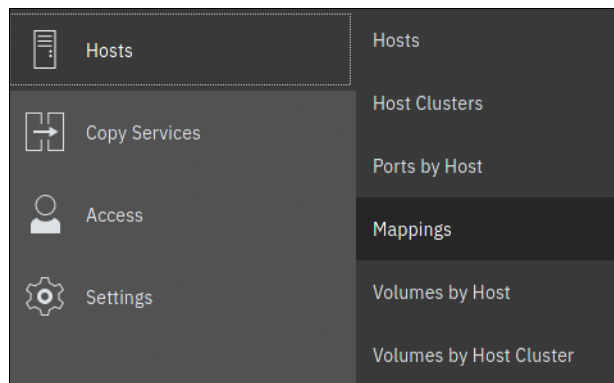


Figure 6-61 Accessing the Hosts Mapping view

A window with a list of volumes that are mapped to all hosts opens, as shown in Figure 6-62.

Host Name	SCSI ID	Volume Name	UID	I/O Group ID	I/O Group Name
iSCSI-host1	1	iscsi-host-vol1	600507640084031DD800000000000061	0	io_grp0
iSCSI-host1	4	itso-mirrored00-Pool0-Pool1	600507640084031DD80000000000006A	0	io_grp0
iSCSI-host1	0	iscsi-host-vol0	600507640084031DD800000000000060	0	io_grp0
SCSI-FC-host1	2	iscsi-cluster-vol	600507640084031DD800000000000065	0	io_grp0

Figure 6-62 List of volume to host mappings

To see volumes that are mapped to clusters instead of hosts, change the value that is shown in the upper left (see Figure 6-62) from **Private Mappings** to **Shared Mappings**.

Note: You can use the filter to display only the hosts or volumes that you want to see.

The host can now access the mapped volume. For more information about discovering the volumes on the host, see Chapter 7, “Hosts” on page 405.

To remove the volume to host mapping, in the **Hosts** → **Mappings** view, select the volume or volumes, right-click, and click **Unmap Volumes**, as shown in Figure 6-63.



Figure 6-63 Unmap Volumes menu item

In the Delete Mapping window, enter the number of volumes that you intend to unmap, as shown in Figure 6-64. This action is as a security measure that minimizes changes that result from an accidental unmap of an invalid volume.

Note: Removing volume to host mapping makes the volume unavailable to the host. Make sure that the host is prepared for the operation. An improperly run volume unmap operation might cause data unavailability or loss.

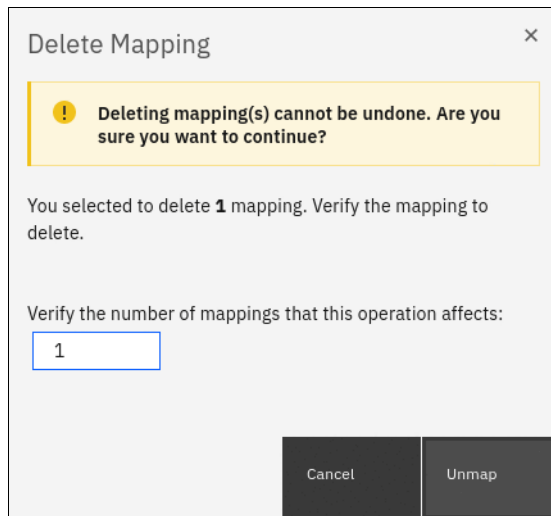


Figure 6-64 Volume unmap confirmation window

Click **Unmap** to complete the operation. Volume mapping is removed and is no longer displayed in the volume map view, as shown in Figure 6-65.

Host Name	SCSI ID	Volume Name	UID	I/O Group ID	I/O Group Name
iSCSI-host1	1	iscsi-host-vol1	600507640084031DD800000000000061	0	io_grp0
iSCSI-host1	0	iscsi-host-vol0	600507640084031DD800000000000060	0	io_grp0
SCSI-FC-host1	2	iscsi-cluster-vol	600507640084031DD800000000000065	0	io_grp0

Figure 6-65 Volume mapping removed

6.5.9 Modify I/O Group or Non-disruptive Volume Move

The NDVM function was introduced in Version 6.3 of the SVC and Storwize code. This function is also used for changing the preferred node within the I/O group of a volume or volumes. Moving volumes between I/O groups is a task that is sometimes needed for workload balancing or migration between clustered enclosures. Because the caching I/O group is the enclosure that mediates the I/O between the host system and the storage, moving a volume to another I/O group also shifts the resource consumption of CPU, memory, and both front-end host and back-end storage traffic. In enclosure-based systems, you generally want to align the I/O group with the enclosure that contains the pool in which the volume is.

The primary use case for this function is upgrading environments from Storwize V7000 to the IBM FlashSystem 7200, IBM FlashSystem 9150, or IBM FlashSystem 9200 because those systems can be clustered with a Storwize V7000.

Note: Certain conditions prevent the changing of I/O group dynamically with NDVM for a volume. If the volume is using data reduction in a DRP, or if a volume is a member of a FlashCopy map and is in an RC relationship, the first command in the sequence, **addvdiskaccess**, fails.

Complete the following steps:

1. To initiate the process, select the volumes, right-click, and then select **Modify I/O Group**, or select the **Actions** drop-down menu, as shown in Figure 6-66.

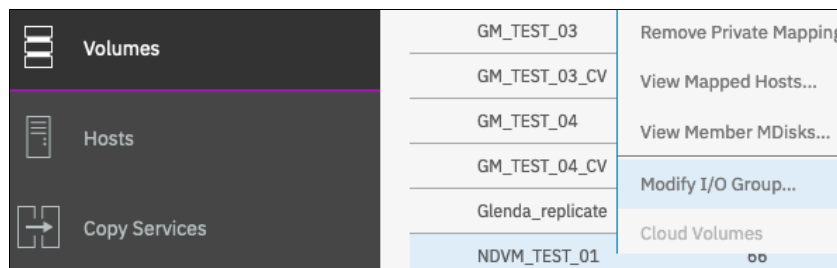


Figure 6-66 Modify I/O Group menu item

If there are no host mappings for the volumes, then the operation immediately displays the target I/O group selection dialog box, as shown in Figure 6-67.

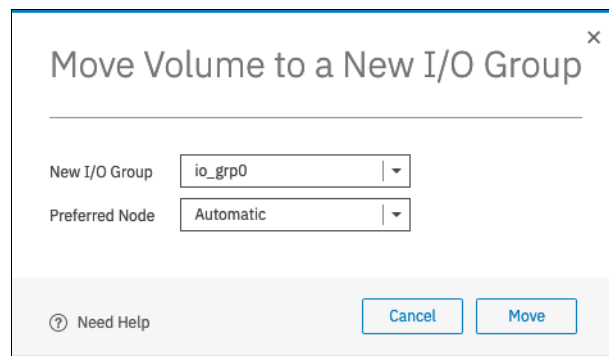


Figure 6-67 I/O Group selection dialog box

2. Select the new I/O group and preferred node and click **Move** to move the volume to the new I/O group and preferred node. This GUI action runs the following commands:
 - **addvdiskaccess -iogrp {new i/o group} {volume}**
Adds the specified I/O group to the set of I/O groups in which the volume can be made accessible to hosts.
 - **movevdisk -iogrp {new i/o group} {volume}**
Moves the preferred node of the volume to the new (target) caching I/O group.
 - **rmvdiskaccess -iogrp {old i/o group} {volume}**
Removes the old (source) I/O group from the set of I/O groups in which the volume can be made accessible to hosts.

In the likely case where the volume is mapped to a host, the GUI detects the host mapping and starts a wizard, as shown in Figure 6-68, to ensure that the correct steps are performed in the correct order. You are required to configure zoning between the host and the new I/O group and ensure that all hosts to which the volume is mapped discovers new paths to the volume. The steps that are required to modify the I/O group of a mapped volume are shown below.

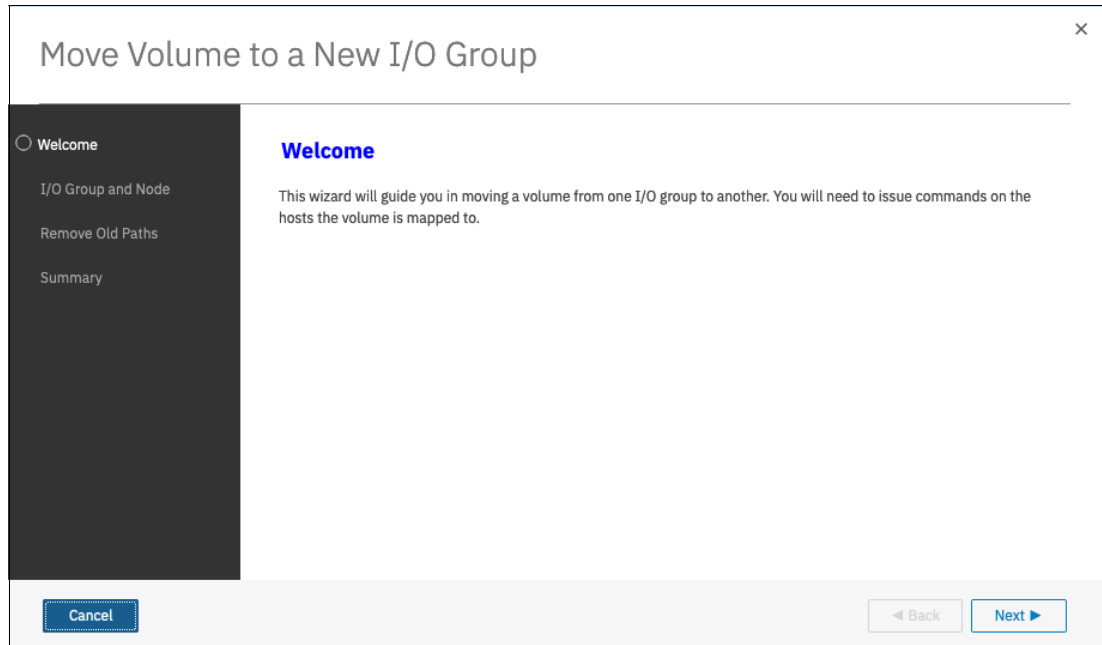


Figure 6-68 Modify I/O Group for a mapped volume wizard: Welcome

1. Verify that all hosts that use the volume are zoned to the target I/O group, and click **Next** to proceed to the new I/O group selection window, as shown in Figure 6-69.

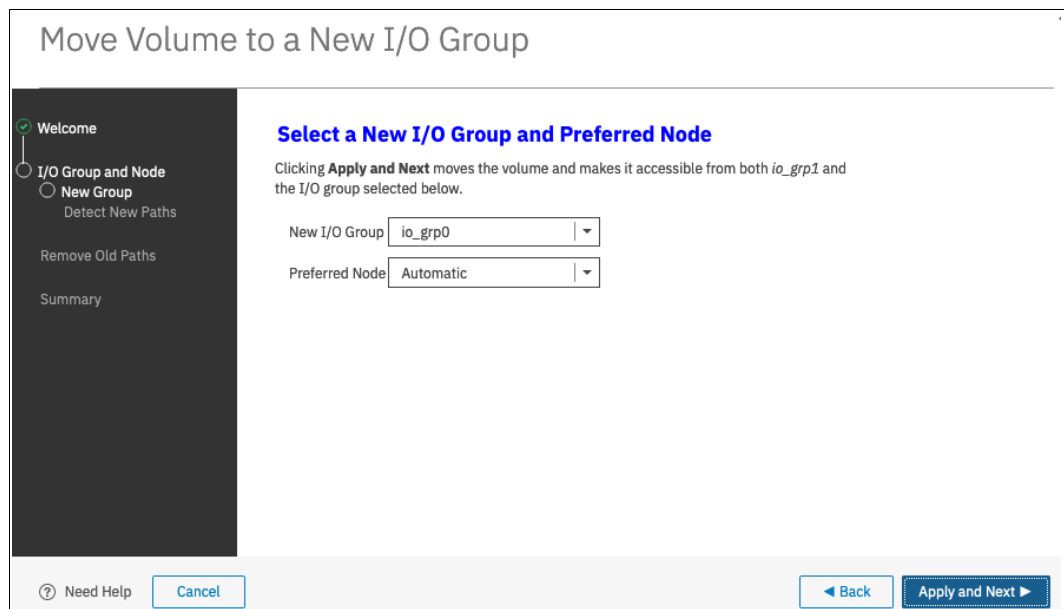


Figure 6-69 Modify I/O Group for a mapped volume wizard: I/O group selection window

2. Select the new (target) I/O group and preferred node, and click **Apply** and **Next**. The GUI runs the following commands:

– **addvdiskaccess -iogrp {new i/o group} {volume}**

Adds the specified I/O group to the set of I/O groups in which the volume can be made accessible to hosts.

– **movevdisk -iogrp {new i/o group} {volume}**

Moves the preferred node of the volume to the new (target) caching I/O group.

A window opens and confirms the successful completion of the commands, as shown in Figure 6-70.

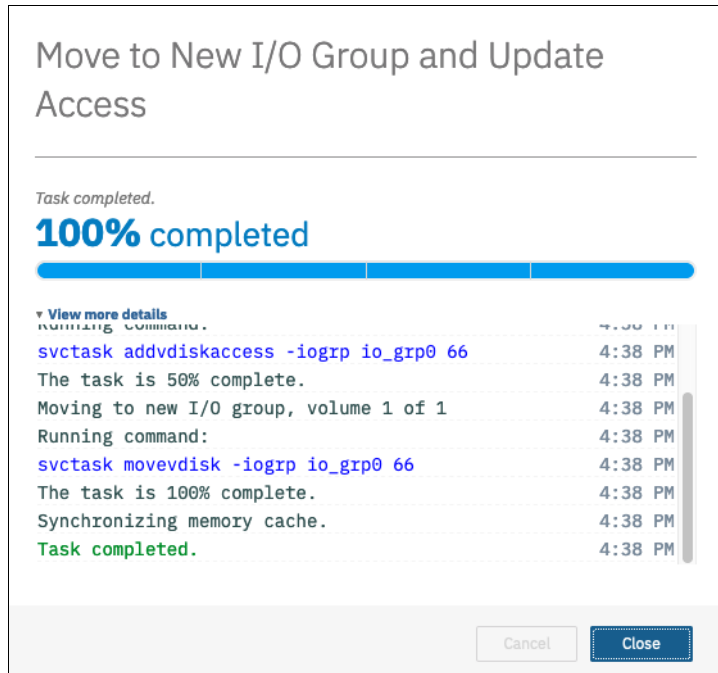


Figure 6-70 Modify I/O Group for a mapped volume wizard: First stage completed (details)

3. Click **Close** to proceed to the validation window, as shown in Figure 6-71 on page 367. Click **Need Help** to see information about how to prepare the host for the volume move. After the host is ready for the volume path change, check the box confirming that path validation was performed on the host.

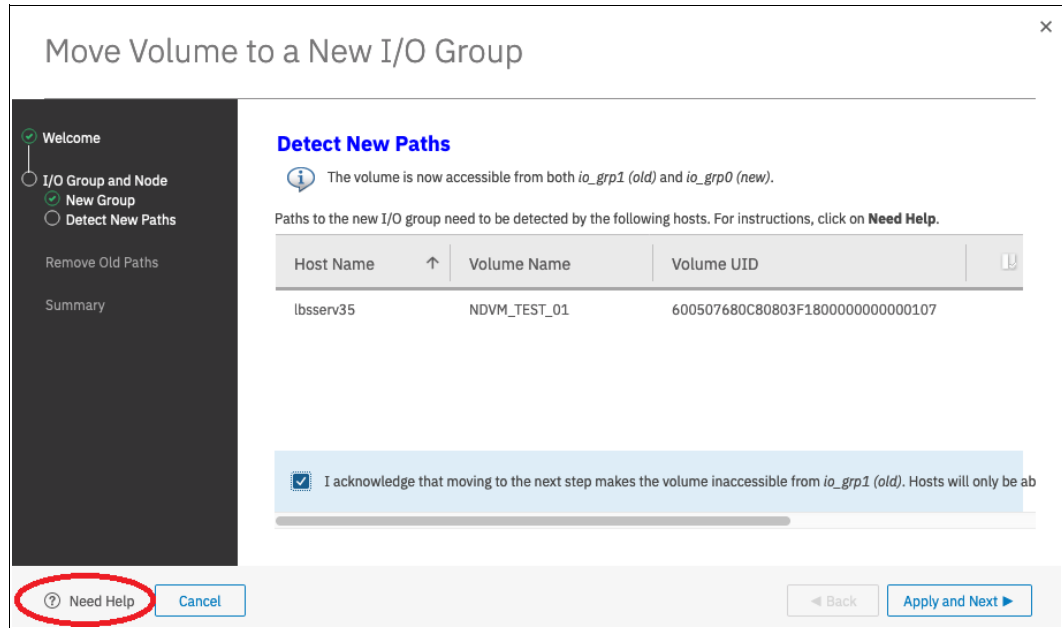


Figure 6-71 Modify I/O Group for a mapped volume wizard: Validation

Note: Failure to ensure that the host discovered the new paths to all the volumes might result in this process being disruptive and cause the host to lose access to the moved volume or volumes.

- After validation is complete and the acknowledgment box is checked, click **Apply** and **Next**, as shown in Figure 6-72.

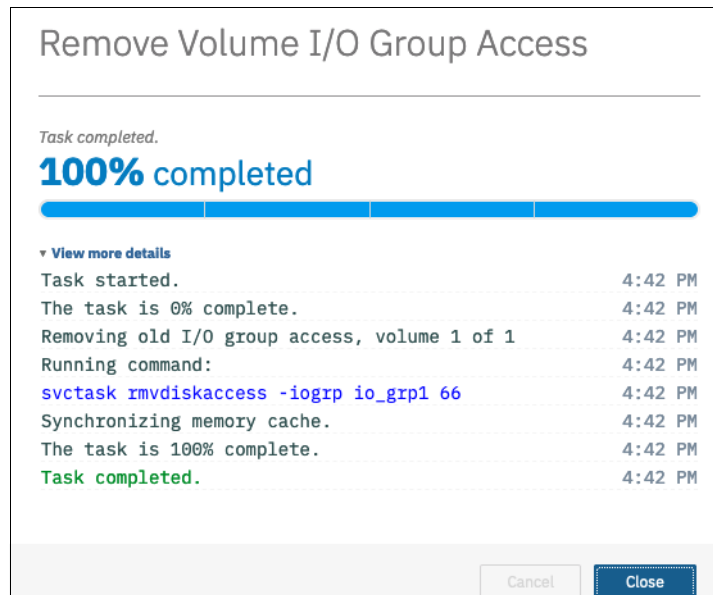


Figure 6-72 Modify I/O Group for a mapped volume wizard: Second stage completes (details)

5. Close the detail window to proceed to the final window of the wizard, as shown in Figure 6-73. Now, hosts cannot access the volume through the old I/O group, and depending on the OS, you might need a restart to remove the dead/stale paths.

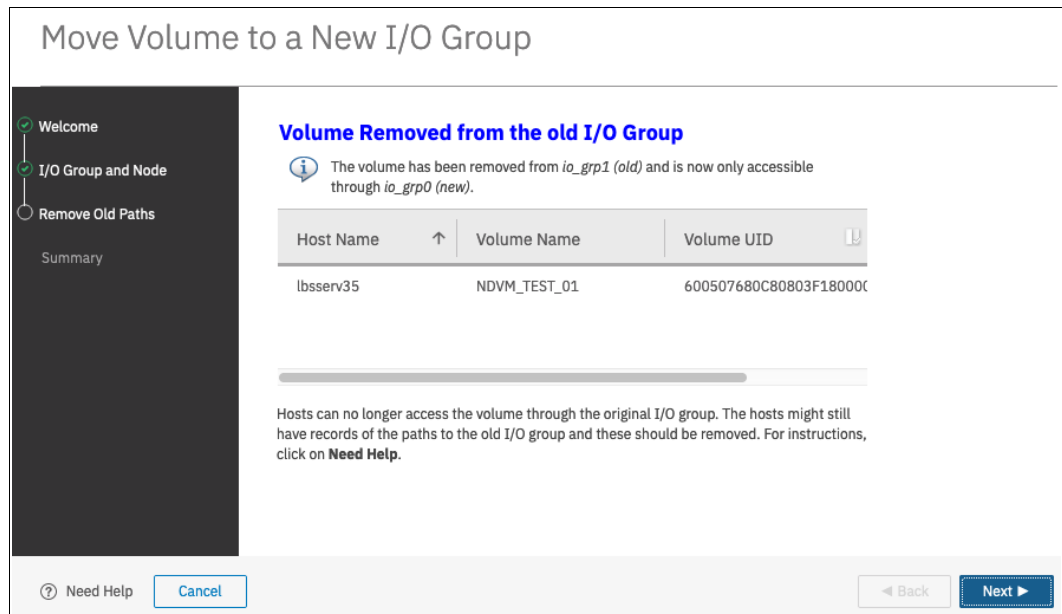


Figure 6-73 Modify I/O Group for a mapped volume wizard: Operation complete

6.5.10 Migrating a volume to another storage pool

IBM Spectrum Virtualize enables online volume migration with no applications downtime. Volumes can be moved between storage pools without affecting business workloads that are running on these volumes.

There are two ways to perform volume migration: by using the volume migration feature and by creating a volume copy.

Volume migration by using the migration feature

Volume migration is a low-priority process that does not affect the performance of the IBM Spectrum Virtualize system. However, as subsequent volume extents are moved to the new storage pool, the performance of the volume is determined more by the characteristics of the new storage pool.

Note: You cannot move a volume copy that is compressed to an I/O group that contains a node that does not support compressed volumes.

To migrate a volume to another storage pool, complete the following steps:

1. In the **Volumes** menu, highlight the volume that you want to migrate. Select **Actions** → **Migrate to Another Pool...**, as shown in Figure 6-74.

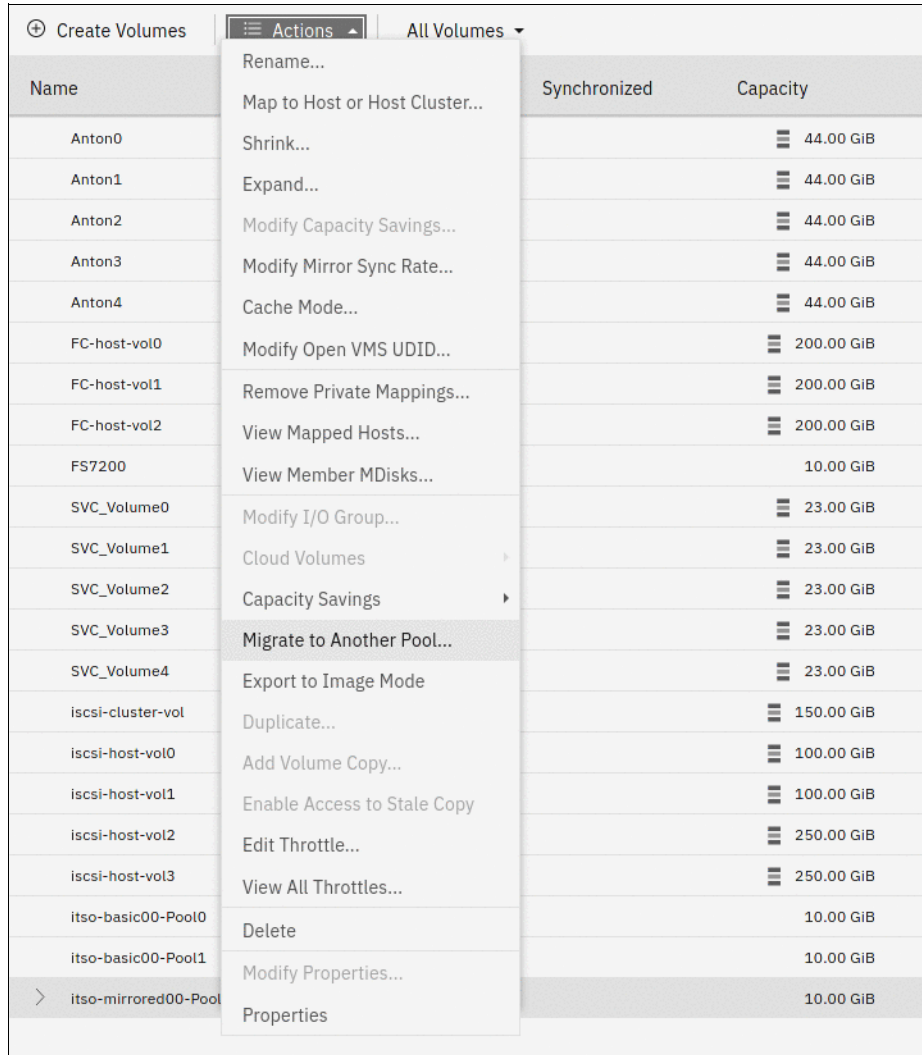


Figure 6-74 Migrate Volume Copy: Selecting a menu item

- The Migrate Volume Copy window opens. If your volume consists of more than one copy, select the copy that you want to migrate to another storage pool, as shown in Figure 6-75. If the selected volume consists of one copy, the volume copy selection window is not displayed.

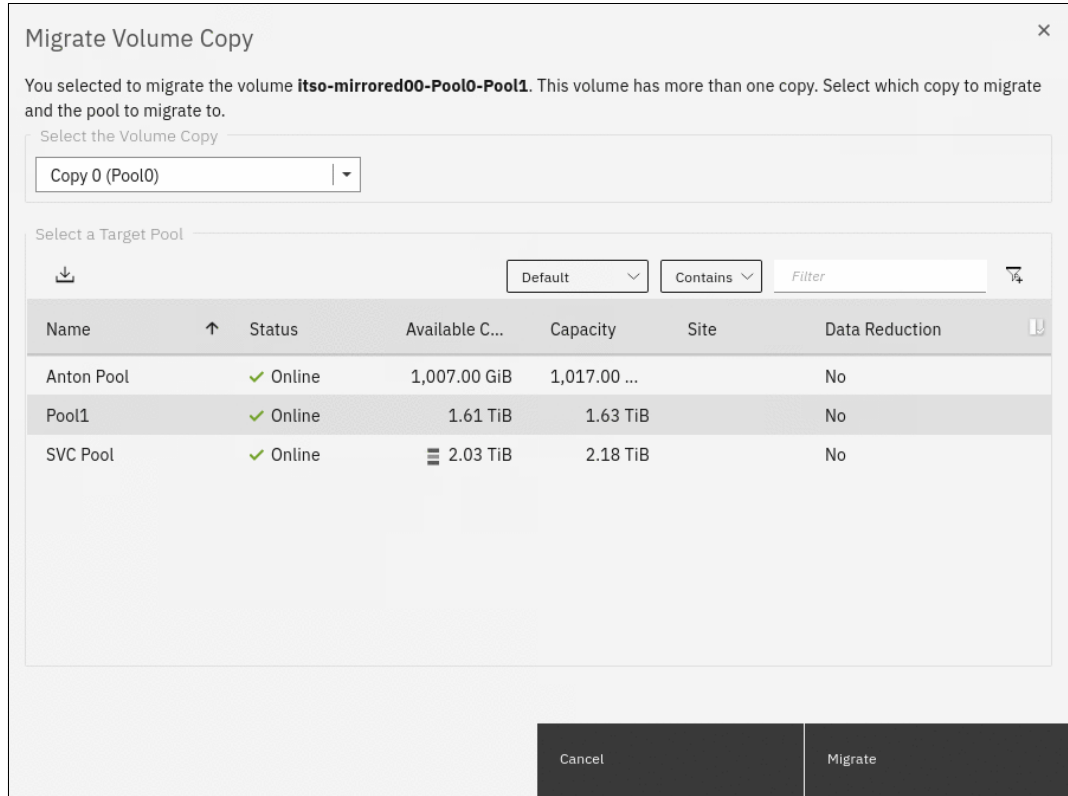


Figure 6-75 Migrate Volume Copy: Selecting the volume copy

- Select the new target storage pool and click **Migrate**, as shown in Figure 6-75. The Select a Target Pool window displays the list of all pools that are a valid migration copy target for the selected volume copy.
- You are returned to the Volumes view. The time that it takes for the migration process to complete depends on the size of the volume. The status of the migration can be monitored by selecting **Monitoring** → **Background Tasks**, as shown in Figure 6-76.



Figure 6-76 Monitoring the volume migration progress

After the migration task completes, the completed migration task is visible in the Recently Completed Task window of the **Background Tasks** menu, as shown in Figure 6-77 on page 371.

Name	Completion Time
Synchronized volume itso-thin02-Pool2, copy 1	44 minutes ago
Migrated volume itso-mirrored00-Pool0-Pool1, copy...	4 minutes ago

Figure 6-77 Volume migration complete

In the **Volumes** → **Volumes** menu, the volume copy is now displayed in the target storage pool, as shown in Figure 6-78.

Name	State	Synchronized	Pool	Capacity
Anton0	✓ Online		SVC Pool	44.00 GiB
Anton1	✓ Online		SVC Pool	44.00 GiB
Anton2	✓ Online		SVC Pool	44.00 GiB
Anton3	✓ Online		SVC Pool	44.00 GiB
Anton4	✓ Online		SVC Pool	44.00 GiB
FC-host-vol0	✓ Online		SVC Pool	200.00 GiB
FC-host-vol1	✓ Online		SVC Pool	200.00 GiB
FC-host-vol2	✓ Online		SVC Pool	200.00 GiB
FS7200	✓ Online		Anton Pool	10.00 GiB
SVC_Volume0	✓ Online		SVC Pool	23.00 GiB
SVC_Volume1	✓ Online		SVC Pool	23.00 GiB
SVC_Volume2	✓ Online		SVC Pool	23.00 GiB
SVC_Volume3	✓ Online		SVC Pool	23.00 GiB
SVC_Volume4	✓ Online		SVC Pool	23.00 GiB
iscsi-cluster-vol	✓ Online		SVC Pool	150.00 GiB
iscsi-host-vol0	✓ Online		SVC Pool	100.00 GiB
iscsi-host-vol1	✓ Online		SVC Pool	100.00 GiB
iscsi-host-vol2	✓ Online		SVC Pool	250.00 GiB
iscsi-host-vol3	✓ Online		SVC Pool	250.00 GiB
itso-basic00-Pool0	✓ Online		Pool0	10.00 GiB
itso-basic00-Pool1	✓ Online		Pool1	10.00 GiB
itso-mirrored00-Pool0-Po...	✓ Online		Pool1	10.00 GiB
Copy 0*	✓ Online	Yes	Pool1	10.00 GiB
Copy 1	✓ Online	Yes	Pool1	10.00 GiB

Figure 6-78 Volume copy after migration

The volume copy is now migrated without any host or application downtime to the new storage pool.

Another way to migrate single-copy volumes to another pool is to use the volume copy feature, as described in “Volume migration by adding a volume copy” on page 372.

Note: Migrating a volume between storage pools with different extent sizes is *not* supported. If you must migrate a volume to a storage pool with a different extent size, use the volume migration by adding a volume copy method.

Volume migration by adding a volume copy

IBM Spectrum Virtualize supports creating, synchronizing, splitting, and deleting volume copies. A combination of these tasks can be used to migrate volumes between storage pools.

The easiest way to migrate volumes is to use the migration feature that is described in 6.5.10, “Migrating a volume to another storage pool” on page 368. However, in some use cases, the preferred or only method of volume migration is to create a copy of the volume in the target storage pool and then remove the old copy.

Note: You can specify storage efficiency characteristics of the new volume copy differently than the ones of the primary copy. For example, you can make a thin-provisioned copy of a standard-provisioned volume.

This volume migration option can be used only for single-copy volumes. If you need to move a copy of a mirrored volume by using this method, you must delete one of the volume copies first and then create a copy in the target storage pool. This process causes a temporary loss of redundancy while the volume copies synchronize.

To migrate a volume by using the volume copy feature, complete the following steps:

1. Select the volume that you want to move, and select **Actions** → **Add Volume Copy**, as shown in Figure 6-79.

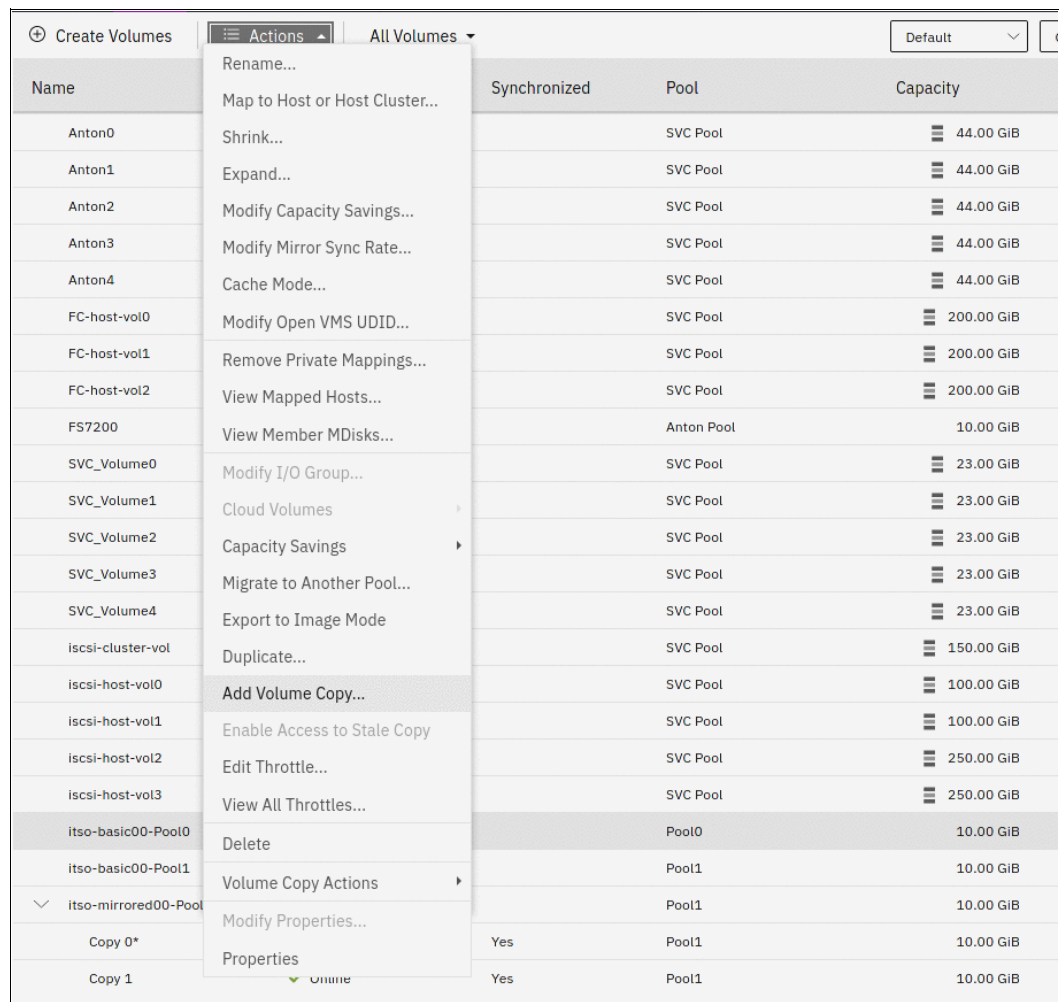


Figure 6-79 Add Volume Copy menu item

2. Create a second copy of your volume in the target storage pool, as shown in Figure 6-80. You can modify the capacity savings options for the new volume copy. In our example, a compressed copy of the volume is created in target pool Pool2. The Deduplication option is not available if either of the volume copies is not in a DRP. Click **Add** to proceed.

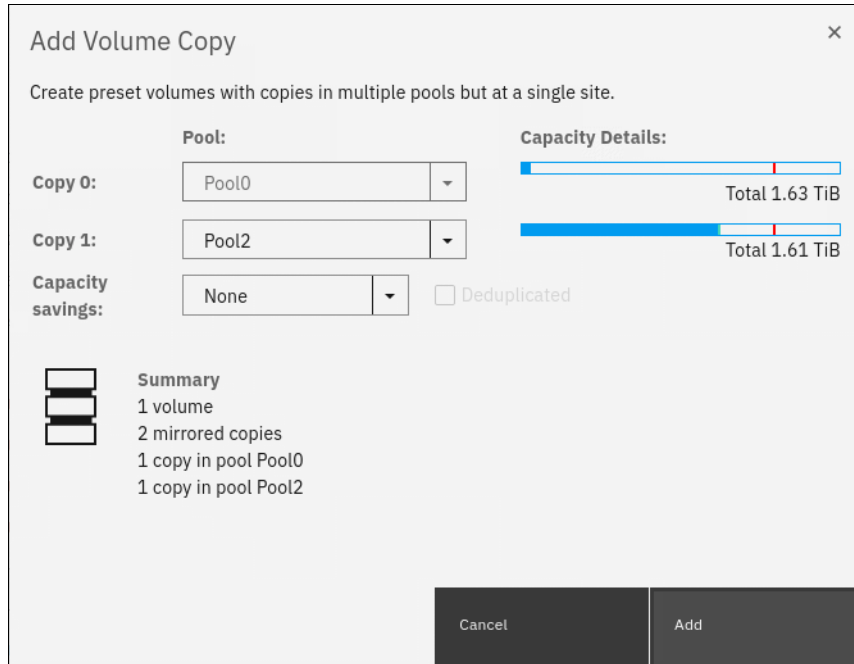


Figure 6-80 Adding a volume copy

Wait until the copies are synchronized, as shown in Figure 6-81.

Name	State	Synchronized	Pool	Capacity
iscsi-host-vol3	Online		SVC Pool	250.00 GiB
itso-basic00-Pool0	Online		Pool0	10.00 GiB
Copy 0*	Online	Yes	Pool0	10.00 GiB
Copy 1	Online	Yes	Pool2	10.00 GiB
itso-basic00-Pool1	Online		Pool1	10.00 GiB

Showing 22 volumes | Selecting 1 volume (10.00 GiB)

Figure 6-81 Verifying that the volume copies are synchronized

- Change the roles of the volume copies by making the new copy the primary copy, as shown in Figure 6-82. The current primary copy is displayed with an asterisk next to its name.

Name	State	Synchronized	Pool	Capacity
SVC_Volume1	✓ Online		SVC Pool	23.00 GiB
SVC_Volume2	✓ Online		SVC Pool	23.00 GiB
SVC_Volume3	✓ Online		SVC Pool	23.00 GiB
SVC_Volume4	✓ Online		SVC Pool	23.00 GiB
iscsi-cluster-vol	✓ Online		Pool	150.00 GiB
iscsi-host-vol0	✓ Online		Pool	100.00 GiB
iscsi-host-vol1	✓ Online		Pool	100.00 GiB
iscsi-host-vol2	✓ Online		Pool	250.00 GiB
iscsi-host-vol3	✓ Online		Pool	250.00 GiB
itso-basic00-Pool0	✓ Online			10.00 GiB
Copy 0*	✓ Online			10.00 GiB
Copy 1	✓ Online			10.00 GiB
itso-basic00-Pool1	✓ Online		Pool1	10.00 GiB
itso-mirrored00-Pool0-Po...	✓ Online		Pool1	10.00 GiB

Showing 22 volumes | Selecting 1 volume (10.00 GiB)

Figure 6-82 Setting the volume copy in the target storage pool as the primary copy

- Split or delete the volume copy in the source pool, as shown in Figure 6-83.

Name	State	Synchronized	Pool	Capacity
SVC_Volume1	✓ Online		SVC Pool	23.00 GiB
SVC_Volume2	✓ Online		SVC Pool	23.00 GiB
SVC_Volume3	✓ Online		SVC Pool	23.00 GiB
SVC_Volume4	✓ Online		SVC Pool	23.00 GiB
iscsi-cluster-vol	✓ Online			150.00 GiB
iscsi-host-vol0	✓ Online			100.00 GiB
iscsi-host-vol1	✓ Online			100.00 GiB
iscsi-host-vol2	✓ Online			250.00 GiB
iscsi-host-vol3	✓ Online			250.00 GiB
itso-basic00-Pool0	✓ Online			10.00 GiB
Copy 0	✓ Online			10.00 GiB
Copy 1*	✓ Online	Yes	Pool2	10.00 GiB
itso-basic00-Pool1	✓ Online		Pool1	10.00 GiB
itso-mirrored00-Pool0-Po...	✓ Online		Pool1	10.00 GiB

Showing 22 volumes | Selecting 1 volume (10.00 GiB)

Figure 6-83 Deleting the volume copy in the source pool

- Confirm the removal of the volume copy, as shown in Figure 6-84 on page 375.

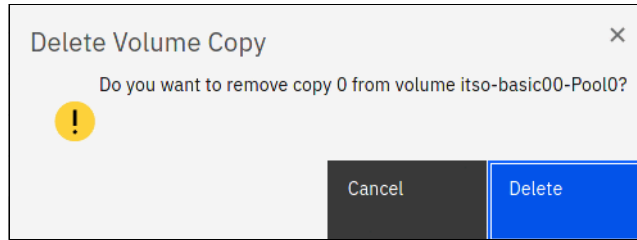


Figure 6-84 Confirming the deletion of a volume copy

- The Volumes view now shows that the volume has a single copy in the target pool, as shown in Figure 6-85.

Name	State	Synchronized	Pool	Capacity
SVC_Volume1	✓ Online		SVC Pool	23.00 GiB
SVC_Volume2	✓ Online		SVC Pool	23.00 GiB
SVC_Volume3	✓ Online		SVC Pool	23.00 GiB
SVC_Volume4	✓ Online		SVC Pool	23.00 GiB
iscsi-cluster-vol	✓ Online		SVC Pool	150.00 GiB
iscsi-host-vol0	✓ Online		SVC Pool	100.00 GiB
iscsi-host-vol1	✓ Online		SVC Pool	100.00 GiB
iscsi-host-vol2	✓ Online		SVC Pool	250.00 GiB
iscsi-host-vol3	✓ Online		SVC Pool	250.00 GiB
itso-basic00-Pool0	✓ Online		Pool2	10.00 GiB
itso-basic00-Pool1	✓ Online		Pool1	10.00 GiB
itso-mirrored00-Pool0-Po...	✓ Online		Pool1	10.00 GiB
Copy 0*	✓ Online	Yes	Pool1	10.00 GiB
Copy 1	✓ Online	Yes	Pool1	10.00 GiB

Showing 22 volumes | Selecting 1 volume (10.00 GiB)

Figure 6-85 Volume copy in the target storage pool

Migrating volumes by using the volume copy feature requires more user interaction, but might be a preferred option for particular use cases. One such example is migrating a volume from a tier 1 storage pool to a lower performance tier 2 storage pool.

First, the volume copy feature can be used to create a copy in the tier 2 pool (steps 1 and 2). All reads are still performed in the tier 1 pool to the primary copy. After the volume copies are synchronized (step 3), all writes are destaged to both pools, but the reads are still done only from the primary copy.

To test the performance of the volume in the new pool, switch the roles of the volume copies to make the new copy the primary (step 4). If the performance is acceptable, the volume copy in tier 1 can be split or deleted. If the tier 2 pool shows unsatisfactory performance, switch the primary volume copy to one that is backed by tier 1 storage.

With this method, you can migrate between storage tiers with a fast and secure back-out option.

6.6 Volume operations by using the CLI

This section describes how to perform various volume configuration and administrative tasks by using the CLI.

For more information about how to set up CLI access, see Appendix C, “Command-line interface setup” on page 925.

6.6.1 Displaying volume information

To display information about all volumes that are defined within the IBM Spectrum Virtualize environment, run the `lsvdisk` command. To display more information about a specific volume, run the command again and provide the volume name or the volume ID as the command parameter, as shown in Example 6-1.

Example 6-1 The `lsvdisk` command

```
IBM_Storwize:ITS0:superuser>lsvdisk -delim ' '
id name IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name capacity type FC_id
FC_name RC_id RC_name vdisk_UID fc_map_count copy_count fast_write_state se_copy_count
RC_change compressed_copy_count parent_mdisk_grp_id parent_mdisk_grp_name formatting
encrypt volume_id volume_name function
0 A_MIRRORED_VOL_1 0 io_grp0 online many many 10.00GB many
6005076400F580049800000000000002 0 2 empty 0 no 0 many many no yes 0 A_MIRRORED_VOL_1
1 COMPRESSED_VOL_1 0 io_grp0 online 1 Poo11 15.00GB striped
6005076400F580049800000000000003 0 1 empty 0 no 1 1 Poo11 no yes 1 COMPRESSED_VOL_1
2 vdisk0 0 io_grp0 online 0 Poo10 10.00GB striped 6005076400F580049800000000000004 0 1
empty 0 no 0 0 Poo10 no yes 2 vdisk0
3 THIN_PROVISION_VOL_1 0 io_grp0 online 0 Poo10 100.00GB striped
6005076400F580049800000000000005 0 1 empty 1 no 0 0 Poo10 no yes 3 THIN_PROVISION_VOL_1
4 COMPRESSED_VOL_2 0 io_grp0 online 1 Poo11 30.00GB striped
6005076400F580049800000000000006 0 1 empty 0 no 1 1 Poo11 no yes 4 COMPRESSED_VOL_2
5 COMPRESS_VOL_3 0 io_grp0 online 1 Poo11 30.00GB striped
6005076400F580049800000000000007 0 1 empty 0 no 1 1 Poo11 no yes 5 COMPRESS_VOL_3
6 MIRRORED_SYNC_RATE_16 0 io_grp0 online many many 10.00GB many
6005076400F580049800000000000008 0 2 empty 0 no 0 many many no yes 6 MIRRORED_SYNC_RATE_16
7 THIN_PROVISION_MIRRORED_VOL 0 io_grp0 online many many 10.00GB many
6005076400F580049800000000000009 0 2 empty 2 no 0 many many no yes 7
THIN_PROVISION_MIRRORED_VOL
8 Tiger 0 io_grp0 online 0 Poo10 10.00GB striped 6005076400F58004980000000000010 0 1
not_empty 0 no 0 0 Poo10 yes yes 8 Tiger
12 vdisk0_restore 0 io_grp0 online 0 Poo10 10.00GB striped
6005076400F58004980000000000000E 0 1 empty 0 no 0 0 Poo10 no yes 12 vdisk0_restore
13 vdisk0_restore1 0 io_grp0 online 0 Poo10 10.00GB striped
6005076400F58004980000000000000F 0 1 empty 0 no 0 0 Poo10 no yes 13 vdisk0_restore1
```

6.6.2 Creating a volume

Running the `mkvdisk` command creates sequential, striped, or image mode volumes. When they are mapped to a host object, these objects are seen as disk drives on which the host can perform I/O operations.

Creating an image mode disk: If you do not specify the `-size` parameter when you create an image mode disk, the entire MDisk capacity is used.

You must know the following information before you start to create the volume:

- ▶ In which storage pool the volume will have its extents.
- ▶ From which I/O group the volume will be accessed.
- ▶ Which IBM Spectrum Virtualize node will be the preferred node for the volume.
- ▶ Size of the volume.
- ▶ Name of the volume.
- ▶ Type of the volume.
- ▶ Whether this volume is to be managed by IBM Easy Tier to optimize its performance.

When you are ready to create your striped volume, run the **mkvdisk** command. The command that is shown in Example 6-2 creates a 10 GB striped volume within the storage pool Pool0 and assigns it to the I/O group io_grp0. Its preferred node is node 1. The volume is given ID 8 by the system.

Example 6-2 The mkvdisk command

```
IBM_Storwize:ITS0:superuser>mkvdisk -mdiskgrp Pool0 -iogrp io_grp0 -size 10 -unit gb -name Tiger
Virtual Disk, id [8], successfully created
```

To verify the results, run the **lsvdisk** command and provide the volume ID as the command parameter, as shown in Example 6-3.

Example 6-3 The lsvdisk command

```
IBM_Storwize:ITS0:superuser>lsvdisk 8
id 8
name Tiger
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name Pool0
capacity 10.00GB
type striped
formatted no
formatting yes
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076400F580049800000000000010
preferred_node_id 2
fast_write_state not_empty
cache readwrite
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
File system
mirror_write_priority latency
RC_change no
compressed_copy_count 0
access_IO_group_count 1
last_access_time
parent_mdisk_grp_id 0
```

```
parent_mdisk_grp_name Pool0
owner_type none
owner_id
owner_name
encrypt yes
volume_id 8
volume_name Tiger
function
throttle_id
throttle_name
IOPs_limit
bandwidth_limit_MB
volume_group_id
volume_group_name
cloud_backup_enabled no
cloud_account_id
cloud_account_name
backup_status off
last_backup_time
restore_status none
backup_grain_size
deduplicated_copy_count 0

copy_id 0
status online
sync yes
auto_delete no
primary yes
mdisk_grp_id 0
mdisk_grp_name Pool0
type striped
mdisk_id
mdisk_name
fast_write_state not_empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status balanced
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 10.00GB
compressed_copy no
uncompressed_used_capacity 10.00GB
parent_mdisk_grp_id 0
parent_mdisk_grp_name Pool0
encrypt yes
deduplicated_copy no
used_capacity_before_reduction
0.00MB
```

The required tasks to create a volume are complete.

6.6.3 Creating a thin-provisioned volume

Example 6-4 shows an example of creating a thin-provisioned volume, which requires the following parameters to be specified:

-rsize	This parameter makes the volume a thin-provisioned volume. If this parameter is missing, the volume is created as standard-provisioned.
-autoexpand	This parameter specifies that thin-provisioned volume copies automatically expand their real capacities by allocating new extents from their storage pool (optional).
-grainsize	This parameter sets the grain size in kilobytes (KB) for a thin-provisioned volume (optional).

Example 6-4 Running the mkvdisk command

```
IBM_Storwize:ITS0:superuser>mkvdisk -mdiskgrp Pool0 -iogrp 0 -vtype striped -size 10 -unit
gb -rsize 50% -autoexpand -grainsize 256
Virtual Disk, id [9], successfully created
```

This command creates a thin-provisioned volume with 10 GB of virtual capacity in a storage pool that is named `Site1_Pool` and is owned by I/O group `io_grp0`. The real capacity is set to automatically expand until the real volume size of 10 GB is reached. The grain size is set to 256 KB, which is the default.

Disk size: When the **-rsize** parameter is used to specify the real physical capacity of a thin-provisioned volume, the following options are available to specify the physical capacity: **disk_size**, **disk_size_percentage**, and **auto**.

Use the **disk_size_percentage** option to define initial real capacity by using a percentage of the disk's virtual capacity that is defined by the **-size** parameter. This option takes as a parameter an integer, or an integer that is immediately followed by the percent (%) symbol.

Use the **disk_size** option to directly specify the real physical capacity by specifying its size in the units that are defined by using the **-unit** parameter (the default unit is MB). The **-rsize** value can be greater than, equal to, or less than the size of the volume.

The **auto** option creates a volume copy that uses the entire size of the MDisk. If you specify the **-rsize auto** option, you must also specify the **-vtype image** option.

An entry of 1 GB uses 1,024 MB.

6.6.4 Creating a volume in image mode

Use an image mode volume to bring a non-virtualized disk (for example, from a pre-virtualization environment) under the control of the IBM Spectrum Virtualize system. After it is managed by the system, you can migrate the volume to the standard MDisk.

When an image mode volume is created, it directly maps to the thus far unmanaged MDisk from which it is created. Therefore, except for a thin-provisioned image mode volume, the volume's LBA x equals MDisk LBA x .

Size: An image mode volume must be at least 512 bytes (the capacity cannot be 0) and always occupies at least one extent.

You must use the `-mdisk` parameter to specify an MDisk that has a mode of unmanaged. The `-fmt disk` parameter cannot be used to create an image mode volume.

Capacity: If you create a mirrored volume from two image mode MDisks without specifying a `-capacity` value, the capacity of the resulting volume is the smaller of the two MDisks. The remaining space on the larger MDisk is inaccessible.

If you do not specify the `-size` parameter when you create an image mode disk, the entire MDisk capacity is used.

Running the `mkvdisk` command to create an image mode volume is shown in Example 6-5.

Example 6-5 The `mkvdisk` (image mode) command

```
IBM_2145:ITSO_CLUSTER:superuser>mkvdisk -mdiskgrp ITSO_Pool1 -iogrp 0 -mdisk mdisk25 -vtype
image -name Image_Volume_A
Virtual Disk, id [6], successfully created
```

As shown in this example, an image mode volume that is named `Image_Volume_A` is created that uses the `mdisk25` MDisk. The MDisk is moved to the storage pool `ITSO_Pool1`, and the volume is owned by the I/O group `io_grp0`.

If you run the `lsvdisk` command, it shows a volume that is named `Image_Volume_A` with the type `image`, as shown in Example 6-6.

Example 6-6 The `lsvdisk` command

```
IBM_2145:ITSO_CLUSTER:superuser>lsvdisk -filtervalue type=image
id name IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name capacity
type FC_id FC_name RC_id RC_name vdisk_UID fc_map_count copy_count
fast_write_state se_copy_count RC_change compressed_copy_count parent_mdisk_grp_id
parent_mdisk_grp_name formatting encrypt volume_id volume_name function
6 Image_Volume_A 0 io_grp0 online 5 ITSO_Pool1 1.00GB
image 6005076801FE80840800000000000021 0 1
empty 0 no 0 5
ITSO_Pool1 no no 6 Image_Volume_A
```

6.6.5 Adding a volume copy

You can add a copy to a volume. If volume copies are defined on different MDisks, the volume remains accessible, even when the MDisk on which one of its copies depends becomes unavailable. You can also create a copy of a volume on a dedicated MDisk by creating an image mode copy of the volume. Although volume copies can increase the availability of data, they are not separate objects.

Volume mirroring can be also used as an alternative method of migrating volumes between storage pools.

To create a copy of a volume, run the `addvdiskcopy` command. This command creates a copy of the chosen volume in the specified storage pool, which changes a non-mirrored volume into a mirrored one.

The following scenario shows how to create a copy of a volume in a different storage pool. As shown in Example 6-7, the volume initially has a single copy with `copy_id 0` that is provisioned in pool `Poo10`.

Example 6-7 The `lsvdisk` command

```
IBM_Storwize:ITS0:superuser>lsvdisk 2
id 2
name vdisk0
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name Poo10
capacity 10.00GB
type striped
formatted yes
formatting no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076400F580049800000000000004
preferred_node_id 2
fast_write_state empty
cache readonly
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
File system
mirror_write_priority latency
RC_change no
compressed_copy_count 0
access_IO_group_count 1
last_access_time
parent_mdisk_grp_id 0
parent_mdisk_grp_name Poo10
owner_type none
owner_id
owner_name
encrypt yes
volume_id 2
volume_name vdisk0
function
throttle_id
throttle_name
IOPs_limit
bandwidth_limit_MB
volume_group_id
volume_group_name
cloud_backup_enabled no
cloud_account_id
cloud_account_name
backup_status off
last_backup_time
restore_status none
backup_grain_size
```

```
deduplicated_copy_count 0

copy_id 0
status online
sync yes
auto_delete no
primary yes
mdisk_grp_id 0
mdisk_grp_name Pool0
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status balanced
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 10.00GB
compressed_copy no
uncompressed_used_capacity 10.00GB
parent_mdisk_grp_id 0
parent_mdisk_grp_name Pool0
encrypt yes
deduplicated_copy no
used_capacity_before_reduction 0.00MB
```

Example 6-8 shows adding the second volume copy by running the **addvdiskcopy** command.

Example 6-8 The addvdiskcopy command

```
IBM_Storwize:ITS0:superuser>addvdiskcopy -mdiskgrp Pool1 -vtype striped -unit gb vdisk0
Vdisk [2] copy [1] successfully created
```

During the synchronization process, you can see the status by running the **lsvdisksyncprogress** command.

As shown in Example 6-9 on page 383, the first time that the status is checked, the synchronization progress is at 48%, and the estimated completion time is 201018232305. The estimated completion time is displayed in the YYMMDDHHMMSS format. In our example, it is 2020, Oct-18 20:23:05. When the command is run again, the progress status is at 100%, and the synchronization is complete.

Example 6-9 Synchronization

```
IBM_Storwize:ITS0:superuser>lsvdisk syncprogress
vdisk_id vdisk_name copy_id progress estimated_completion_time
2        vdisk0     1        0        201018202305
IBM_Storwize:ITS0:superuser>lsvdisk syncprogress
vdisk_id vdisk_name copy_id progress estimated_completion_time
2        vdisk0     1        100
```

As shown in Example 6-10, the new volume copy (copy_id 1) was added and appears in the output of the **lsvdisk** command.

Example 6-10 The *lsvdisk* command

```
IBM_Storwize:ITS0:superuser>lsvdisk vdisk0
id 2
name vdisk0
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id many
mdisk_grp_name many
capacity 10.00GB
type many
formatted yes
formatting no
mdisk_id many
mdisk_name many
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076400F5800498000000000000004
preferred_node_id 2
fast_write_state empty
cache readonly
udid
fc_map_count 0
sync_rate 50
copy_count 2
se_copy_count 0
File system
mirror_write_priority latency
RC_change no
compressed_copy_count 0
access_IO_group_count 1
last_access_time
parent_mdisk_grp_id many
parent_mdisk_grp_name many
owner_type none
owner_id
owner_name
encrypt yes
volume_id 2
volume_name vdisk0
function
throttle_id
throttle_name
IOPs_limit
bandwidth_limit_MB
volume_group_id
```

```

volume_group_name
cloud_backup_enabled no
cloud_account_id
cloud_account_name
backup_status off
last_backup_time
restore_status none
backup_grain_size
deduplicated_copy_count 0

copy_id 0
status online
sync yes
auto_delete no
primary yes
mdisk_grp_id 0
mdisk_grp_name Pool0
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status balanced
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 10.00GB
compressed_copy no
uncompressed_used_capacity 10.00GB
parent_mdisk_grp_id 0
parent_mdisk_grp_name Pool0
encrypt yes
deduplicated_copy no
used_capacity_before_reduction 0.00MB

copy_id 1
status online
sync yes
auto_delete no
primary no
mdisk_grp_id 1
mdisk_grp_name Pool1
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB

```



```
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status balanced
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 10.00GB
compressed_copy no
uncompressed_used_capacity 10.00GB
parent_mdisk_grp_id 1
parent_mdisk_grp_name Pool1
encrypt yes
deduplicated_copy no
used_capacity_before_reduction 0.00MB
```

When adding a volume copy, you can define it with different parameters than the original volume copy. For example, you can create a thin-provisioned copy of a standard-provisioned volume to migrate a thick-provisioned volume to a thin-provisioned volume. The migration can be also done in the opposite direction.

Volume copy mirror parameters: To change the parameters of a volume copy, you must delete the volume copy and redefine it with the new values.

In Example 6-11, the volume name is changed from VOL_NO_MIRROR to VOL_WITH_MIRROR.

Example 6-11 Volume name changes

```
IBM_Storwize:ITS0:superuser>chvdisk -name VOL_WITH_MIRROR VOL_NO_MIRROR
IBM_Storwize:ITS0:superuser>
```

Using the `-autodelete` flag to migrate a volume

This section shows how to run the `addvdiskcopy` command with the `-autodelete` flag set. The `-autodelete` flag causes the primary copy to be deleted after the secondary copy is synchronized.

Example 6-12 shows a shortened `lsvdisk` output for an decompressed volume with a single volume copy.

Example 6-12 An decompressed volume

```
IBM_Storwize:ITS0:superuser>lsvdisk UNCOMPRESSED_VOL
id 9
name UNCOMPRESSED_VOL
IO_group_id 0
IO_group_name io_grp0
status online
...
```

```
copy_id 0
status online
sync yes
auto_delete no
primary yes
...
compressed_copy no
...
```

Example 6-13 adds a compressed copy with the **-autodelete** flag set.

Example 6-13 Compressed copy

```
IBM_Storwize:ITS0:superuser>addvdiskcopy -autodelete -rsize 2 -mdiskgrp 0 -compressed
UNCOMPRESSED_VOL
Vdisk [9] copy [1] successfully created
```

Example 6-14 shows the **lsvdisk** output with another compressed volume (copy 1) and volume copy 0 being set to **auto_delete yes**.

Example 6-14 The lsvdisk command output

```
IBM_Storwize:ITS0:superuser>lsvdisk UNCOMPRESSED_VOL
id 9
name UNCOMPRESSED_VOL
IO_group_id 0
IO_group_name io_grp0
status online
...
compressed_copy_count 2
...

copy_id 0
status online
sync yes
auto_delete yes
primary yes
...

copy_id 1
status online
sync no
auto_delete no
primary no
...
```

When copy 1 is synchronized, copy 0 is deleted. You can monitor the progress of volume copy synchronization by running the **lsvdisksyncprogress** command.

6.6.6 Splitting a mirrored volume

Running the **splitvdiskcopy** command creates an independent volume in the specified I/O group from a volume copy of the specified mirrored volume. In effect, the command changes a volume with two copies into two independent volumes, each with a single copy.

If the copy that you are splitting is not synchronized, you must use the **-force** parameter. If you are attempting to remove the only synchronized copy of the source volume, the command fails. However, you can run the command when either copy of the source volume is offline.

Example 6-15 shows the `splitvdiskcopy` command, which is used to split a mirrored volume. It creates a volume that is named `SPLIT_VOL` from a copy with ID 1 of the volume that is named `VOLUME_WITH_MIRRORED_COPY`.

Example 6-15 Splitting a volume

```
IBM_Storwize:ITS0:superuser>splitvdiskcopy -copy 1 -iogrp 0 -name SPLIT_VOL
VOLUME_WITH_MIRRORED_COPY
Virtual Disk, id [1], successfully created
```

As you can see in Example 6-16, the new volume is created as an independent volume.

Example 6-16 The `lsvdisk` command

```
IBM_Storwize:ITS0:superuser>lsvdisk SPLIT_VOL
id 1
name SPLIT_VOL
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 1
mdisk_grp_name Pool1
capacity 10.00GB
type striped
formatted yes
formatting no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076400F580049800000000000012
preferred_node_id 1
fast_write_state empty
cache readwrite
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
File system
mirror_write_priority latency
RC_change no
compressed_copy_count 0
access_IO_group_count 1
last_access_time
parent_mdisk_grp_id 1
parent_mdisk_grp_name Pool1
owner_type none
owner_id
owner_name
encrypt yes
volume_id 1
volume_name SPLIT_VOL
function
throttle_id
throttle_name
IOPs_limit
bandwidth_limit_MB
```

```
volume_group_id
volume_group_name
cloud_backup_enabled no
cloud_account_id
cloud_account_name
backup_status off
last_backup_time
restore_status none
backup_grain_size
deduplicated_copy_count 0

copy_id 0
status online
sync yes
auto_delete no
primary yes
mdisk_grp_id 1
mdisk_grp_name Pool1
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status balanced
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 10.00GB
compressed_copy no
uncompressed_used_capacity 10.00GB
parent_mdisk_grp_id 1
parent_mdisk_grp_name Pool1
encrypt yes
deduplicated_copy no
used_capacity_before_reduction 0.00MB
```

6.6.7 Modifying a volume

Running the `chvdisk` command modifies a single property of a volume. Only one property can be modified at a time. Therefore, changing the volume name and modifying its I/O group requires two invocations of the command.

Tips: Changing the I/O group with which this volume is associated requires a flush of the cache within the nodes in the current I/O group to ensure that all data is written to disk. I/O must be suspended at the host level before you perform this operation.

If the volume has a mapping to any hosts, it is impossible to move the volume to an I/O group that does not include any of those hosts.

This operation fails if insufficient space exists to allocate bitmaps for a mirrored volume in the target I/O group.

If the **-force** parameter is used and the system cannot destage all write data from the cache, the contents of the volume are corrupted by the loss of the cached data.

If the **-force** parameter is used to move a volume that has out-of-sync copies, a full resynchronization is required.

6.6.8 Deleting a volume

To delete a volume, run the **rmvdisk** command. When this command is run on a managed mode volume, any data on the volume is lost, and the extents that made up this volume are returned to the pool of free extents in the storage pool.

If any RC, IBM FlashCopy, or host mappings still exist for the target of **rmvdisk** command, the delete fails unless the **-force** flag is specified. This flag causes the deletion of the volume and any volume to host mappings and copy mappings.

If the volume is being migrated to image mode, the delete fails unless the **-force** flag is specified. Using the **-force** flag halts the migration and then deletes the volume.

If the command succeeds (without the **-force** flag) for an image mode volume, the write cache data is flushed to the storage before the volume is removed. Therefore, the underlying LU is consistent with the disk state from the point of view of the host that uses the image mode volume (crash-consistent file system). If the **-force** flag is used, consistency is not ensured, that is, the data that the host believes to be written might not be present on the LU.

If any non-destaged data exists in the fast write cache for the target of **rmvdisk** command, the deletion of the volume fails unless the **-force** flag is specified, in which case, any non-destaged data in the fast write cache is deleted.

Example 6-17 shows how to run the **rmvdisk** command to delete a volume from your IBM Spectrum Virtualize configuration.

Example 6-17 The rmvdisk command

```
IBM_2145:ITSO_CLUSTER:superuser>rmvdisk volume_A
```

This command deletes the `volume_A` volume from the IBM Spectrum Virtualize configuration. If the volume is assigned to a host, you must use the **-force** flag to delete the volume, as shown in Example 6-18.

Example 6-18 The rmvdisk -force command

```
IBM_2145:ITSO_CLUSTER:superuser>rmvdisk -force volume_A
```

6.6.9 Volume protection

To prevent active volumes or host mappings from being deleted inadvertently, the system supports a global setting that prevents these objects from being deleted if the system detects recent I/O activity to these objects.

To set the time interval for which the volume must be idle before it can be deleted from the system, run the `chsystem` command. This setting affects the following commands:

- ▶ `rmvdisk`
- ▶ `rmvolume`
- ▶ `rmvdiskcopy`
- ▶ `rmvdiskhostmap`
- ▶ `rmmdiskgrp`
- ▶ `rmhostiogr`
- ▶ `rmhost`
- ▶ `rmhostport`

These commands fail unless the volume was idle for the specified interval or the `-force` parameter was used.

To enable volume protection by setting the required inactivity interval, run the following command:

```
svctask chsystem -vdiskprotectionenabled yes -vdiskprotectiontime 60
```

The `-vdiskprotectionenabled yes` parameter enables volume protection and the `-vdiskprotectiontime` parameter specifies for how long a volume must be inactive (in minutes) before it can be deleted. In this example, volumes can be deleted only if they were inactive for over 60 minutes.

To disable volume protection, run the following command:

```
svctask chsystem -vdiskprotectionenabled no
```

6.6.10 Expanding a volume

Expanding a volume presents a larger capacity disk to your OS. Although this expansion can be easily performed by using IBM Spectrum Virtualize, you must ensure that your OS supports expansion before this function is used.

Assuming that your OS supports expansion, you can run the `expandvdisksize` command to increase the capacity of a volume, as shown in Example 6-19.

Example 6-19 The `expandvdisksize` command

```
IBM_2145:ITS0_CLUSTER:superuser>expandvdisksize -size 5 -unit gb volume_C
```

This command expands the `volume_C` volume (which was 35 GB) by another 5 GB to give it a total size of 40 GB.

To expand a thin-provisioned volume, you can use the `-rsize` option, as shown in Example 6-20 on page 391. This command changes the real size of the `volume_B` volume to a real capacity of 55 GB. The capacity of the volume is unchanged.

Example 6-20 The `lsdisk` command

```
IBM_Storwize:ITS0:superuser>lsdisk volume_B
id 26
capacity 100.00GB
type striped
.
.
copy_id 0
status online
used_capacity 0.41MB
real_capacity 50.02GB
free_capacity 50.02GB
overallocation 199
autoexpand on
warning 80
grainsize 32
se_copy yes
```

```
IBM_Storwize:ITS0:superuser>expandvdisksize -rsize 5 -unit gb volume_B
IBM_Storwize:ITS0:superuser>lsdisk volume_B
id 26
name volume_B
capacity 100.00GB
type striped
.
.
copy_id 0
status online
used_capacity 0.41MB
real_capacity 55.02GB
free_capacity 55.02GB
overallocation 181
autoexpand on
warning 80
grainsize 32
se_copy yes
```

Important: If a volume is expanded, its type becomes striped, even if it was previously sequential or in image mode.

If not enough extents are available to expand your volume to the specified size, the following error message is displayed:

```
CMMVC5860E The action failed because there were not enough extents in the
storage pool.
```

6.6.11 HyperSwap volume modification with CLI

The following new CLI commands for administering volumes were released in IBM Spectrum Virtualize V7.6. However, the GUI uses the new commands only for HyperSwap volume creation (`mkvolume`) and deletion (`rmvolume`):

- ▶ `mkvolume`
- ▶ `mkimagevolume`
- ▶ `addvolumecopy`
- ▶ `rmvolumecopy`
- ▶ `rmvolume`

In addition, the `lsvdisk` output shows more fields: `volume_id`, `volume_name`, and `function`, which help to identify the individual VDisks that make up a HyperSwap volume. This information is used by the GUI to provide views that reflect the client's view of the *HyperSwap* volume and its site-dependent copies, as opposed to the "low-level" VDisks and VDisk Change Volumes.

The following individual commands are related to HyperSwap:

► **mkvolume**

Creates an empty volume by using storage from a storage pool. The type of volume that is created is determined by the system topology and the number of storage pools that is specified. The volume is always formatted (zeroed). The `mkvolume` command can be used to create the following objects:

- Basic volume: Any topology
- Mirrored volume: Standard topology
- Stretched volume: Stretched topology
- HyperSwap volume: HyperSwap topology

► **rmvolume**

Removes a volume. For a HyperSwap volume, this process includes deleting the active-active relationship and the change volumes.

The `-force` parameter that is used by `rmvdisk` is replaced by a set of override parameters, one for each operation-stopping condition, which makes it clearer to the user exactly what protection they are bypassing.

► **mkimagevolume**

Creates an image mode volume. This command can be used to import a volume, which preserves data. It can be implemented as a separate command to provide greater differentiation between the action of creating an empty volume and creating a volume by importing data on an MDisk.

► **addvolumecopy**

Adds a copy to a volume. The new copy is always synchronized from the existing copy. For stretched and HyperSwap topology systems, this command creates a HA volume. This command can be used to create the following volume types:

- Mirrored volume: Standard topology
- Stretched volume: Stretched topology
- HyperSwap volume: HyperSwap topology

► **rmvolumecopy**

Removes a copy of a volume. This command leaves the volume intact. It also converts a Mirrored, Stretched, or HyperSwap volume to a basic volume. For a HyperSwap volume, this command includes deleting the active-active relationship and the change volumes.

This command enables a copy to be identified by its site.

The `-force` parameter that is used by `rmvdiskcopy` is replaced by a set of override parameters, one for each operation-stopping condition, making it clearer to the user exactly what protection they are bypassing.

6.6.12 Mapping a volume to a host

To map a volume to a host, run the `mkvdiskhostmap` command. This mapping makes the volume available to the host for I/O operations. A host can perform I/O operations only on volumes that are mapped to it.

When the host bus adapter (HBA) on the host scans for devices that are attached to it, the HBA discovers all of the volumes that are mapped to its FC ports and their SCSI identifiers (SCSI LUN IDs).

For example, the first disk that is found is generally SCSI LUN 1. You can control the order in which the HBA discovers volumes by assigning the SCSI LUN ID as required. If you do not specify a SCSI LUN ID when mapping a volume to the host, the storage system automatically assigns the next available SCSI LUN ID based on any mappings that exist with that host.

Note: The SCSI-3 standard requires LUN 0 to exist on every SCSI target. This LUN must implement a number of standard commands, including Report LUNs. However, this LUN does not have to provide any storage capacity.

Example 6-21 shows how to map volumes `volume_B` and `volume_C` to the defined host `Almaden` by running the `mkvdiskhostmap` command.

Example 6-21 The `mkvdiskhostmap` command

```
IBM_Storwize:ITS0:superuser>mkvdiskhostmap -host Almaden volume_B
Virtual Disk to Host map, id [0], successfully created
IBM_Storwize:ITS0:superuser>mkvdiskhostmap -host Almaden volume_C
Virtual Disk to Host map, id [1], successfully created
```

Example 6-22 shows the output of the `lshostvdiskmap` command, which shows that the volumes are mapped to the host.

Example 6-22 The `lshostvdiskmap -delim` command

```
IBM_2145:ITS0_CLUSTER:superuser>lshostvdiskmap -delim :
id:name:SCSI_id:vdisk_id:vdisk_name:vdisk_UID
2:Almaden:0:26:volume_B:6005076801AF813F1000000000000020
2:Almaden:1:27:volume_C:6005076801AF813F1000000000000021
```

Assigning a specific LUN ID to a volume: The optional `-scsi scsi_lun_id` parameter can help assign a specific LUN ID to a volume that is to be associated with a host. The default (if nothing is specified) is to assign the next available ID based on the current volume that is mapped to the host.

Certain HBA device drivers stop when they find a gap in the sequence of SCSI LUN IDs, as shown in the following examples:

- ▶ Volume 1 is mapped to Host 1 with SCSI LUN ID 1.
- ▶ Volume 2 is mapped to Host 1 with SCSI LUN ID 2.
- ▶ Volume 3 is mapped to Host 1 with SCSI LUN ID 4.

When the device driver scans the HBA, it might stop after discovering volumes 1 and 2 because no SCSI LUN is mapped with ID 3.

Important: Ensure that the SCSI LUN ID allocation is contiguous.

If you are using host clusters, run the `mkvolumehostclustermap` command to map a volume to a host cluster instead (see Example 6-23).

Example 6-23 The `mkvolumehostclustermap` command

```
BM_Storwize:ITS0:superuser>mkvolumehostclustermap -hostcluster vmware_cluster
UNCOMPRESSED_VOL
Volume to Host Cluster map, id [0], successfully created
```

6.6.13 Listing volumes that are mapped to the host

To show the volumes that are mapped to the specific host, run the `lshostvdiskmap` command, as shown in Example 6-24.

Example 6-24 The `lshostvdiskmap` command

```
IBM_2145:ITS0_CLUSTER:superuser>lshostvdiskmap -delim , Siam
id,name,SCSI_id,vdisk_id,vdisk_name,wwpn,vdisk_UID
3,Siam,0,0,volume_A,210000E08B18FF8A,60050768018301BF280000000000000C
```

In the output of the command, you can see that only one volume (`volume_A`) is mapped to the host `Siam`. The volume is mapped with SCSI LUN ID 0.

If no hostname is specified by the `lshostvdiskmap` command, it returns all defined host-to-volume mappings.

Specifying the flag before the hostname: Although the `-delim` flag normally comes at the end of the command string, you must specify this flag before the hostname in this case. Otherwise, it returns the following message:

```
CMMVC6070E An invalid or duplicated parameter, unaccompanied argument, or
incorrect argument sequence has been detected. Ensure that the input is as per
the help.
```

You can also run the `lshostclustervolumemap` command to show the volumes that are mapped to a specific host cluster, as shown in Example 6-25.

Example 6-25 The `lshostclustervolumemap` command

```
IBM_Storwize:ITS0:superuser>lshostclustervolumemap
id name          SCSI_id volume_id volume_name      volume_UID
IO_group_id IO_group_name
0 vmware_cluster 0          9          UNCOMPRESSED_VOL 6005076400F580049800000000000011 0
io_grp0
```

6.6.14 Listing hosts that are mapped to the volume

To identify the hosts to which a specific volume was mapped, run the `lsvdiskhostmap` command, as shown in Example 6-26.

Example 6-26 The `lsvdiskhostmap` command

```
IBM_2145:ITS0_CLUSTER:superuser>lsvdiskhostmap -delim , volume_B
id,name,SCSI_id,host_id,host_name,vdisk_UID
26,volume_B,0,2,Almaden,6005076801AF813F1000000000000020
```

This command shows the list of hosts to which the volume `volume_B` is mapped.

Specifying the `-delim` flag: Although the optional `-delim` flag normally comes at the end of the command string, you must specify this flag before the volume name in this case. Otherwise, the command does not return any data.

6.6.15 Deleting a volume to host mapping

Deleting a volume mapping does not affect the volume. Instead, it removes only the host's ability to use the volume. To unmap a volume from a host, run the `rmvdiskhostmap` command, as shown in Example 6-27.

Example 6-27 The `rmvdiskhostmap` command

```
IBM_2145:ITSO_CLUSTER:superuser>rmvdiskhostmap -host Tiger volume_D
```

This command unmaps the volume that is called `volume_D` from the host that is called `Tiger`.

You can also run the `rmvolumehostclustermap` command to delete a volume mapping from a host cluster, as shown in Example 6-28.

Example 6-28 The `rmvolumehostclustermap` command

```
IBM_Storwize:ITSO:superuser>rmvolumehostclustermap -hostcluster vmware_cluster  
UNCOMPRESSED_VOL
```

This command unmaps the volume that is called `UNCOMPRESSED_VOL` from the host cluster that is called `vmware_cluster`.

Note: Removing a volume that is mapped to the host makes the volume unavailable for I/O operations. Ensure that the host is prepared for this situation before removing a volume mapping.

6.6.16 Migrating a volume

You might want to migrate volumes from one set of MDisks to another set of MDisks to decommission an old disk subsystem to better distribute load across your virtualized environment, or to migrate data into the IBM Spectrum Virtualize environment by using image mode. For more information about migration, see Chapter 8, "Storage migration" on page 485.

Important: After migration is started, it continues until it completes unless it is stopped or suspended by an error condition or the volume that is being migrated is deleted.

As you can see from the parameters that are shown in Example 6-29, before you can migrate your volume, you must determine the name of the volume that you want to migrate and the name of the storage pool to which you want to migrate it. To list the names of volumes and storage pools, run the `lsvdisk` and `lsmdiskgrp` commands.

The command that is shown in Example 6-29 moves `volume_C` to the storage pool that is named `STGPool_DS5000-1`.

Example 6-29 The migratevdisk command

```
IBM_2145:ITSO_CLUSTER:superuser>migratevdisk -mdiskgrp STGPool_DS5000-1 -vdisk volume_C
```

Note: If insufficient extents are available within your target storage pool, you receive an error message. Ensure that the source MDisk group and target MDisk group have the same extent size.

You can use the optional `threads` parameter to control priority of the migration process. The default is 4, which is the highest priority setting. However, if you want the process to take a lower priority over other types of I/O, you can specify 3, 2, or 1.

You can run the `lsmigrate` command at any time to see the status of the migration process, as shown in Example 6-30.

Example 6-30 The lsmigrate command

```
IBM_2145:ITSO_CLUSTER:superuser>lsmigrate
migrate_type MDisk_Group_Migration
progress 0
migrate_source_vdisk_index 27
migrate_target_mdisk_grp 2
max_thread_count 4
migrate_source_vdisk_copy_id 0
```

```
IBM_2145:ITSO_CLUSTER:superuser>lsmigrate
migrate_type MDisk_Group_Migration
progress 76
migrate_source_vdisk_index 27
migrate_target_mdisk_grp 2
max_thread_count 4
migrate_source_vdisk_copy_id 0
```

Progress: The progress is shown in terms of percentage complete. If no output is displayed when running the command, all volume migrations are finished.

6.6.17 Migrating a fully managed volume to an image mode volume

Migrating a fully managed volume to an image mode volume enables the IBM Spectrum Virtualize system to be removed from the data path. This feature might be useful when the IBM Spectrum Virtualize system is used as a data mover.

To migrate a fully managed volume to an image mode volume, the following rules apply:

- ▶ Cloud snapshots must not be enabled on the source volume.
- ▶ The destination MDisk must be greater than or equal to the size of the volume.
- ▶ The MDisk that is specified as the target must be in an unmanaged state.

- ▶ Regardless of the mode in which the volume starts, it is reported as a managed mode during the migration.
- ▶ If the migration is interrupted by a system recovery or cache problem, the migration resumes after the recovery completes.

Example 6-31 shows running the **migratetoimage** command to migrate the data from `volume_A` onto `mdisk10`, and to put the MDisk `mdisk10` into the `STGPool_IMAGE` storage pool.

*Example 6-31 The **migratetoimage** command*

```
IBM_2145:ITSO_CLUSTER:superuser>migratetoimage -vdisk volume_A -mdisk mdisk10 -mdiskgrp
STGPool_IMAGE
```

6.6.18 Shrinking a volume

The **shrinkvdisksize** command reduces the capacity that is allocated to the particular volume by the specified amount. You cannot shrink the real size of a thin-provisioned volume to less than its used size. All capacities (including changes) must be in multiples of 512 bytes. An entire extent is reserved even if it is only partially used. The default capacity unit is MB.

You can use this command to shrink the physical capacity of a volume or to reduce the virtual capacity of a thin-provisioned volume without altering the physical capacity that is assigned to the volume. To change the volume size, use the following parameters:

- ▶ For a standard-provisioned volume, use the **-size** parameter.
- ▶ For a thin-provisioned volume's real capacity, use the **-rsize** parameter.
- ▶ For a thin-provisioned volume's virtual capacity, use the **-size** parameter.

When the virtual capacity of a thin-provisioned volume is changed, the warning threshold is automatically scaled.

If the volume contains data that is being used, do not shrink the volume without backing up the data first. The system reduces the capacity of the volume by removing arbitrarily chosen extents, or extents from those sets that are allocated to the volume. You cannot control which extents are removed. Therefore, you cannot assume that it is unused space that is removed.

Image mode volumes cannot be reduced in size. To reduce their size, first they must be migrated to fully managed mode.

Before the **shrinkvdisksize** command is used on a mirrored volume, all copies of the volume must be synchronized.

Important: Consider the following guidelines when you are shrinking a disk:

- ▶ If the volume contains data or host-accessible metadata (for example, an empty physical volume of an LVM), do not shrink the disk.
- ▶ This command can shrink a FlashCopy target volume to the same capacity as the source.
- ▶ Before you shrink a volume, validate that the volume is not mapped to any host objects.
- ▶ You can determine the exact capacity of the source or master volume by running the **svcinfo lsvdisk -bytes vdiskname** command.

Shrink the volume by the required amount by running the following command:

```
shrinkvdisksize -size disk_size -unit b | kb | mb | gb | tb | pb vdisk_name |  
vdisk_id.
```

Example 6-32 shows running the **shrinkvdisksize** command to reduce the size of volume `volume_D` from a total size of 80 GB by 44 GB to the new total size of 36 GB.

Example 6-32 The shrinkvdisksize command

```
IBM_2145:ITS0_CLUSTER:superuser>shrinkvdisksize -size 44 -unit gb volume_D
```

6.6.19 Listing volumes that use MDisks

To identify which volumes use space on the specified MDisk, run the **lsmdiskmember** command. Example 6-33 displays a list of volume IDs of all volume copies that use `mdisk8`. To correlate the IDs that are displayed in this output to volume names, run the **lsvdisk** command.

Example 6-33 The lsmdiskmember command

```
IBM_2145:ITS0_CLUSTER:superuser>lsmdiskmember mdisk8  
id copy_id  
24 0  
27 0
```

6.6.20 Listing MDisks that are used by the volume

To list MDisks that supply space that is used by the specified volume, run the **lsvdiskmember** command. Example 6-34 lists the MDisk IDs of all MDisks that are used by the volume with ID 0.

Example 6-34 The lsvdiskmember command

```
IBM_2145:ITS0_CLUSTER:superuser>lsvdiskmember 0  
id  
4  
5  
6  
7
```

If you want to know more about these MDisks, you can run the **lsmdisk** command and provide the MDisk ID that is listed in the output of the **lsvdiskmember** command as a parameter.

6.6.21 Listing volumes that are defined in the storage pool

To list volumes that are defined in the specified storage pool, run the **lsvdisk -filtervalue** command. Example 6-35 shows how to use the **lsvdisk -filtervalue** command to list all volumes that are defined in the storage pool that is named `Poo10`.

Example 6-35 The lsvdisk -filtervalue command: Volumes in the pool

```
IBM_Storwize:ITS0:superuser>lsvdisk -filtervalue mdisk_grp_name=Poo10 -delim ,  
id,name,IO_group_id,IO_group_name,status,mdisk_grp_id,mdisk_grp_name,capacity,type,FC_id,FC  
_name,RC_id,RC_name,vdisk_UID,fc_map_count,copy_count,fast_write_state,se_copy_count,RC_cha  
nge,compressed_copy_count,parent_mdisk_grp_id,parent_mdisk_grp_name,formatting,encrypt,volu  
me_id,volume_name,function
```

```

0,A_MIRRORED_VOL_1,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F58004980000000000
00002,0,1,empty,0,no,0,0,Pool0,no,yes,0,A_MIRRORED_VOL_1,
2,VOLUME_WITH_MIRRORED_COPY,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F5800498
00000000000004,0,1,empty,0,no,0,0,Pool0,no,yes,2,VOLUME_WITH_MIRRORED_COPY,
3,THIN_PROVISION_VOL_1,0,io_grp0,online,0,Pool0,100.00GB,striped,,,,,6005076400F58004980000
0000000005,0,1,empty,1,no,0,0,Pool0,no,yes,3,THIN_PROVISION_VOL_1,
6,MIRRORED_SYNC_RATE_16,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F58004980000
0000000008,0,1,empty,0,no,0,0,Pool0,no,yes,6,MIRRORED_SYNC_RATE_16,
7,THIN_PROVISION_MIRRORED_VOL,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F58004
9800000000000009,0,1,empty,1,no,0,0,Pool0,no,yes,7,THIN_PROVISION_MIRRORED_VOL,
8,Tiger,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F580049800000000000010,0,1,e
mpty,0,no,0,0,Pool0,no,yes,8,Tiger,
9,UNCOMPRESSED_VOL,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F58004980000000000
00011,0,1,empty,0,no,1,0,Pool0,no,yes,9,UNCOMPRESSED_VOL,
12,vdisk0_restore,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F5800498000000000000
000E,0,1,empty,0,no,0,0,Pool0,no,yes,12,vdisk0_restore,
13,vdisk0_restore1,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F5800498000000000000
0000F,0,1,empty,0,no,0,0,Pool0,no,yes,13,vdisk0_restore1,

```

6.6.22 Listing storage pools in which a volume has its extents

To show to which storage pool a specific volume belongs, run the `lsvdisk` command, as shown in Example 6-36.

Example 6-36 The `lsvdisk` command: Storage pool ID and name

```

IBM_Storwize:ITS0:superuser>lsvdisk 0
id 0
name A_MIRRORED_VOL_1
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name Pool0
capacity 10.00GB
type striped
formatted yes
formatting no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076400F5800498000000000000002
preferred_node_id 2
fast_write_state empty
cache readwrite
udid 4660
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
File system
mirror_write_priority latency
RC_change no
compressed_copy_count 0
access_IO_group_count 1
last_access_time

```

```
parent_mdisk_grp_id 0
parent_mdisk_grp_name Pool0
owner_type none
owner_id
owner_name
encrypt yes
volume_id 0
volume_name A_MIRRORED_VOL_1
function
throttle_id 1
throttle_name throttle1
IOPs_limit 233
bandwidth_limit_MB 122
volume_group_id
volume_group_name
cloud_backup_enabled no
cloud_account_id
cloud_account_name
backup_status off
last_backup_time
restore_status none
backup_grain_size
deduplicated_copy_count 0

copy_id 0
status online
sync yes
auto_delete no
primary yes
mdisk_grp_id 0
mdisk_grp_name Pool0
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status measured
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 10.00GB
compressed_copy no
uncompressed_used_capacity 10.00GB
parent_mdisk_grp_id 0
parent_mdisk_grp_name Pool0
encrypt yes
deduplicated_copy no
used_capacity_before_reduction0.00MB
```

To learn more about these storage pools, run the `lsmdiskgrp` command as described in Chapter 5, “Storage pools” on page 237.

6.6.23 Tracing a volume from a host back to its physical disks

In some cases, you might need to verify exactly which physical disks are used to store the data of a volume. This information is not directly available to the host, but it might be obtained by using a sequence of queries.

Before you trace a volume, you must unequivocally map a logical device that is seen by the host to a volume that is presented by the storage system. The best volume characteristic for this purpose is the volume ID. This ID is available to the OS in the Vendor Specified Identifier field of page 0x80 or 0x83 (vital product data (VPD)), which the storage device sends in response to SCSI **INQUIRY** command from the host.

In practice, the ID can be obtained from the multipath driver in the OS. After you know the volume ID, you can use it to identify the physical location of data.

Note: For sequential and image mode volumes, a volume copy is mapped to exactly one MDisk. This configuration usually is not used for striped volumes unless the volume size is lesser than the extent sizes. Therefore, a single striped volume uses multiple MDisks in a typical case.

For example, on a Linux host running a native multipath driver, you can use the output of the command `multipath -ll` to find the volume ID, as shown in Example 6-37.

Example 6-37 Volume ID returned by the multipath -ll command

```
mpath1 (360050768018301BF280000000000004) IBM,2145
[size=2.0G][features=0][hwandler=0]
\_ round-robin 0 [prio=200][ enabled]
\_ 4:0:0:1 sdd 8:48 [active][ready]
\_ 5:0:0:1 sdt 65:48 [active][ready]
\_ round-robin 0 [prio=40][ active]
\_ 4:0:2:1 sdak 66:64 [active][ready]
\_ 5:0:2:1 sda1 66:80 [active][ready]
```

Note: the volume ID that is shown in the output of `multipath -ll` is generated by the Linux `scsi_id`. For systems that provide the VPD by using page 0x83 (such as IBM Spectrum Virtualize devices), the ID that is obtained from the VPD page is prefixed by the number 3, which is the Network Address Authority (NAA) type identifier. Therefore, the volume NAA identifier (that is, the volume ID that is obtained by running the SCSI **INQUIRY** command) starts at the second displayed digit. In Example 6-37, the volume ID starts with digit 6.

After you know the volume ID, complete the following steps:

1. To list volumes that are mapped to the host, run the **lshostvdiskmap** command. Example 6-38 lists the volumes that are mapped to host Almaden.

Example 6-38 The lshostvdiskmap command

```
IBM_2145:ITSO_CLUSTER:superuser>lshostvdiskmap -delim , Almaden
id,name,SCSI_id,vdisk_id,vdisk_name,vdisk_UID
2,Almaden,0,26,volume_B,60050768018301BF2800000000000005
2,Almaden,1,27,volume_A,60050768018301BF28000000000000004
2,Almaden,2,28,volume_C,60050768018301BF28000000000000006
```

Look for the VDisk unique identifier (UID) that matches volume UID that was identified and note the volume name (or ID) for a volume with this UID.

2. To list the MDisks that contain extents that are allocated to the specified volume, run the **lsvdiskmember vdiskname** command, as shown in Example 6-39.

Example 6-39 The lsvdiskmember command

```
IBM_2145:ITSO_CLUSTER:superuser>lsvdiskmember volume_A
id
0
1
2
3
4
10
11
13
15
16
17
```

3. For each of the MDisk IDs that were obtained in step 2, run the **lsmdisk mdiskID** command to discover the MDisk controller and LUN information. Example 6-40 shows the output for mdisk0. The output displays the back-end storage controller name and the controller LUN ID to help you to track back to a LUN within the disk subsystem.

Example 6-40 The lsmdisk command

```
IBM_2145:ITSO_CLUSTER:superuser>lsmdisk 0
id 0
name mdisk0
status online
mode managed
mdisk_grp_id 0
mdisk_grp_name STGPool_DS3500-1
capacity 128.0GB
quorum_index 1
block_size 512
controller_name ITSO-DS3500
ctrl_type 4
ctrl_WWNN 20080080E51B09E8
controller_id 2
path_count 4
max_path_count 4
ctrl_LUN_# 0000000000000000
UID 60080e50001b0b62000007b04e731e4d00000000000000000000000000000000
preferred_WWPN 20580080E51B09E8
```

```
active_WPN 20580080E51B09E8
fast_write_state empty
raid_status
raid_level
redundancy
strip_size
spare_goal
spare_protection_min
balanced
tier generic_hdd
```

You can identify the back-end storage that is presenting the LUN by using the value of the `controller_name` field that was returned for the MDisk.

On the back-end storage, you can identify which physical disks make up the LUN that was presented to the Storage Virtualize system by using the volume ID that is displayed in the `UID` field.



Hosts

This chapter describes the host configuration procedures that are required to attach supported hosts to the storage systems and documents the available ways of host attachment, including Non-Volatile Memory Express (NVMe) over Fabric (NVMe-oF), Fibre Channel Small Computer System Interface (SCSI) (FC-SCSI), serial-attached SCSI (SAS), and internet Small Computer Systems Interface (iSCSI).

This chapter also explains host clustering representation in the storage system and N_Port ID Virtualization (NPIV) support for a host-to-storage system communication.

This chapter includes the following topics:

- ▶ 7.1, “Host attachment overview” on page 406
- ▶ 7.2, “Host objects overview” on page 407
- ▶ 7.3, “NVMe over Fibre Channel” on page 408
- ▶ 7.4, “N_Port ID Virtualization support” on page 410
- ▶ 7.5, “Hosts operations by using the GUI” on page 418
- ▶ 7.6, “Performing hosts operations by using CLI” on page 461
- ▶ 7.7, “Host attachment practical examples” on page 471

7.1 Host attachment overview

IBM FlashSystem family storage systems (further referred as storage systems) support a wide range of host types (by IBM and other vendors). With flexible internal storage capabilities, and starting with IBM FlashSystem 5100, the ability to virtualize external storages by IBM and other vendors, it is possible to consolidate storage capacities in an open systems environment in a common storage pool or pools with centralized management, control, and configuration. Thus, storage resources can be used more efficiently and in a precisely controlled manner from a central point in the storage area network (SAN).

The ability to consolidate storage for attached open systems hosts provides the following benefits:

- ▶ Easier storage management.
- ▶ Increased utilization rate of the installed storage capacity.
- ▶ Advanced Copy Services functions that are offered by storage systems, which are independent from host and external storage (if external storage virtualization is used) vendors.
- ▶ Only a multipath driver is required for attached hosts. You do not need a specialized storage vendor-specific driver.

Hosts can be connected to the storage systems by using any of the following protocols:

- ▶ Fibre Channel Protocol (FCP)
- ▶ Fibre Channel over Ethernet (FCoE)
- ▶ iSCSI
- ▶ SAS
- ▶ iSCSI Extensions for Remote Direct Memory Access (RDMA) (iSER)

Starting with IBM FlashSystem 5100 and IBM Spectrum Virtualize V8.2 and later, NVMeoF using Fibre Channel (FC-NVMe) or NVMeoF or Fibre Channel (FC)) is supported.

Hosts that connect to the storage systems by using fabric switches that use the FC, FCoE, or FC-NVMe protocols must be zoned correctly, as described in Chapter 2, “Planning” on page 71. N-Port ID Virtualization support (supported from IBM Spectrum Virtualize V7.7 onwards) plays a central role because it is required for FC-NVMe connectivity.

Hosts that connect to the systems by using the iSCSI protocol must be configured correctly, as described in Chapter 2, “Planning” on page 71.

Note: Certain host operating systems (OSs) can be directly connected to the IBM FlashSystem storage systems without FC fabric switches. For more information, see the [IBM System Storage Interoperation Center \(SSIC\)](#).

For correct volumes representation and access through several access paths from the host side, you must install a multipathing driver in the connected host. IBM FlashSystem family storage systems are supported by several OS-native multipathing drivers. Multipathing drivers also serve the following purposes:

- ▶ Protection from fabric paths failures, including port failures on IBM Spectrum Virtualize system nodes.
- ▶ Protection from a host bus adapter (HBA) failure (if two HBAs are used).

- ▶ Protection from fabric failures if the host is connected through two HBAs to two separate fabrics.
- ▶ To provide load balancing across the host HBAs.

For more information about the various host multipath driver solutions that are native to OSs and versions that are supported, and support an IBM FlashSystem system, see the [SSIC](#).

For more information about how to attach various supported host OSs to the systems, see the “Host Attachment” section of [IBM Documentation](#).

If your host OS is not mentioned in the SSIC, you can ask your IBM representative to submit a special request for support by contacting your IBM Business Partner, account manager, or IBM Support.

7.2 Host objects overview

To provide storage capacity to a specific host, it is necessary to first present the host to the storage system as a host object and then present the chosen capacity in the form of a volume to that host.

On the storage system, a host is represented by host objects, which must be configured by using the GUI or command-line interface (CLI) and contain the necessary credentials for host-to-storage communications. A real world host receives access to the storage capacity through a host object that is configured on a storage system and a storage space that is mapped to the host object in the form of a volume.

IBM Spectrum Virtualize V 8.4 supports configuring the following host objects:

- ▶ Host
- ▶ Host cluster (Supported since Version 7.7.1 and later)

Each *host object* has attributes that should be configured and provide the status of the host as it is visible by storage system.

A *host cluster* object groups multiple hosts that are working as a cluster. A host cluster object is treated as a single entity so that multiple hosts can access the same volumes with a single shared mapping.

Volumes that are mapped to a host cluster are assigned to all members of the host cluster with the same SCSI ID.

A typical use case for a host cluster object is to group all the hosts that are the members of a host OS-based cluster, such as IBM PowerHA®, and Microsoft Cluster Server, and present it as a single entity sharing access to the volumes and with improved and simplified control. The following commands deal with host and host cluster objects:

- ▶ Commands that provide information about defined hosts and host clusters (start with **l s** (list)):
 - **lshostcluster**
 - **lshostclustermember**
 - **lshostclustervolumemap**
 - **lshost**
 - **lshostiogrp**

- **lshostiplogin**
- **lsciscsiauth**
- ▶ Commands that define or configure a host object on a storage system (start with **mk** (make)):
 - **mkhost** (Can be used to define an individual host and put the host into a host cluster on creation, which defines a host cluster.)
 - **mkhostcluster** (Defines a cluster host object, and during the creation of the object, the host list can be provided to add specific hosts to the cluster object.)
- ▶ Commands to remove or delete defined host objects from a storage system configuration (start with **rm** (remove)):
 - **rmhostclustermember**
 - **rmhostcluster**
 - **rmvolumehostclustermap**
 - **rmhost**
 - **rmhostiogrps**
 - **rmhostport**
- ▶ Commands to change defined host objects:
 - **chhostcluster** (Changes the name, type, or site of a host cluster object that is part of a host cluster.)
 - **chhost** (Changes the name or type.)
- ▶ Commands to manipulate (add) host to a host cluster or to an I/O group, or add a port to an existing host object:
 - **addhostclustermember**
 - **addhostiogrps**
 - **addhostport**

For more information about each command, see [IBM Documentation](#) and select **Command-line interface** → **Host commands**. The instructions to perform basic tasks on hosts and host clusters are provided in 7.6, “Performing hosts operations by using CLI” on page 461.

7.3 NVMe over Fibre Channel

IBM FlashSystem 5100, IBM FlashSystem 7200/H, IBM FlashSystem 9200, and IBM FlashSystem9200R running IBM Spectrum Virtualize V8.2 or later can attach NVMe hosts by using NVMe-oF. NVMe-oF in IBM Spectrum Virtualize V8.2.1 uses the FCP (FC-NVMe) as its underlying transport, which puts the data transfer in control of the target and transfers data direct from host memory like RDMA. In addition, with FC-NVMe, a host can send commands and data together (first burst), which eliminates the first data that is “read” by the target and provides better performance at distances. Since Version 8.3.1.0, IBM FlashSystem 5100 also provides full support for NVMe-oF, along with many feature improvements on all systems.

The limitation in Version 8.2 of being able to attach only one SCSI or NVMe per I/O group was removed in Version 8.3. You can now run SCSI and NVMe in parallel. The limit of 512 host objects per I/O group remains in place. When you run SCSI and NVMe in parallel, there are limits for each protocol, as shown in Table 7-1.

Table 7-1 Defined host object limits per I/O group

SCSI host objects	NVMe host objects	Total host objects
496	16	512 ^a

a. IBM Storwize V5100 systems have a maximum of 256 hosts.

Note: Although the specifications that are shown in Table 7-1 are the maximum amount of each type of host attachment that you can have, if you do not have the maximum NVMe Host Objects defined, then these objects do not reduce the number of total host objects that you can have. For example, if you had 10 NVMe host objects that are defined, you might have up to 502 SCSI host objects defined. The only hard limit is 16 NVMe host objects per I/O group. You should be diligent when planning a parallel SCSI and NVMe deployment as described in Chapter 2, “Planning” on page 71 because it can be *resource-intensive*, especially with large deployments (many hosts). This situation occurs because NVMe is more sensitive to delays than SCSI. It is a best practice to check the [Configuration Limits page](#) for your product and 7.3, “NVMe over Fibre Channel” on page 408.

Ensure that you select the correct product.

To avoid any potential interoperability problems, a volume can be mapped to a host only by using one protocol. IBM FlashCopy, volume mirroring, Remote Copy (RC), and Data Reduction Pools (DRPs) are all supported by NVMe-oF. Starting with Version 8.3.1, there is support for stretched cluster configurations, and starting with Version 8.4, HyperSwap is supported for NVMe-oF attached hosts.

Note: In Version 8.4, HyperSwap and Non-disruptive Volume Move (NDVM) support is available for FC-NVMe hosts because IBM Spectrum Virtualize is using Asymmetric Namespace Access (ANA) reporting. The following features are available for FC-NVMe attached hosts:

- ▶ Sites can be defined to facilitate awareness of HyperSwap volume site properties.
- ▶ It is possible to map HyperSwap volumes by using multiple I/O groups on the same and different sites.
- ▶ Hosts can use I/O through a non-optimized path even if the primary site is available.
- ▶ The ability to fail over to the secondary site if the primary site is down.

For more information about NVMe, see *IBM Storage and the NVM Express Revolution*, REDP-5437.

7.4 N_Port ID Virtualization support

The operation model for all IBM Spectrum Virtualize products is based on a two-way active/active node model.

Note: IBM Flash System 5010 can have only one I/O group and does not support clustering with other IBM Flash System 5010 control enclosures.

This model has a pair of distinct control modules that are known as *nodes* that share active/active access for any specific volume within the same I/O group. Each of these nodes has their own FC worldwide node name (WWNN). Ports from each node's network adapter or HBA have a set of worldwide port names (WWPNs) that are presented to the fabric.

These ports are used for following purposes:

- ▶ Internode communication (communication between storage system nodes, for example, to share the cache and exchange the status of the nodes)
- ▶ Back-end controllers communication (except for the IBM Flash System 5010 and IBM Flash System 5030)
- ▶ Host communications

Traditionally, if one node fails or is removed for some reason, the paths that are presented for volumes from that node to a host go offline. In this case, it is up to the native OS multipathing software to fail over from using both sets of WWPNs to only those nodes and paths that remain online.

Although this scenario is the one that multipathing software is designed for, occasionally it can be problematic, particularly if paths are not seen as coming back online for some reason and also relies on correct configuration and implementation of specific multipath driver.

Starting with Version 7.7, the implementation of NPIV mode is available on the storage systems.

When NPIV mode is enabled on the storage systems, target ports (also known as host attach ports) dedicated *only* to host communication become available, which efficiently separates internode communication and host I/O. You can move host attachment ports between the nodes in the same I/O group transparently for the host and make sure that host-dedicated ports do not come online until they are ready to service I/O, which improves host behavior when storage nodes leave or join the storage cluster for any reason. If one node in I/O group is offline, moving host attach ports to the online node in the same I/O group masks the path failures that are caused by the offline node from hosts, and the multipathing driver does not need to perform any path recovery.

When NPIV is enabled on the storage system, each physical WWPN reports up to four virtual WWPNs, as listed in Table 7-2.

Table 7-2 IBM Spectrum Virtualize NPIV ports

NPIV port	Port description
Primary port	The WWPN that communicates with back-end storage. It can be used for node to node traffic (local or remote).
Primary SCSI host attach port	The WWPN that communicates with hosts. It is a target port only. It is the primary port, so it is based on this local node's WWNN.

NPIV port	Port description
Failover SCSI host attach port	A standby WWPN that communicates with hosts that is brought online only if the partner node within the I/O group goes offline. This WWPN is the same as the primary host attach WWPN of the partner node.
Primary NVMe host attach port	The WWPN that communicates with hosts. It is a target port only. This WWPN is the primary port, so it is based on this local node's WWNN.
Failover NVMe host attach port	A standby WWPN that communicates with hosts that is brought online only if the partner node within the I/O group goes offline. This WWPN is the same as the primary host attach WWPN of the partner node.

Figure 7-1 shows the five WWPNs that are associated with a port when NPIV is enabled.

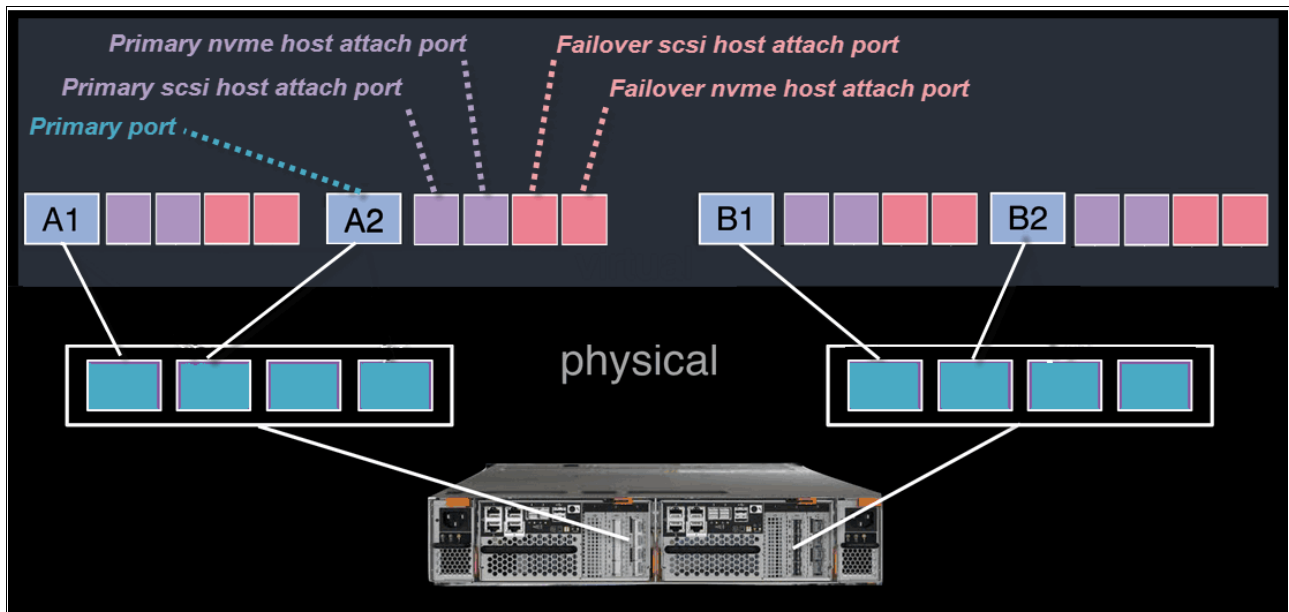


Figure 7-1 Allocation of NPIV virtual WWPN ports per physical port

The *failover host attach port* is not active. Figure 7-2 shows what happens when the partner node fails. After the node failure, the *failover* host attach ports on the remaining node become active and take on the WWPN of the failed node's *primary* host attach port.

Note: Figure 7-2 shows only two ports per node in detail, but the same situation applies for all physical ports. The effect is the same for NVMe ports because they use the same NPIV structure, but with the NVMe topology instead of regular SCSI.

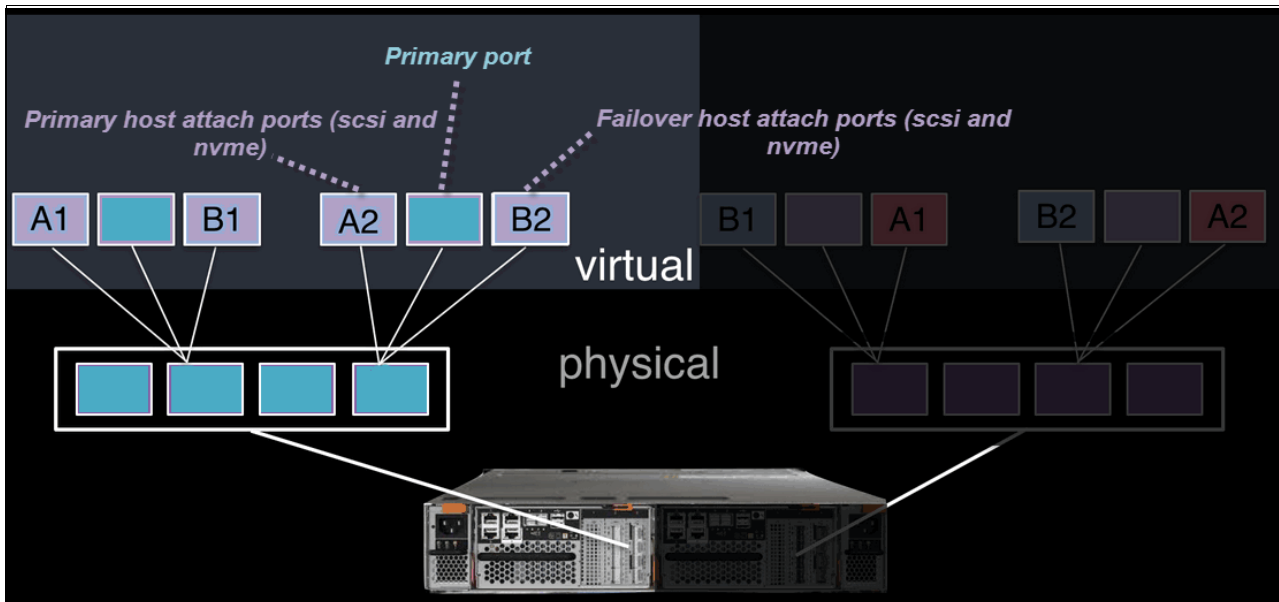


Figure 7-2 Allocation of NPIV virtual WWPN ports per physical port after a node failure

Since Version 7.7 and later, this process happens automatically when NPIV is enabled at a system level in the storage systems. This failover happens only between the two nodes in the same I/O group.

The NPIV mode states are:

- ▶ Disabled: Virtualized host attach ports (NPIV target ports) may not be used for I/O. Only physical ports can be used for I/O. Volumes are presented to the hosts through physical ports only.
- ▶ Transitional: Both virtualized host attach ports (NPIV target ports) and physical ports can be used for I/O. Volumes are presented to the hosts through both physical and NPIV target ports.
- ▶ Enabled: Only virtualized host attach ports (NPIV target ports) may service host I/O. Volumes are presented to the hosts only through NPIV target ports.

A transitional state enables migration of hosts from previous non-NPIV enabled systems to enabled NPIV systems, which enables a transition state as hosts are rezoned to the *primary* host attach WWPNs.

The process to enable NPIV on a new system is slightly different than on an existing system. For more information, see [IBM Documentation](#).

Note: NPIV is supported for FC-based communication only. It is not supported for the FCoE or iSCSI protocols.

7.4.1 NPIV prerequisites

Consider the following key points for NPIV enablement:

- ▶ The system must be running Version 7.7 or later.
- ▶ A Version 7.7 or later system with NPIV enabled as back-end storage for a system that is earlier than Version 7.7 is not supported.
- ▶ Both nodes within an I/O group should have identical hardware to enable failover to work as expected.
- ▶ The FC switches to which the system ports are attached must support NPIV and have this feature enabled.
- ▶ Node connectivity should be done according to “Zoning requirements for N_Port ID virtualization” at [IBM Documentation](#). Both nodes in one I/O group should have their equivalent ports connected to their equivalent fabrics (switch), for example, port 1 of node1 should be on the same fabric as port 1 of the node2.

7.4.2 Verifying the NPIV mode state for a new system installation

New systems with IBM Spectrum Virtualize V7.7 or later are NPIV-enabled by default. You verify whether NPIV is enabled by completing the following steps, and if necessary turn on NPIV (see step 2):

1. Run the `lsiogrp` command to list the I/O groups that are present in the system, as shown in Example 7-1.

Example 7-1 Listing the I/O groups in the system

```
IBM_IBM FlashSystem:FS9100:superuser>lsiogrp
id name          node_count vdisk_count host_count site_id site_name
0  io_grp0        2          10          0          0
1  io_grp1        0          0           0          0
2  io_grp2        0          0           0          0
3  io_grp3        0          0           0          0
4  recovery_io_grp 0          0           0          0
```

Example 7-1 shows that in our example that we have one full I/O group with ID 0, two nodes in it, and 10 virtual disks (VDisks). The other I/O groups are empty.

2. Run the `lsiogrp <id> | grep fctargetportmode` command for the specific I/O group ID to display the `fctargetportmode` setting. If this setting is enabled, as shown in Example 7-2, NPIV host target port mode is enabled. If NPIV mode is disabled, the `fctargetportmode` parameter reports as disabled.

Example 7-2 Checking the NPIV mode by viewing the fctargetportmode field

```
IBM_IBM FlashSystem:FS9100:superuser>lsiogrp 0|grep fctargetportmode
fctargetportmode enabled
```

- The virtual WWPNs can be listed by running the `lstargetportfc` command, as shown in Example 7-3. The `host_io_permitted` and `virtualized` columns should be `yes`, meaning that the WWPNs in those lines are a primary host attach port and should be used when zoning the hosts to the system.

Example 7-3 Listing the virtual WWPNs

```
IBM_IBM FlashSystem:FS9100:superuser>lstargetportfc
```

id	WWPN	WWNN	port_id	..	host_io_permitted	virtualized	protocol
1	500507681011024F	500507681000024F	1		no	no	scsi
2	500507681015024F	500507681000024F	1		yes	yes	scsi
3	500507681019024F	500507681000024F	1		yes	yes	nvme
...							
10	500507681014024F	500507681000024F	4		no	no	scsi
11	500507681018024F	500507681000024F	4		yes	yes	scsi
12	50050768101C024F	500507681000024F	4		yes	yes	nvme

- Now, it is necessary to configure zones for host-to-storage communication by using the primary host attach ports (virtual WWPNs), as shown in the output of the command. The virtualized ports are marked **bold** in Example 7-3.
- If the status of `fctargetportmode` is disabled and this is a new installation, run the `chiogrp` command to set the NPIV mode to the transitional state and then to enabled, as shown in Example 7-4.

Example 7-4 Changing the NPIV mode to enabled

```
IBM_IBM FlashSystem:FS9100:superuser>chiogrp -fctargetportmode transitional 0
IBM_IBM FlashSystem:FS9100:superuser>chiogrp -fctargetportmode enabled 0
```

7.4.3 Enabling NPIV on an existing system

When IBM Spectrum Virtualize systems that are running code earlier than Version 7.7.1 are upgraded to the latest version of the code, the NPIV feature is not turned on by default because it requires changes to host-to-storage zoning, so enabling it by default on a configured system might cause the loss of access to the volumes for all hosts.

To enable NPIV mode on a storage system, it is necessary to complete the following actions:

- Audit your SAN fabric layout and zoning rules because NPIV usage has strict requirements. Ensure that equivalent ports are on *the same fabric* and *in the same zone*.
- Check the path count between your hosts and the IBM Spectrum Virtualize system to ensure that the number of paths is half of the usual supported maximum.
- Run the `lstargetportfc` command to discover the primary host attach WWPNs (virtual WWPNs), as shown in **bold** in Example 7-5. Those virtualized ports are not enabled for host I/O communication yet (see the `host_io_permitted` column).

Example 7-5 Running the lstargetportfc command to get the primary host WWPNs (virtual WWPNs)

```
IBM_IBM FlashSystem:FS9100:superuser>lstargetportfc
```

id	WWPN	WWNN	port_id	owning_node_id	current_node_id	nportid	host_io_permitted	virtualized	protocol
1	500507680140A288	500507680100A288	1	1	1	010A00	yes	no	scsi
2	500507680142A288	500507680100A288	1	1	1	000000	no	yes	scsi
3	500507680144A288	500507680100A288	1	1	1	000000	no	yes	nvme
4	500507680130A288	500507680100A288	2	1	1	010400	yes	no	scsi
5	500507680132A288	500507680100A288	2	1	1	000000	no	yes	scsi
6	500507680134A288	500507680100A288	2	1	1	000000	no	yes	nvme
7	500507680110A288	500507680100A288	3	1	1	010500	yes	no	scsi

8	500507680112A288	500507680100A288	3	1		000000	no	yes	scsi
9	500507680114A288	500507680100A288	3	1		000000	no	yes	nvme
10	500507680120A288	500507680100A288	4	1	1	010A00	yes	no	scsi
11	500507680122A288	500507680100A288	4	1		000000	no	yes	scsi
12	500507680124A288	500507680100A288	4	1		000000	no	yes	nvme
...									
58	500507680C140009	500507680C000009	4	2	2	010900	yes	no	scsi
59	500507680C180009	500507680C000009	4	2		000000	no	yes	scsi
60	500507680C1C0009	500507680C000009	4	2		000000	no	yes	nvme

- To enable virtualized ports for host I/O communication and still keep access to the hosts that are using hardware-defined ports (not in bold in Example 7-5 on page 414), you must enable transitional mode for NPIV on the system (see Example 7-6).

Example 7-6 NPIV in transitional mode

```
IBM_IBM FlashSystem:FS9100:superuser>chlogrp -fctargetportmode transitional 0
IBM_IBM FlashSystem:FS9100:superuser>lsiogrp 0 |grep fctargetportmode
fctargetportmode transitional
```

Alternatively, to activate NPIV in transitional mode by using the GUI, go to the GUI and select **Settings** → **System** → **I/O Groups**, as shown in Figure 7-3.

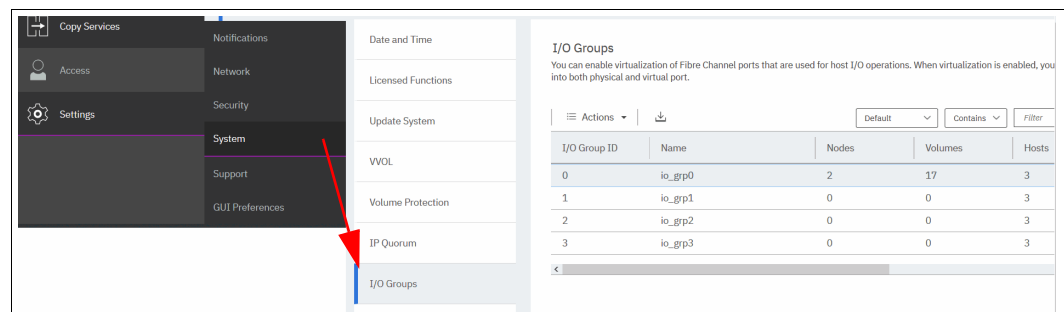


Figure 7-3 I/O Groups menu

Then, check the current NPIV setting by viewing the NPIV column, which shows “disabled” if NPIV is not enabled. Select the I/O group on which you want to enable NPIV and select **Actions** → **Change NPIV Settings**, as shown in Figure 7-4.

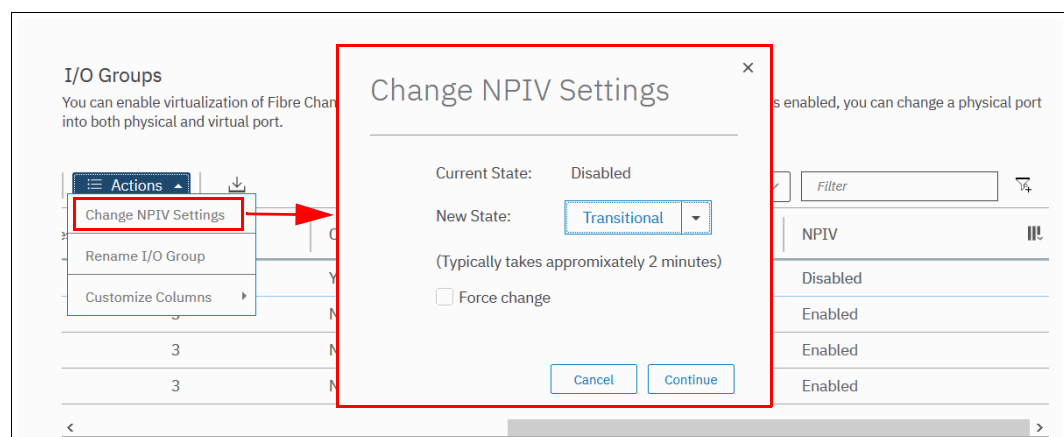


Figure 7-4 Change NPIV Settings

After you select **Continue**, NPIV is enabled in Transitional Mode.

- Ensure that the primary host attach WWPNs (virtual WWPNs) now allow host traffic, as shown in **bold** in Example 7-7.

Example 7-7 Host attach WWPNs (virtual WWPNs) permitting host traffic

```
IBM_IBM FlashSystem:FS9100:superuser>lstargetportfc
```

id	WWPN	WWNN	port_id	owning_node_id	current_node_id	nportid	host_io_permitted	virtualized	protocol
1	500507680140A288	500507680100A288	1	1	1	010A00	yes	no	scsi
2	500507680142A288	500507680100A288	1	1	1	010A02	yes	yes	scsi
3	500507680144A288	500507680100A288	1	1	1	010A01	yes	yes	nvme
4	500507680130A288	500507680100A288	2	1	1	010400	yes	no	scsi
5	500507680132A288	500507680100A288	2	1	1	010401	yes	yes	scsi
6	500507680134A288	500507680100A288	2	1	1	010402	yes	yes	nvme
7	500507680110A288	500507680100A288	3	1	1	010500	yes	no	scsi
8	500507680112A288	500507680100A288	3	1	1	010501	yes	yes	scsi
9	500507680114A288	500507680100A288	3	1	1	010502	yes	yes	nvme
...									
58	500507680C140009	500507680C000009	4	2	2	010900	yes	no	scsi
59	500507680C180009	500507680C000009	4	2	2	010901	yes	yes	scsi
60	500507680C1C0009	500507680C000009	4	2	2	010902	yes	yes	nvme

- Add the primary host attach ports (virtual WWPNs) to the host zones, but do not remove the IBM FlashSystem WWPNs that are in the zones. Example 7-8 shows a host zone to the primary port WWPNs of the IBM FlashSystem nodes.

Example 7-8 Legacy host zone

```
zone: WINDOWS_HOST_01_IBM_FS9100
      10:00:00:05:1e:0f:81:cc
      50:05:07:68:01:40:A2:88
      50:05:07:68:0C:11:00:09
```

Example 7-9 shows that we added the primary host attach ports (virtual WWPNs) to our example host zone so that we can change the host without disrupting its availability.

Example 7-9 Transitional host zone (added host attach ports are in bold)

```
zone: WINDOWS_HOST_01_IBM_FS9100
      10:00:00:05:1e:0f:81:cc
      50:05:07:68:01:40:A2:88
      50:05:07:68:0C:11:00:09
      50:05:07:68:01:42:A2:88
      50:05:07:68:0C:15:00:09
```

- With the transitional zoning active in the fabrics, ensure that the host is using the new NPIV ports for host I/O. Example 7-10 on page 417 shows the pathing for our host before and after adding the new host attach ports by using the old IBM Subsystem Device Driver (SDD) Device Specific Module (SDDDSM) multipathing driver. The select count increases on the new paths and stops on the old paths.

Note: SDDDSM, which is a multipathing driver, is not recommended or supported. The Recommended multipathing driver for the Microsoft Windows platform is Microsoft Device Specific Module (MSDSM).

Example 7-10 Host device pathing: Before and after

```
C:\Program Files\IBM\SDDDSM>datapath query device
```

```
Total Devices : 1
```

```
DEV#: 0 DEVICE NAME: Disk3 Part0 TYPE: 2145 POLICY: OPTIMIZED  
SERIAL: 60050764008680083800000000000000 LUN SIZE: 20.0GB
```

```
=====
```

Path#	Adapter/Hard Disk	State	Mode	Select	Errors
0 *	Scsi Port2 Bus0/Disk1 Part0	OPEN	NORMAL	3991778	0
1 *	Scsi Port2 Bus0/Disk1 Part0	OPEN	NORMAL	416214	0
2 *	Scsi Port3 Bus0/Disk1 Part0	OPEN	NORMAL	22255	0
3 *	Scsi Port3 Bus0/Disk1 Part0	OPEN	NORMAL	372785	0

```
C:\Program Files\IBM\SDDDSM>datapath query device
```

```
Total Devices : 1
```

```
DEV#: 0 DEVICE NAME: Disk3 Part0 TYPE: 2145 POLICY: OPTIMIZED  
SERIAL: 60050764008680083800000000000000 LUN SIZE: 20.0GB
```

```
=====
```

Path#	Adapter/Hard Disk	State	Mode	Select	Errors
0 *	Scsi Port2 Bus0/Disk1 Part0	OPEN	NORMAL	3991778	2
1 *	Scsi Port2 Bus0/Disk1 Part0	OPEN	NORMAL	416214	1
2 *	Scsi Port3 Bus0/Disk1 Part0	OPEN	NORMAL	22255	0
3 *	Scsi Port3 Bus0/Disk1 Part0	OPEN	NORMAL	372785	2
4 *	Scsi Port2 Bus0/Disk1 Part0	OPEN	NORMAL	22219	0
5	Scsi Port2 Bus0/Disk1 Part0	OPEN	NORMAL	95109	0
6 *	Scsi Port3 Bus0/Disk1 Part0	OPEN	NORMAL	2	0
7	Scsi Port3 Bus0/Disk1 Part0	OPEN	NORMAL	91838	0

Note: Consider the following points:

- ▶ You can verify that you are logged in to the NPIV ports by running the `lsfabric -host host_id_or_name` command. If I/O activity is occurring, each host has at least one line in the command output that corresponds to a host port and shows `active` in the activity field:
 - Hosts where no I/O happen in the past 5 minutes show `inactive` for any login.
 - Hosts that do not adhere to preferred paths *might still* be processing I/O to primary ports.
- ▶ Depending on the host OS, rescanning of storage might be required on some hosts to recognize more paths that are now provided by using primary host attach ports (virtual WWPNs).

8. After all hosts are rezoned and the pathing is validated, change the system NPIV to enabled mode by running the command that is shown in Example 7-11.

Example 7-11 Enabling the NPIV

```
IBM_IBM FlashSystem:FS9100:superuser>chiogrp -fctargetportmode enabled 0
```

Alternatively, to enable NPIV by using the GUI, go to the I/O Groups window, as shown in Step 4, select the I/O group, click **Actions**, and then click **Change NPIV Settings**. The NPIV Settings window opens, as shown in Figure 7-5.

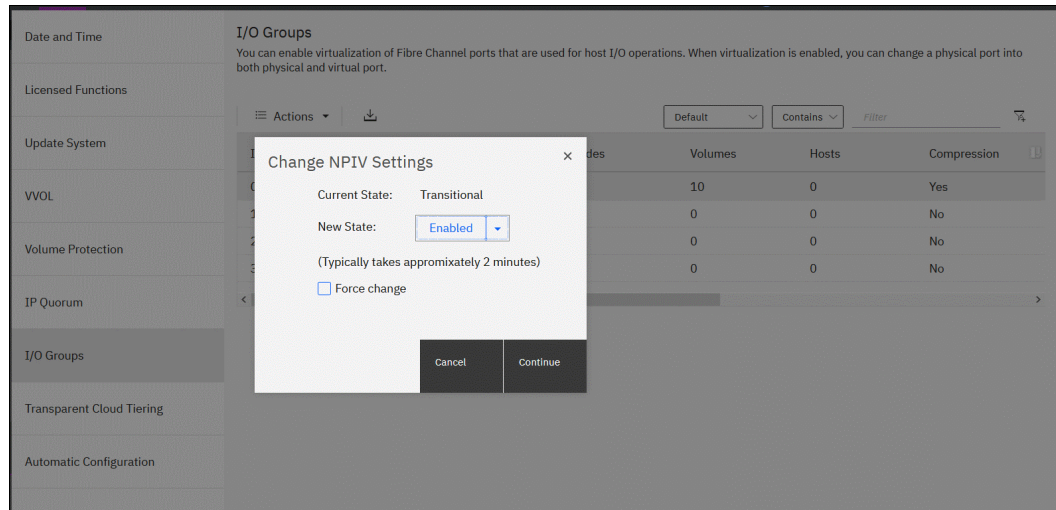


Figure 7-5 NPIV Settings: Enabling

Click **Continue**, and NPIV is enabled on the I/O group.

NPIV is enabled on the system, and the hosts use the virtualized WWPNs for I/O. To complete the NPIV implementation, you can modify the host zones to remove the old primary attach port WWPNs. Example 7-12 shows the final zone with the host HBA and the IBM FlashSystem virtual WWPNs.

Example 7-12 Final host zone

```
zone: WINDOWS_HOST_01_IBM_FS9100
      10:00:00:05:1e:0f:81:cc
      50:05:07:68:01:42:A2:88
      50:05:07:68:0C:15:00:09
```

Note: If any hosts are still configured to use the physical ports on the system, the system prevents you from changing `fctargetportmode` from `transitional` to `enabled` and shows the following error:

```
CMMVC8019E Task could interrupt I/O and force flag not set.
```

7.5 Hosts operations by using the GUI

This section describes performing the following host operations by using the IBM Spectrum Virtualize GUI:

- ▶ Creating hosts
- ▶ Advanced host administration
- ▶ Adding and deleting host ports
- ▶ Host mappings overview

7.5.1 Creating hosts

This section describes how to create FC, iSCSI, and NVMe connected host objects by using a GUI. It is assumed that hosts are prepared for attachment and that the host WWPNs, iSCSI initiator names, or NVMe Qualified Names (NQN) are known. For more information, see the “Host Attachment” section of [IBM Documentation](#).

To create a host, complete the following steps:

1. Open the host configuration window by clicking **Hosts** (see Figure 7-6).

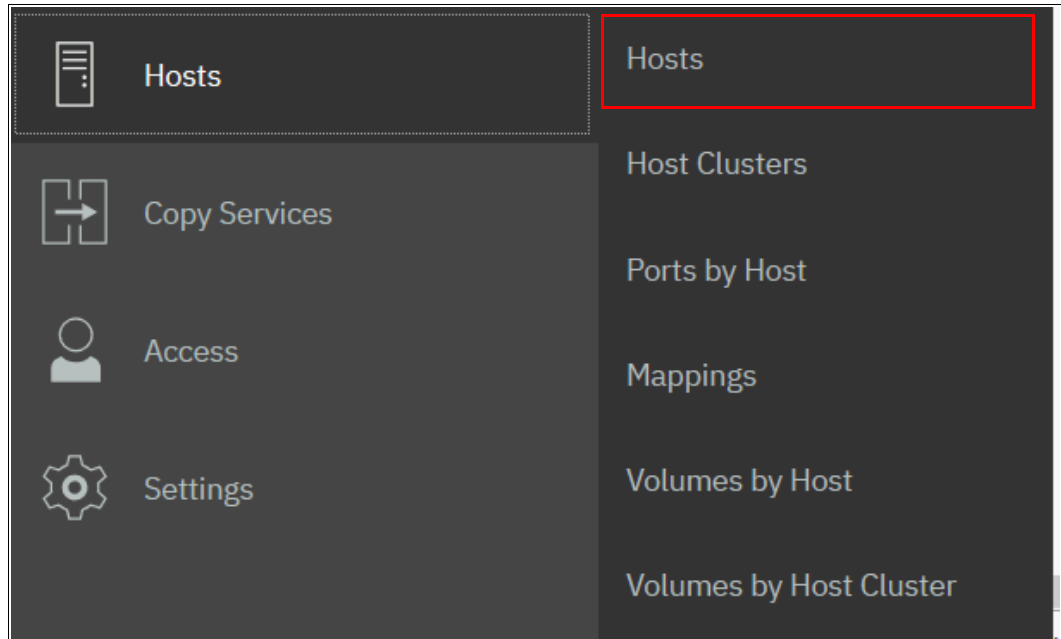


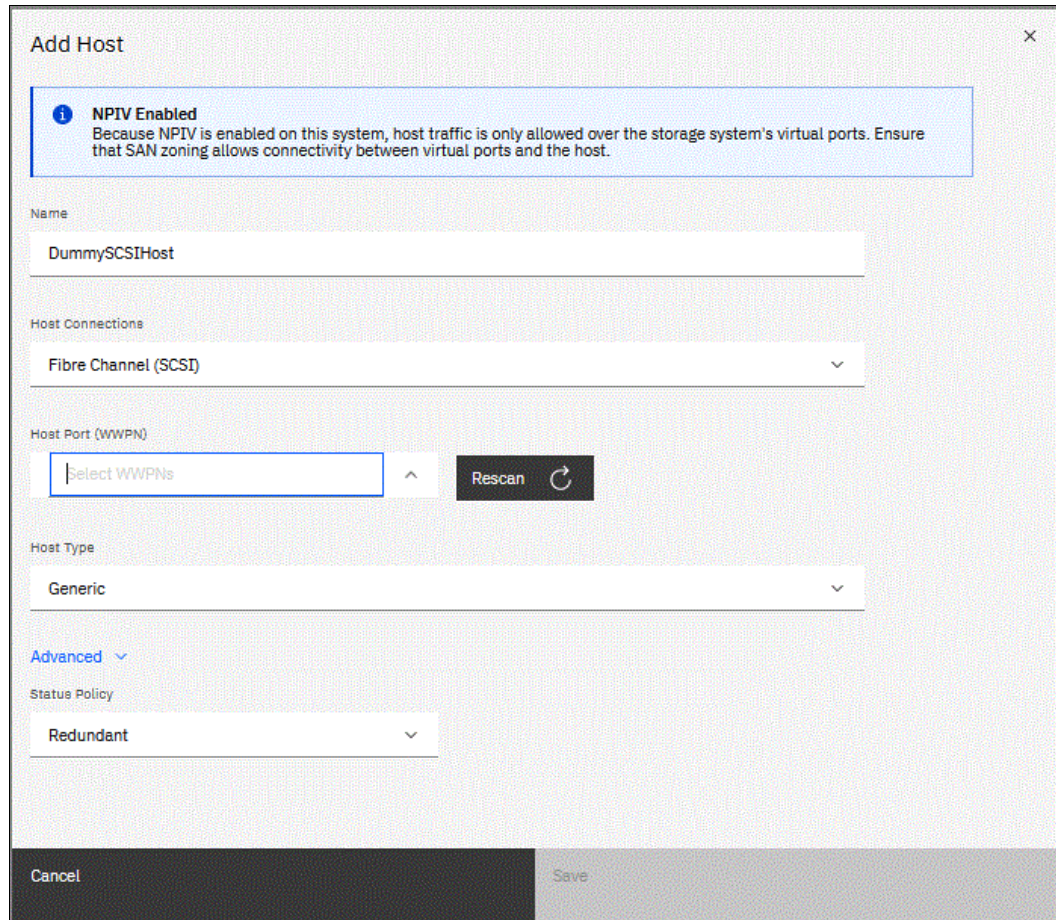
Figure 7-6 Opening the Hosts window

2. To create a host, click **Add Host**. If you want to create an FC host, go to “Creating Fibre Channel host objects”. To create an iSCSI host, go to “Creating iSCSI host objects” on page 429. To create an NVMe host, go to “Creating NVMe host objects” on page 430.

Creating Fibre Channel host objects

To create FC hosts, complete the following steps:

1. Select **Fibre Channel** in the Host Connections list. The FC configuration fields appear (see Figure 7-7).



The screenshot shows the 'Add Host' configuration window. At the top right is a close button (X). Below the title bar is a blue information box with an 'i' icon and the text: 'NPIV Enabled. Because NPIV is enabled on this system, host traffic is only allowed over the storage system's virtual ports. Ensure that SAN zoning allows connectivity between virtual ports and the host.' Below this is the 'Name' field with the value 'DummySCSIHost'. The 'Host Connections' dropdown menu is set to 'Fibre Channel (SCSI)'. The 'Host Port (WWPN)' section has a text input field containing 'select WWPNs', an up arrow button, and a 'Rescan' button with a circular arrow icon. The 'Host Type' dropdown menu is set to 'Generic'. Below this is an 'Advanced' section with a dropdown arrow, containing a 'Status Policy' dropdown menu set to 'Redundant'. At the bottom are 'Cancel' and 'Save' buttons.

Figure 7-7 Fibre Channel host configuration view

2. Enter a hostname and click the **Host Port** menu to get a list of all discovered WWPNs (see Figure 7-8).

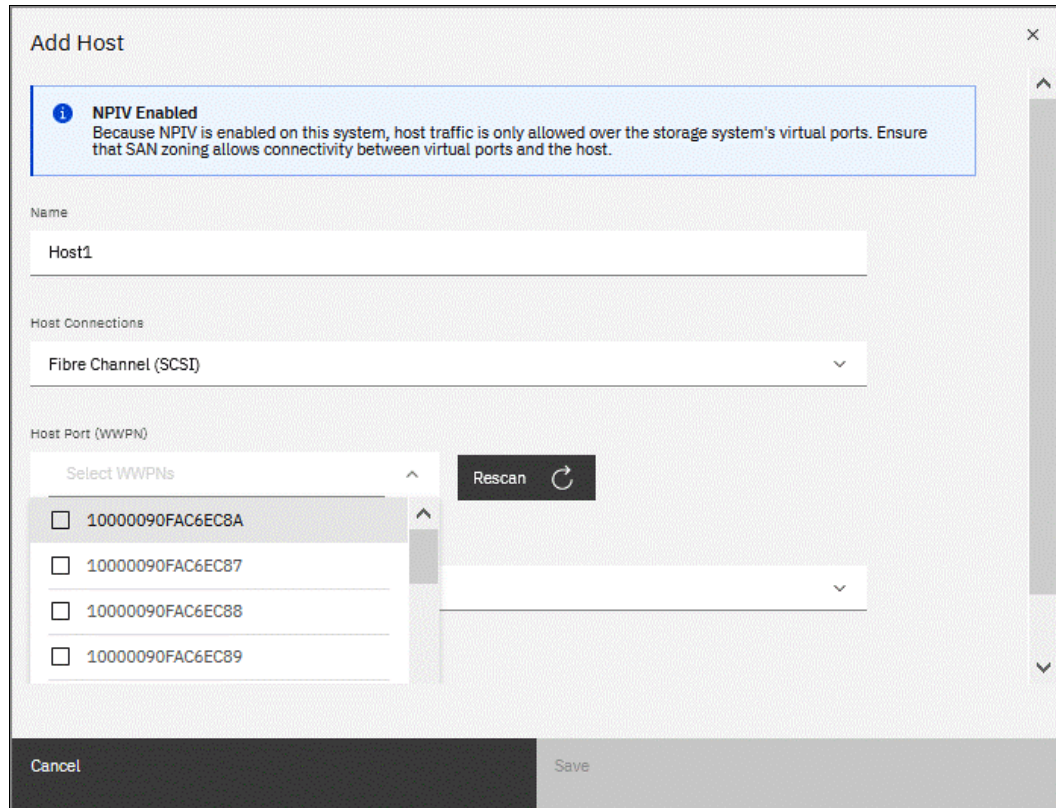


Figure 7-8 Selecting the host WWPNs

3. Select one or more WWPNs for your host from the list. The host WWPNs should be visible on IBM FlashSystem storage if the hosts were zoned and presented to the storage system correctly. If the hosts do not appear in the list, scan for new paths as required on the respective OS and click the **Rescan** icon next to the WWPN box. If they still do not appear, check the SAN zoning, make sure that hosts are connected and running, and then repeat the scanning.

Creating offline hosts: If you want to create hosts that are offline or not connected at the moment, it is also possible to enter the WWPNs manually. Enter them into the **Host Ports** field to add them to the list.

4. If you want to add more ports to your Host, choose several WWPNs from the list to add all ports that belong to the specific host.

5. If you are creating a Hewlett-Packard UNIX (HP-UX) or Target Port Group Support (TPGS) host, click the **Host type** list (see Figure 7-9). Select your host type. If your specific host type is not listed, select **Generic**.

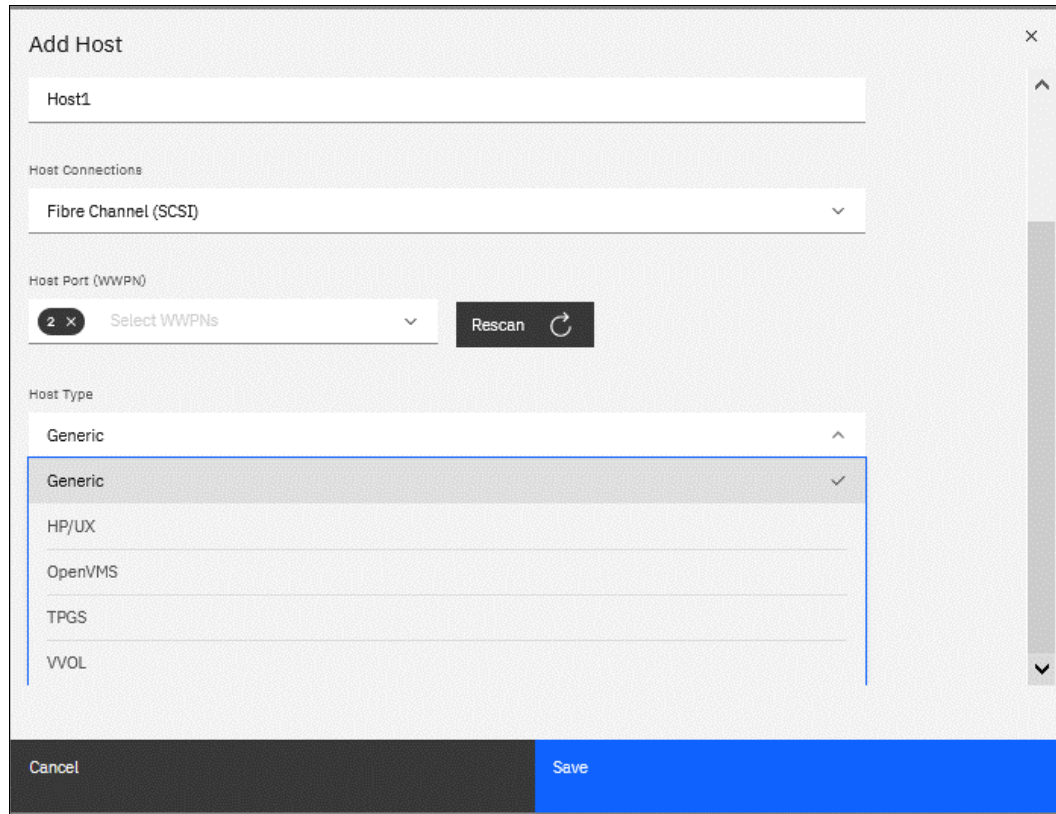


Figure 7-9 Host types selection

6. If you set up object-based access control (OBAC) as described in Chapter 11, "Ownership groups" on page 723, then select the **Advanced** section and choose the ownership group that you want the host to be a part of from the **Ownership Group** menu, as shown in Figure 7-10 on page 423.

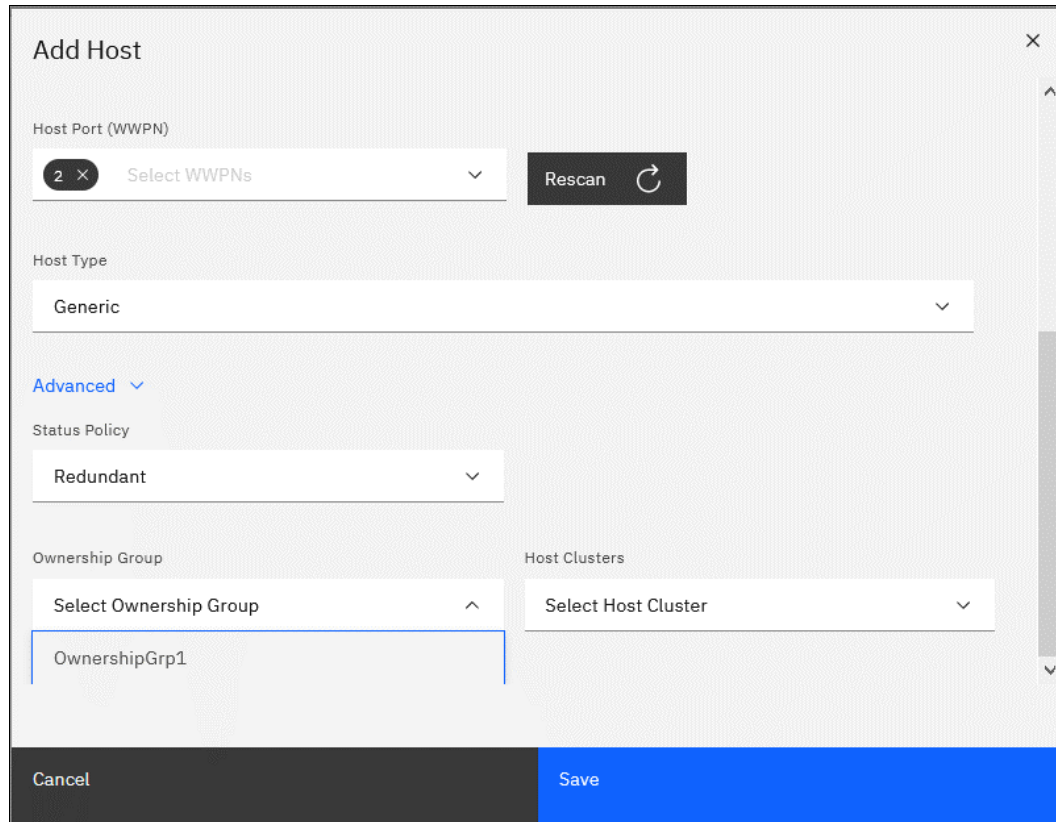


Figure 7-10 Adding a host to an ownership group

Note: If the host cluster object was created, then the Host Clusters list appears in the Advanced section, as shown in Figure 7-11. Use this list to add a host to the cluster.

7. Click **Save** to create the host object.
8. Repeat these steps for all of your FC hosts. Figure 7-11 shows the All Hosts view after creating a host.

Name	Status	Host Type	# of Ports	Host Mappings	Host Cluster ID	Host Cluster Name	Protocol Type
Hosts	Online	Generic	2	No			SCSI

Figure 7-11 Hosts view after creating a host

After defining the FC hosts, you can create volumes and map them to the created hosts, which is described in Chapter 6, “Volumes” on page 299.

Preparing for iSCSI connection

Before configuring or creating iSCSI host objects on an IBM Spectrum Virtualize system, you must make sure that the iSCSI connectivity configuration is done.

First, the iSCSI configuration should be checked and modified in accordance with the planned configuration, and the Ethernet ports must be configured to enable iSCSI communication.

To enable iSCSI connectivity, complete the following steps:

1. Select **Settings** → **Network**, and then click the **iSCSI** tab (see Figure 7-12).

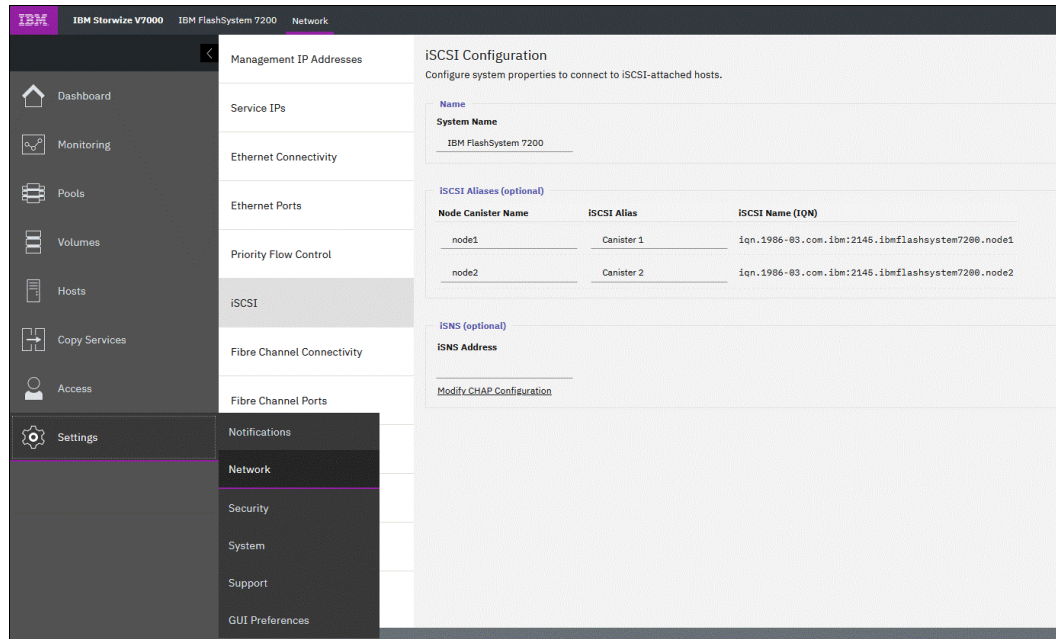


Figure 7-12 Network: iSCSI Configuration view

2. In the iSCSI Configuration window, you can modify the system name, node names, and provide an optional iSCSI Alias for each node, if needed (see Figure 7-13 on page 425).

iSCSI Configuration

Configure system properties to connect to iSCSI-attached hosts.

Name

System Name
IBM FlashSystem 7200

iSCSI Aliases (optional)

! **Renaming a node**
Changing a node name also changes the iSCSI-qualified name (IQN) of the node and might require reconfiguration of all iSCSI-attached hosts for the node.

Node Canister Name	iSCSI Alias	iSCSI Name (IQN)
node_1	Canister 1	iqn.1986-03.com.ibm:2145.ibmflashsystem7200.node1
node_2	Canister 2	iqn.1986-03.com.ibm:2145.ibmflashsystem7200.node2

i **Pending changes** Changes have not yet been applied to the system for the node alias or name. [Apply Changes](#)

iSNS (optional)

iSNS Address

[Modify CHAP Configuration](#)

Figure 7-13 iSCSI Configuration modification

- The interface shows an Apply Changes prompt to apply any changes that are made before continuing.

In the lower left of the configuration window, it is possible to configure internet Storage Name Service (iSNS) addresses and Challenge Handshake Authentication Protocol (CHAP) if they are needed in your environment.

Note: The authentication of hosts is optional. By default, it is disabled. The user can choose to enable CHAP or *CHAP authentication*, which involves sharing a CHAP secret between the cluster and the host. If the correct key is not provided by the host, the IBM FlashSystem system does not allow it to perform I/O to volumes. Also, you can assign a CHAP secret to the cluster.

- Configure the Ethernet ports that you plan to use for iSCSI communication by selecting **Settings** → **Network**, and then clicking the **Ethernet Ports** tab to see the list of ports to configure an Ethernet port for iSCSI communication with planned IP addresses (see Figure 7-14).

Name	Port	State	IP	Speed	Host Attach	IPv4 Remote Copy	Storage Port IPv4	Storage Port IPv6
io_grp0								
Ethernet Ports								
node1	1	Unconfigured		1Gb/s	No	Disabled	Disabled	Disabled
node2	1	Unconfigured		1Gb/s	No	Disabled	Disabled	Disabled
Priority Flow Control								
node1	2	Unconfigured			No	Disabled	Disabled	Disabled
node2	2	Unconfigured			No	Disabled	Disabled	Disabled
iSCSI								
node1	3	Unconfigured			No	Disabled	Disabled	Disabled
node2	3	Unconfigured			No	Disabled	Disabled	Disabled

Figure 7-14 Ethernet ports for iSCSI connectivity

- Select the port to set the iSCSI IP information. Select **Actions** → **Modify IP Settings**. The dialog box that is shown in Figure 7-15 opens.

Figure 7-15 Modifying the IP settings

- After the IP address is configured for a port, click **Modify** to enable the configuration.
- You can see that iSCSI is enabled for host I/O on the required interfaces by the presence of “yes” in the Host Attach column (see Figure 7-16 on page 427).

Ethernet Ports						
The Ethernet ports can be used for iSCSI connections, host attachment, and remote copy.						
☰ Actions ▾						
Name	Port	↑	State	IP	Speed	Host Attach
▽ io_grp0						
node1	1		✓ Configured	9.71.42.61	1Gb/s	Yes
node2	1		✓ Configured	9.71.42.67	1Gb/s	Yes
node1	2		ⓘ Unconfigured		1Gb/s	No
node2	2		ⓘ Unconfigured		1Gb/s	No
node1	3		ⓘ Unconfigured			No
node2	3		ⓘ Unconfigured			No

Figure 7-16 Ports that are configured for the iSCSI connection

- Repeat the above steps to configure all Ethernet ports that are planned for host communication.
- The iSCSI host connection is enabled *after setting* the IP address by default. There are several actions that can be done with already configured ports, as shown in Figure 7-17. For example, to disable any interfaces that you do not want to be used for host connections and might be used for replication only, select the configured port, and then select **Actions (or right mouse button while hovering over the chosen port) → Modify iSCSI Hosts**.

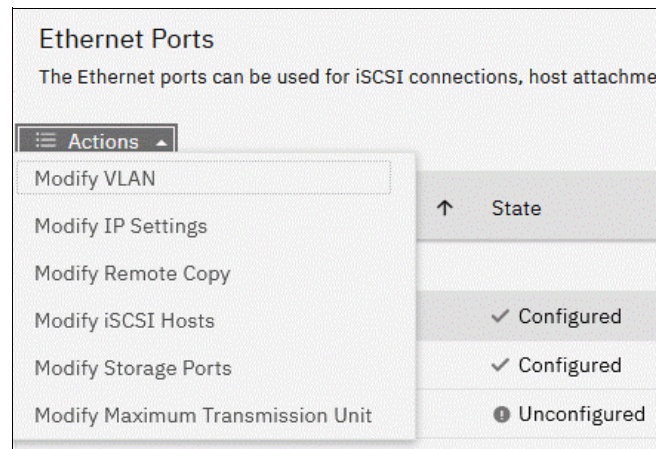


Figure 7-17 Available actions with configured ports

10. The dialog box that is shown in Figure 7-18 opens. Make any necessary changes and click **Modify**.

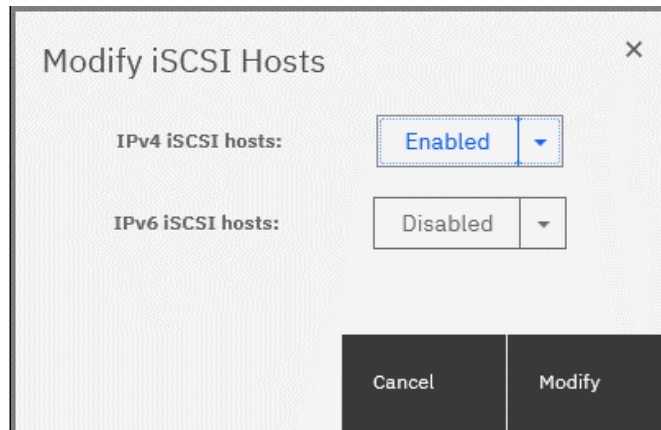


Figure 7-18 Modifying the port for host connectivity

11. As a best practice, it is always good to isolate iSCSI traffic in a separate subnet. It is also possible to set a virtual local area network (VLAN) for the iSCSI traffic. To enable the VLAN, select **Actions** → **Modify VLAN**, as shown in Figure 7-19. The system informs you that at least two ports will be affected by change. To see the details about the effect, click **2 ports affected** (see Figure 7-20 on page 429). Make any necessary changes and click **Modify**.

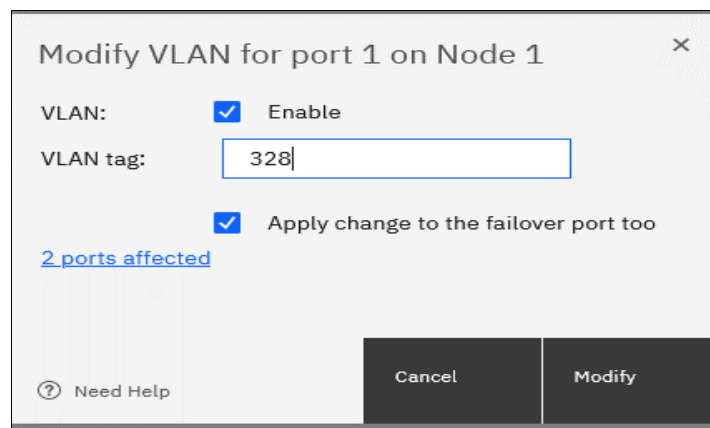


Figure 7-19 VLAN settings modification interface

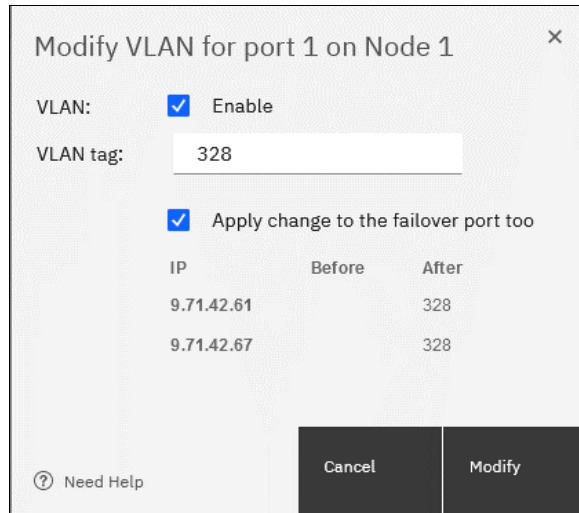


Figure 7-20 VLAN settings: Details

The system is now configured and ready for iSCSI host use. Note the initiator iSCSI Qualified Name (IQN) names of storage node canisters (see Figure 7-13 on page 425) because they are necessary to configure access from host to the storage. For more information about creating volumes and mapping them to a host, see Chapter 6, “Volumes” on page 299.

Creating iSCSI host objects

When creating an iSCSI-attached host, consider the following points:

- ▶ iSCSI IP addresses can fail over to the partner node in the I/O group if a node fails. This design reduces the need for multipathing support in the iSCSI host.
- ▶ The IQN of the host is added to an IBM FlashSystem host object the same way that you add FC WWPNs. For more information about how to obtain the IQN from the host, see the examples in 7.7.3, “iSCSI host connectivity and capacity allocation” on page 475.
- ▶ Host objects can have WWPNs and IQNs.
- ▶ Standard iSCSI host connection procedures can be used to discover and configure the IBM FlashSystem systems as an iSCSI target.
- ▶ The IBM FlashSystem system supports the CHAP authentication methods for iSCSI.
- ▶ The name `iqn.1986-03.com.ibm:2076.<cluster_name>.<node_name>` is the IQN for an IBM FlashSystem node. Because the IQN contains the clustered system name and the node name, do *not* change these names after iSCSI is deployed. It is possible to check the IQN name and iSCSI configuration in the cluster’s GUI by selecting **Settings** → **Network** → **iSCSI**, as shown in Figure 7-21 on page 430.

Note: Check the iSCSI configuration and make the modifications *before* creating iSCSI host objects (configuring the hosts) because in some modifications it might be necessary to redefine the host object or change the configuration on the host.

- ▶ Each node can be given an iSCSI alias as an alternative to the IQN.

To create iSCSI host objects, complete the following steps:

1. In the left pane, select **Hosts** → **Hosts** → **Add Host** (in the host view) to open the host creation window (see Figure 7-21). Choose **iSCSI** in **Host Connections** list. Note that the CHAP authentication trigger is on, so the fields for CHAP credentials are shown in the interface. If CHAP is not used, turn off the trigger.

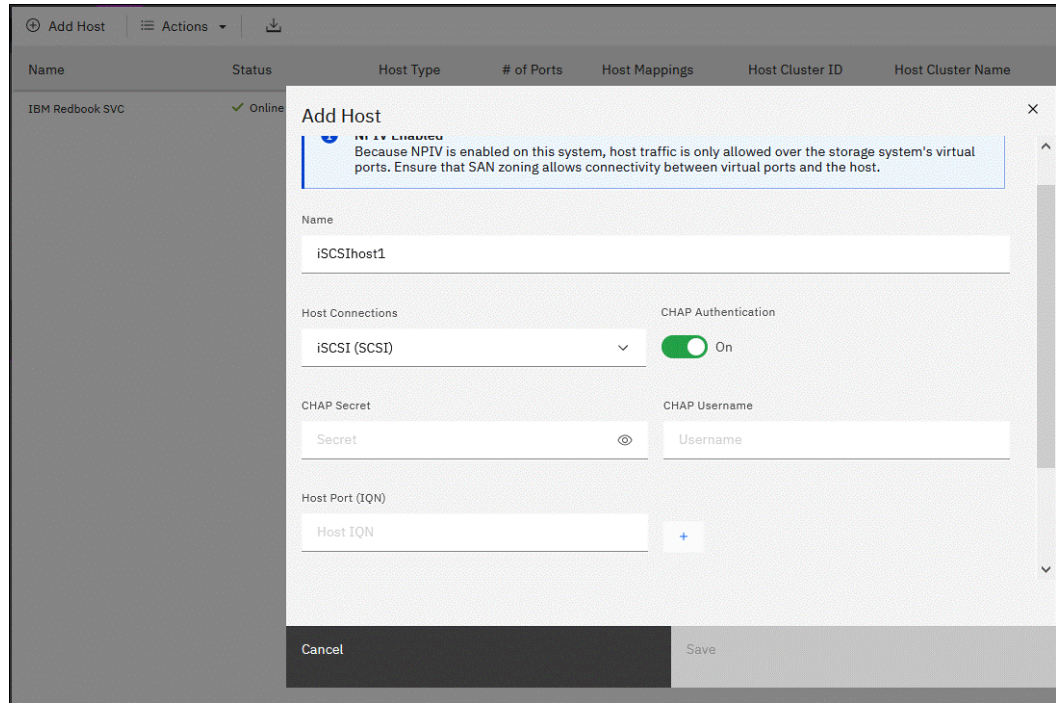


Figure 7-21 Adding the iSCSI host object to the configuration

2. Enter CHAP authentication credentials (if using it), then the hostname into the **Name** field, and then the iSCSI initiator name into the **iSCSI host IQN** field. Click the plus sign (+) if you want to add more initiator names to the host.
3. If you are connecting to an HP-UX or TPGS host, click the **Host type** field (you might need to scroll down the window), and then select the correct host type. For our VMware Elastic Sky X (ESX) host, we select **VMVOL**. However, you can select **Generic** if you are not using VMware vSphere Virtual Volumes (VVOLs).
4. Click **Save** to complete the host object creation.
5. Repeat the above steps for every iSCSI host that must be created. Figure 7-22 shows the Hosts view window after creating the FC host and iSCSI host.

Name	Status	Host Type	# of Ports	Host Mappings	Host Cluster ID	Host Cluster Name	Protocol Type
IBM Redbook SVC	Online	Generic	4	No			SCST
iSCSI-host-1	Offline	Generic	1	No			SCST

Figure 7-22 Defined hosts list

Creating NVMe host objects

The process for creating NVMe hosts is like creating SCSI FC hosts, but instead of using WWPNs, it is necessary to use a host port NQN instead.

Note: To see whether your hosts and IBM FlashSystem system are compatible, see the [SSIC](#).

To configure an NVMe host, complete the following steps:

1. Go to the Host view by selecting **Hosts** → **Hosts** and clicking **Add Host**. Then, in the **Host connections** menu, select **Fibre Channel (NVMe)**, as shown in Figure 7-23.

Add Host [Close]

NPIV Enabled
Because NPIV is enabled on this system, host traffic is only allowed over the storage system's virtual ports. Ensure that SAN zoning allows connectivity between virtual ports and the host.

Name
NVMEHost1

Host Connections
Fibre Channel (NVMe)

Host Port (NQN)
Host NQN +

Host Type
Generic

Advanced [Down Arrow]

Status Policy
Redundant

Cancel Save

Figure 7-23 Creating an NVMe host

2. Enter the hostname and NQN of the host, as shown in Figure 7-24. For more information about how to obtain the host NQN, see 7.7.4, “NVMe over Fabric host connectivity example” on page 478. It is possible to add multiple NQNs by using the + button next to the field.

Figure 7-24 Defining the NQN

3. Click **Save**. Your host appears in the defined host list, as shown in Figure 7-25.

Name	Status	Host Type	# of Ports	Host Mappings	Host Cluster ID	Host Cluster Name	Protocol Type
NVMEHost1	Offline	Generic	1	No			NVMe

Figure 7-25 NVMe host created

Note: As shown in Figure 7-25, it is possible to add the hosts that are not yet connected to the system or are offline by using their known NQN. In this case, their status is *Offline* until they are connected or turned on.

4. The storage system I/O group NQN must be configured on the host so that it can access the mapped capacity. Also, you can use automatic discovery from the host to find the NQN of the I/O group if the connection and zoning is done correctly. To discover the I/O group NQN, run the `lsiogrp` command, as shown in Example 7-13 on page 433.

Example 7-13 The `lsiogrp` command

```
IBM_IBM FlashSystem:GLTLoaner:superuser>lsiogrp 0
id 0
name io_grp0
node_count 2
vdisk_count 8
host_count 1
flash_copy_total_memory 20.0MB
flash_copy_free_memory 20.0MB
remote_copy_total_memory 20.0MB
remote_copy_free_memory 20.0MB
mirroring_total_memory 20.0MB
mirroring_free_memory 20.0MB
raid_total_memory 350.0MB
raid_free_memory 310.2MB
maintenance no
compression_active no
accessible_vdisk_count 8
compression_supported yes
max_enclosures 20
encryption_supported yes
flash_copy_maximum_memory 2048.0MB
site_id
site_name
fctargetportmode enabled
compression_total_memory 0.0MB
deduplication_supported yes
deduplication_active no
nqn nqn.1986-03.com.ibm:nvme:2145.000002042140049E
```

You can now configure your NVMe host to use the IBM SAN Volume Controller (SVC) as a target.

Note: For more information about a compatibility matrix and supported hardware, see [IBM Documentation](#) and the [SSIC](#).

7.5.2 Host clusters

IBM Spectrum Virtualize V7.7 introduced the concept of a *host cluster*. A host cluster object enables a user to collect individual hosts in a group, which is treated as one entity instead of dealing with all the hosts individually in the cluster.

The host cluster object is useful for hosts that are clustered on OS levels. Examples are Microsoft Clustering Server, IBM PowerHA®, Red Hat Cluster Suite, and VMware ESX. By defining a host cluster object, a user can map one or more volumes to this host cluster object.

As a result, the volume or set of volumes are mapped, and access is shared by all individual host objects that are included into the host cluster object. Note that each of the volumes is mapped by using the same SCSI ID to each host that is part of the host cluster by running a single command.

Although a host can be a part of a host cluster object, volumes can still be assigned to an individual host in a *non-shared manner*. A policy can be devised that can pre-assign a standard set of SCSI IDs for volumes to be assigned to the host cluster object and shared with all host objects in the host cluster, and another set of SCSI IDs to be used for individual non-shared assignments to hosts.

Note: For example, SCSI IDs 0 - 100 for individual host assignment and SCSI IDs that are greater than 100 can be used for host cluster. By using such a policy, specific volumes are not shared, and common volumes for the host cluster can be shared. For example, the boot volume of each host can be kept private while data and application volumes can be shared.

Creating a host cluster

This section describes how to create a host cluster. It is assumed that individual hosts were created. Complete the following steps:

1. From the menu on the left, select **Hosts** → **Host Clusters** (see Figure 7-26).

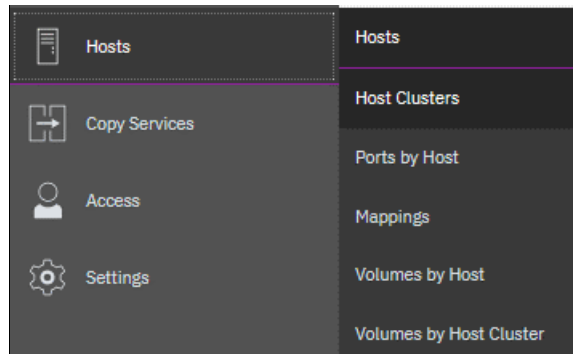


Figure 7-26 Host clusters menu

2. Click **Create Host Cluster** to open the wizard that is shown in Figure 7-27.

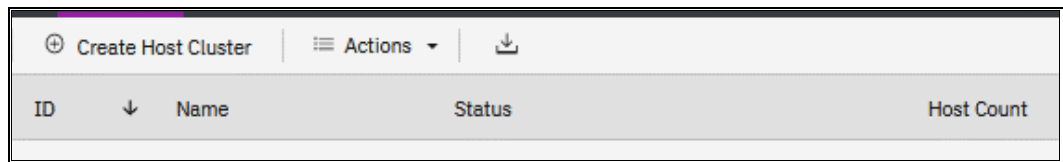


Figure 7-27 Create Host Cluster

3. Enter a cluster name, and if applicable choose the ownership group that the hosts are a part of. Then, you can select the individual hosts that you want in the cluster object by pressing the Ctrl or Shift keys and selecting them, as shown in Figure 7-28 on page 435. Click **Next** after you are done.

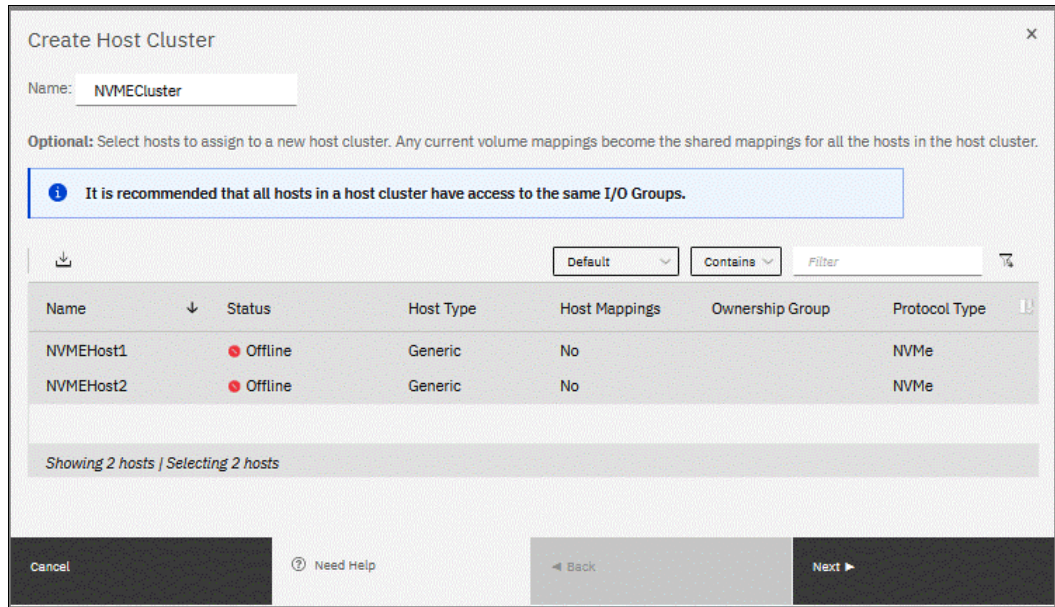


Figure 7-28 Host Cluster details definition

4. A summary opens in which you can confirm that you selected the correct hosts. Click **Make Host Cluster** (see Figure 7-29).

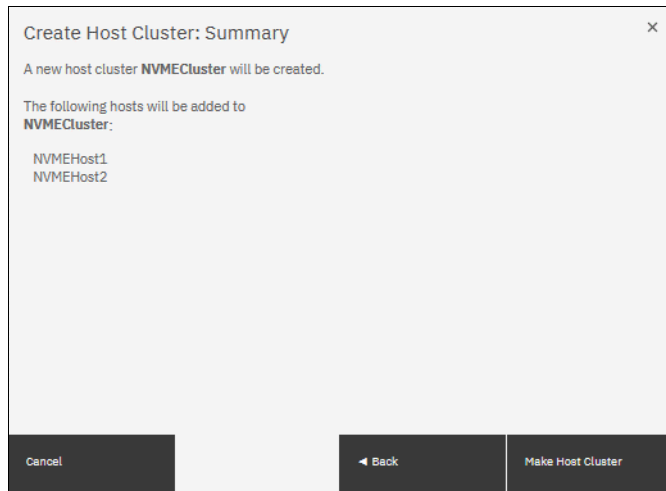


Figure 7-29 Create Host Cluster: Summary

5. After the task completes, the cluster that was created can be seen in the Host Clusters view (see Figure 7-30).

ID	Name	Status	Host Count	Mappings Count	Ports Count	Protocol Type
0	NVMECluster	Offline	2	0	2	NVMe

Figure 7-30 Host Clusters view

Note: The host cluster status depends on its member hosts. One offline or degraded host sets the host cluster status as Degraded. If all member hosts are offline, the cluster status is set to Offline.

From the Host Clusters view, many options are available to manage and configure the host cluster. These options are accessed by selecting a cluster and clicking **Actions** (see Figure 7-31).

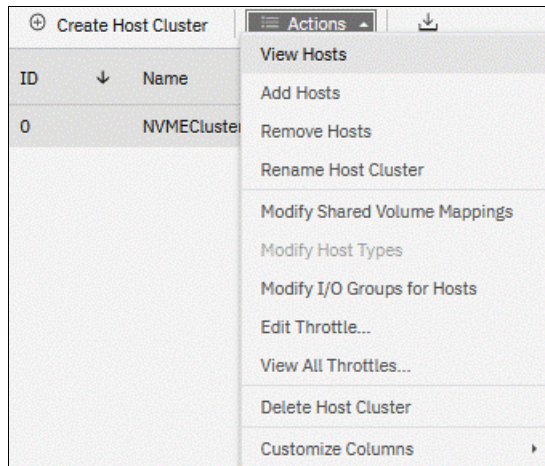


Figure 7-31 Host Clusters Actions menu

From the Actions menu, you can do the following tasks:

View Hosts	View the hosts status within the cluster.
Add Hosts or Remove Hosts	Add or remove hosts from the cluster.
Rename Host Cluster	Rename the host cluster.
Modify Shared Volume Mappings	Add or remove volumes that are mapped to all hosts in the cluster while maintaining the same SCSI ID for all hosts.
Modify Host Types	Change the host type from, for example, generic to VVOLs.
Modify I/O Groups for Hosts	Assign or restrict volume access to specific I/O groups.
Edit Throttle	Restrict the megabytes per second (MBps) or input/output operations per second (IOPS) bandwidth for the host cluster.
View All Throttles	Show all throttling settings, and allow for changing, deleting, or refining throttle settings.
Delete Host Cluster	Delete a host cluster.
Customize Columns	Modify which columns are displayed that show the properties of the host cluster.

For more information about these actions, see 7.5.4, “Actions on host clusters” on page 448.

7.5.3 Actions on hosts

This section covers host administration, including host modification, host mappings, and deleting hosts. The basic host creation process is described in 7.5.1, “Creating hosts” on page 419.

Select **Hosts** → **Hosts** view and right-click one of the existing hosts, or expand the **Actions** menu. You see a list of actions that can be performed on a host, as shown in Figure 7-32.

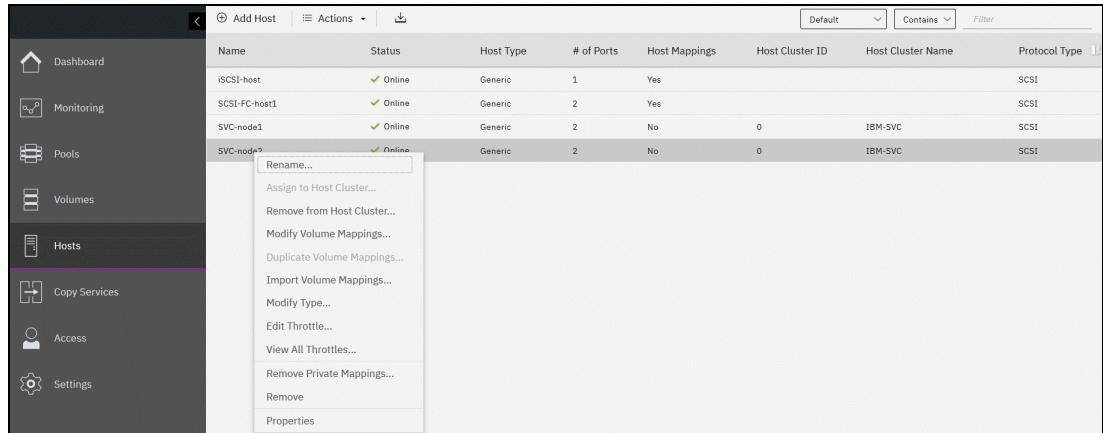


Figure 7-32 Actions on hosts

All the available actions are described in this section:

- ▶ Renaming a host
- ▶ Assigning or removing a host to or from a host cluster
- ▶ Modifying volume mappings
- ▶ Duplicating and importing mappings
- ▶ Modifying the host type
- ▶ Viewing and editing throttles
- ▶ Removing private mappings from a host
- ▶ Removing a host
- ▶ Viewing IP logins
- ▶ Viewing the host properties

Renaming a host

To rename a host, complete the following steps:

1. Select the host, right-click it, and select **Rename**.
2. Enter a new name and click **Rename** (see Figure 7-33). If you click **Reset**, the changes are reset to the original hostname.

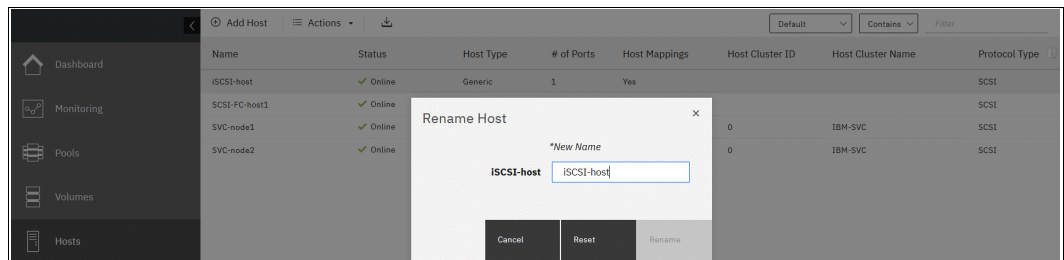


Figure 7-33 Renaming a host

Assigning or removing a host to or from a host cluster

You can assign a stand-alone host to a host cluster or remove a host from a cluster to make it a stand-alone cluster.

The **Assign to Host Cluster** action is active only if you select a host that does not belong to the cluster and if at least one host cluster object exists. If there are no host cluster objects that are configured, you must create one.

To assign a host to existing cluster, perform the following actions:

1. Right-click on the host or a set of hosts you want to add and select **Assign to Host Cluster**.

Note: To select multiple objects, press and hold the Ctrl key and click each host that you need, or press and hold the Shift key and click the first and the last objects that must be selected.

2. Select the existing cluster to which you want to add the host, as shown in Figure 7-34, and click **Next**.

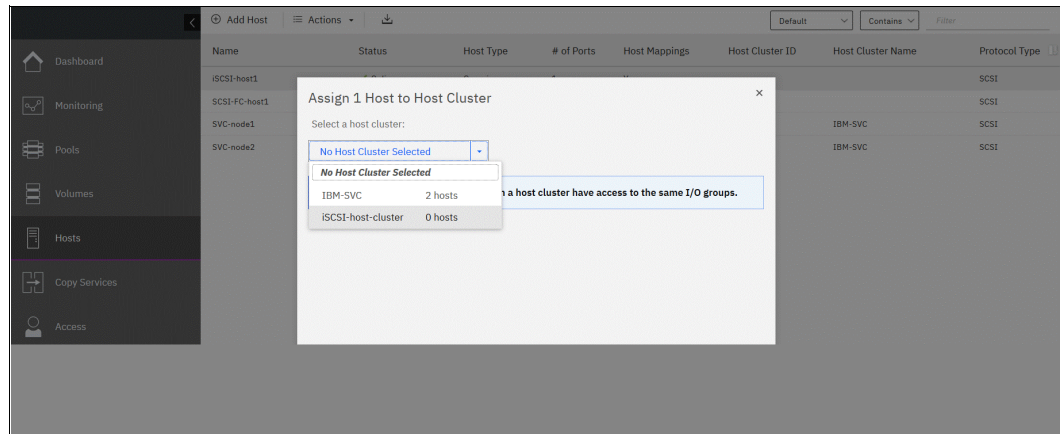


Figure 7-34 Assigning a host to a cluster

3. Your storage system checks for SCSI ID conflicts. In a host cluster, all hosts must have the same SCSI IDs for a mapped volume. For example, a single volume cannot be mapped with SCSI ID 0 to one host and with SCSI ID 1 to another host.

If no SCSI ID conflict is detected, the system provides a list of configuration settings for you to verify, as shown in Figure 7-35 on page 439. Click **Assign** to complete the operation. When the operation completes, the host is included in all existing host cluster volume mappings.

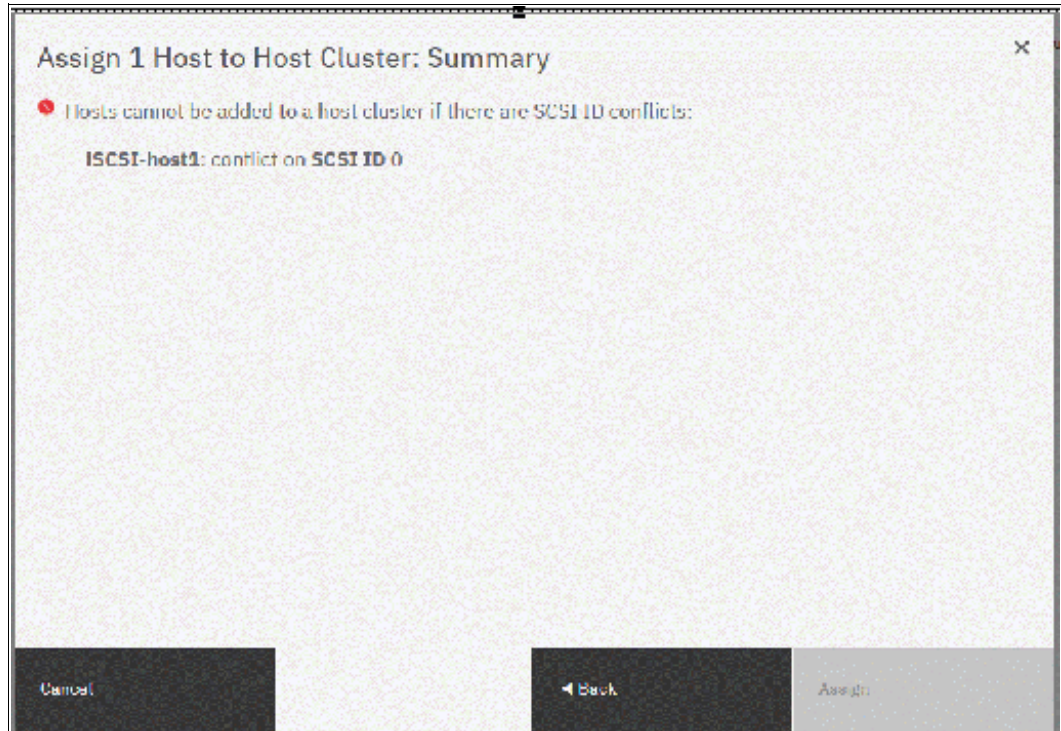


Figure 7-35 Assign host to host cluster confirmation

If a host already has private volume mappings that use SCSI IDs that are used in host cluster shared mappings, a SCSI ID conflict is raised, as shown in Figure 7-36. In this case, you cannot assign this host to the host cluster. First, you must resolve the ID conflict by removing the private host volume mappings or by changing the assigned SCSI IDs for conflicting mappings.

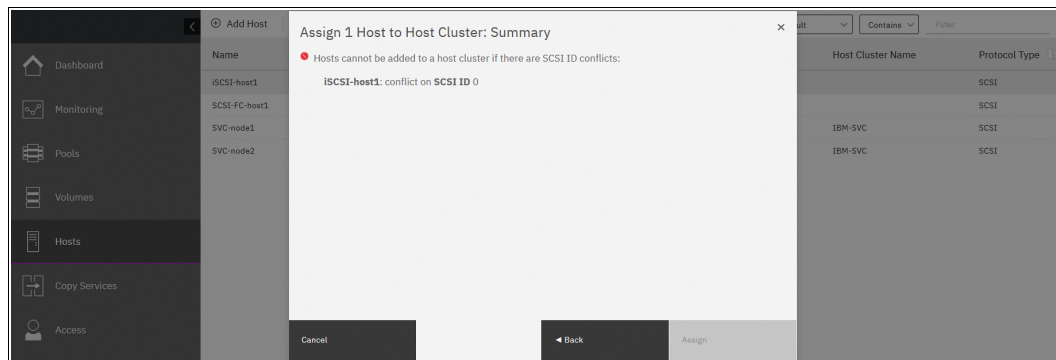


Figure 7-36 Conflict between private and shared volume mappings

To remove a host from a host cluster, complete the following steps:

1. Select a host or a group of hosts, right-click them, and click **Remove from Host Cluster**.
2. A dialog window opens, as shown in Figure 7-37.

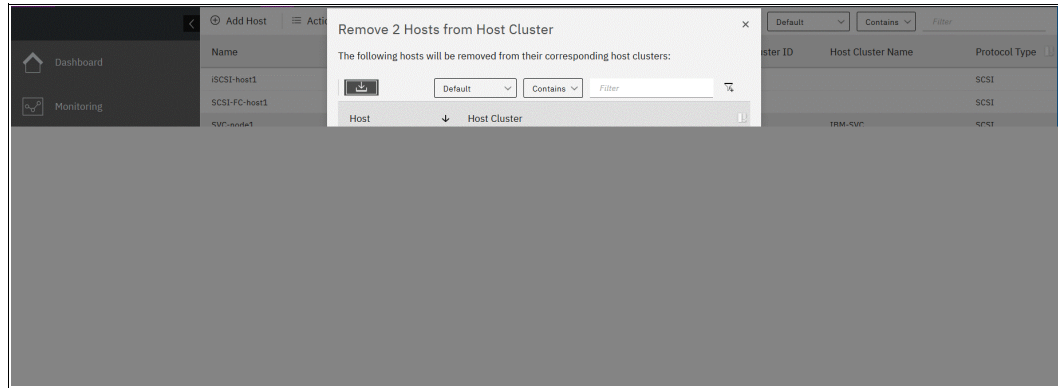


Figure 7-37 Removing a host from a host cluster

In this window, you can verify a list of hosts to be removed and make a choice about what to do with the volume mappings of the hosts that are deleted. They can be removed, or retained and converted from shared to private mappings. Click **Remove Hosts** to complete the operation.

Modifying volume mappings

By using the **Modify Volume Mappings** action, you can modify private volume mappings for a single host. To do so, complete the following steps:

1. Right-click a host and select **Modify Volume Mappings**.
2. A window that shows a list of existing private volume mappings is shown (see Figure 7-38).

Note: Host cluster shared mappings are not shown in this view. Only host private mappings are listed. To modify the share host cluster mappings, use another GUI view, as described 7.5.4, “Actions on host clusters” on page 448.

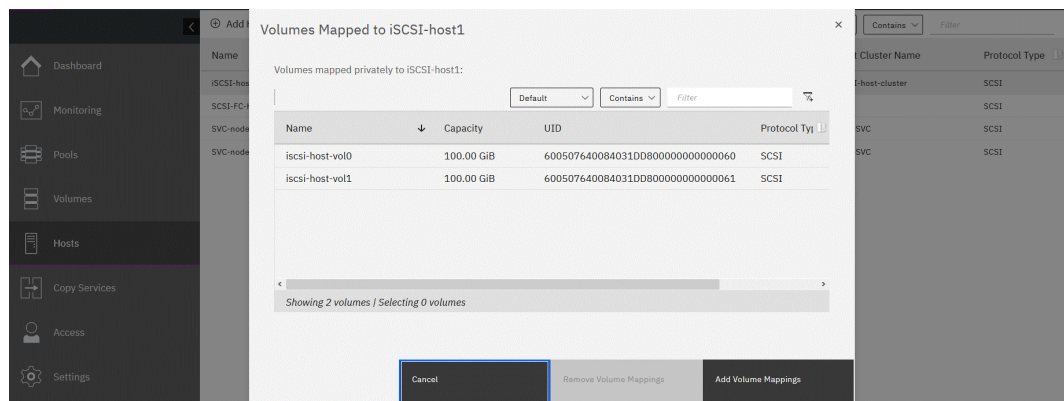


Figure 7-38 Modifying the host volume mappings

3. To remove volume mappings, select the ones that need to be deleted, and click **Remove Volume Mappings**. The next window prompts you to verify your changes and complete the removal procedure.
4. If you intend to add a private mapping, click **Add Volume Mappings**. A list of volumes appears, as shown in Figure 7-39. If a volume already has a private mapping to this host, or it has a shared mapping with a host cluster that includes this host, it is not listed.

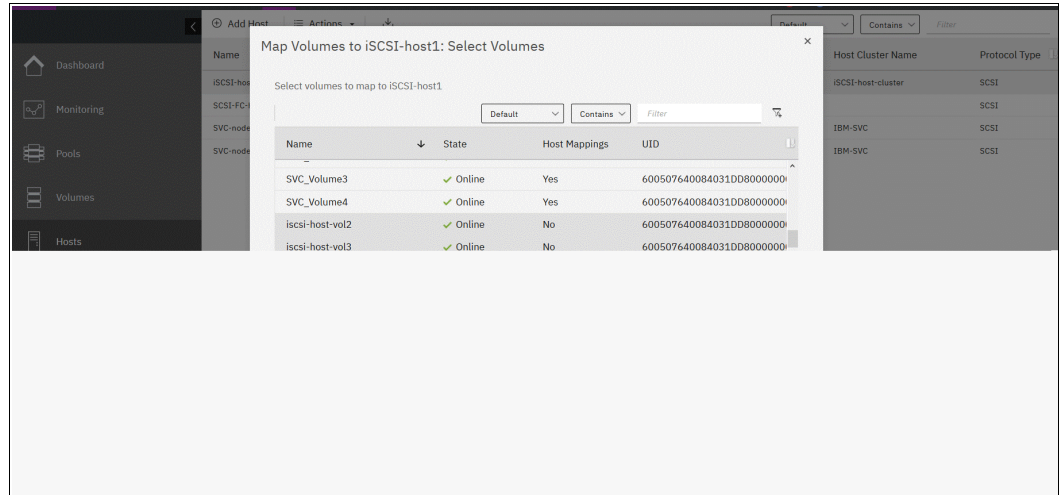


Figure 7-39 Adding private mappings

If a volume that you want to map is already mapped to another host or host cluster, you see Yes in Host Mappings column. If you attempt to map that volume to the host, a warning is shown (Figure 7-40). You can still continue to add a mapping if access is coordinated at the host side.

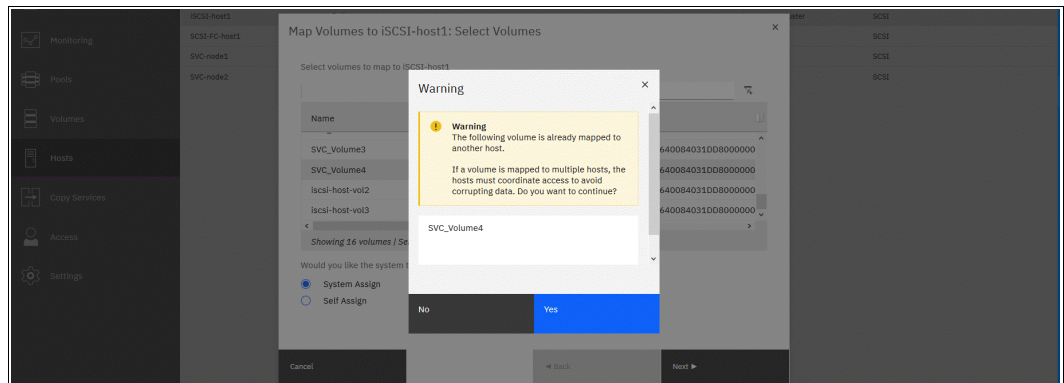


Figure 7-40 Warning that a mapping to another host exists

- By default, the storage system automatically assigns the next available SCSI ID for new mappings. However, you also can click **Self Assign** to manually assign SCSI logical unit number (LUN) IDs, as shown in Figure 7-41. Note that in Figure 7-38 on page 440 only two existing mappings were shown for this host, but Figure 7-41 shows three mappings because the third mapping is a shared host cluster mapping, which was not shown in previous views.

Note: The SCSI ID of the volume can be changed only before it is mapped to a host. Changing it afterward is a disruptive operation because the volume must be unmapped from the host and mapped again with a new SCSI ID.

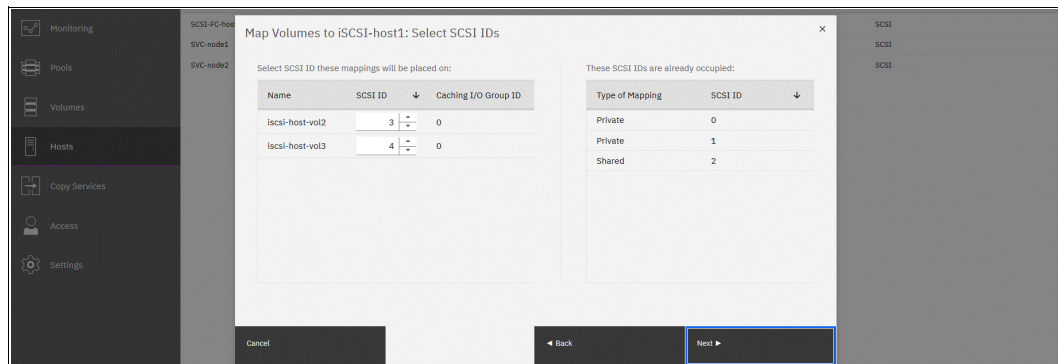


Figure 7-41 Choosing SCSI IDs manually

- When the assignments are done, click **Next** to verify the prepared changes, and click **Map Volumes** to complete the operation.

Duplicating and importing mappings

Volumes that are assigned to one host can be quickly and easily mapped to another host object. You might make this assignment, for example, when replacing an aging host's hardware and want to ensure that the replacement host node can access the same set of volumes as the old host.

You can accomplish this process in two ways:

- ▶ Duplicating the mappings from the selected host to the new host object.
- ▶ Selecting a new host and importing host mappings from another host object.

Notes:

- ▶ When duplicating or importing mappings, all existing mappings are copied, both private and shared. The shared mappings of an old host become the private mappings of a new host.
- ▶ You can duplicate mappings only for a host that does not have volumes mapped, or import mappings only for a host that has no mappings.

To duplicate the mappings, complete the following steps:

- Right-click the host that you want to duplicate (source host) and click **Duplicate Volume Mappings**.
- The Duplicate Mappings window opens. With it, you can select a target host to which you want to map all the existing source host volumes (see Figure 7-42 on page 443). In this example, only the target candidate is a host that has no existing mappings.

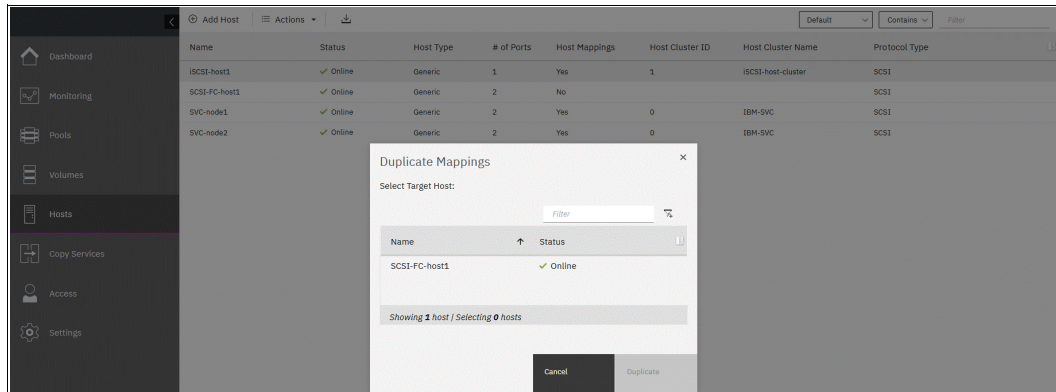


Figure 7-42 Duplicate Mappings window

3. Select a target host and click **Duplicate**. After the operation completes, the target host has the same volume mappings that the source host has. Both private and shared mappings are duplicated. Mappings on the source host also remain, and they can be deleted manually if necessary.

To import hosts mappings from an existing host to a new host, complete the following steps:

1. Right-click the new host that has no mapped volumes and select **Import Volume Mappings**. Note that if the host already has private or shared mappings, this action is inactive (disabled) in an Actions menu.
2. The Import Mappings window opens. Select the source host from which you want to import the volume mappings, as shown in Figure 7-43, and click **Import**.

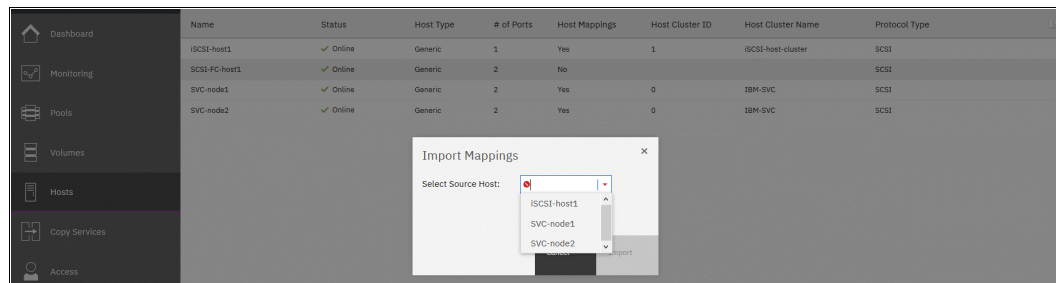


Figure 7-43 Import volume mappings source host selection

3. After the task completes, the host has all the volume mappings as the source host. Shared mappings in which the source host participates are imported as private. Mappings on the source host also remain, and they can be deleted manually if necessary.

Note: You can import mappings only from a source host that is in the same ownership group as your target host. If they are not in the same ownership group, the import fails with “The command failed because the objects are in different ownership groups” message.

Modifying the host type

During the host creation process, the host type is specified. If it must be changed, use the **Modify Type** action.

To change the host type, perform the following actions:

1. Select a host or several hosts that need to be modified, right-click, and select **Modify Type**.
2. A Modify Type dialog opens, as shown in Figure 7-44.

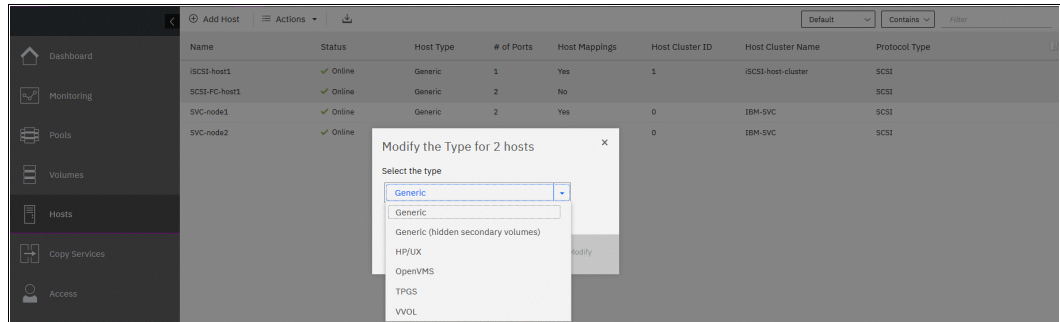


Figure 7-44 Changing the host type

3. Select one of the available host types:
 - **Generic**: The default host type. It is used in most cases, and for all NVMe hosts.
 - **Generic (hidden secondary volumes)**: If this host type is set, all RC relationship secondary volumes are unavailable to that host.
 - **HP/UX, OpenVMS, TPGS**: Set when IBM Documentation requires the setting for the appropriate host OS types.
 - **VVOL**: Set if the host is configured to work with VVOLs.

For more information about host type selection, see [IBM Documentation](#).

4. Click **Modify** to complete the task.

Viewing and editing throttles

Throttles are a mechanism to control the amount of resources that are used when the system is processing I/O for a specific host or host cluster. If a throttle is defined, the system either processes the I/O, or delays the processing of the I/O to free resources for more critical I/O.

A host throttle sets the limit for combined read and write I/O to all mapped volumes. Other hosts accessing the same set of volumes are not affected by a host throttle.

To create a host throttle, or change or remove an existing host throttle, complete the following steps:

1. Select one host or several hosts, right-click, and select **Edit Throttle**.
2. The Edit Throttle for Host dialog opens, as shown in Figure 7-45 on page 445.

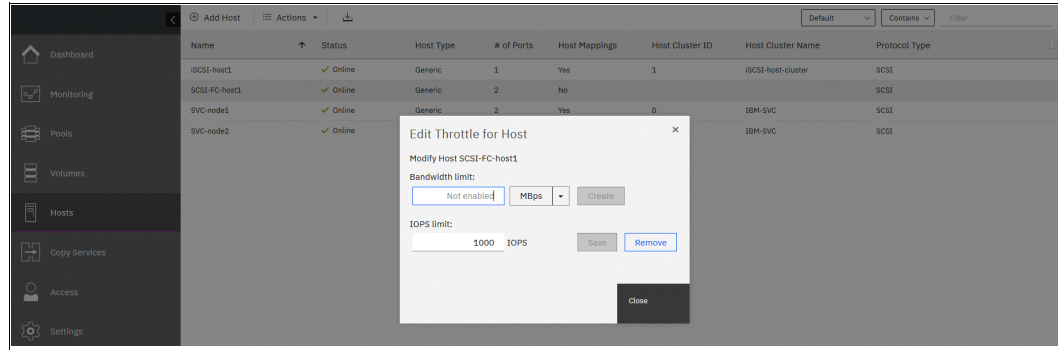


Figure 7-45 Edit Throttle for Host dialog

3. Specify the **IOPS limit**, **Bandwidth limit**, or both. Click **Create** to create a host throttle, change the throttle limit and click **Save** to edit an existing throttle, or click **Remove** to delete a host throttle.
4. When done editing or creating, click **Close**.

To view and edit all the throttles that are configured on the system, right-click any of the hosts and select **View All Throttles**. As shown in Figure 7-46, a list of all throttles that are configured on the system appears. You can switch between throttle types by clicking the drop-down menu next to the **Actions** menu. You can also change the view to see all the system's throttles in one list.

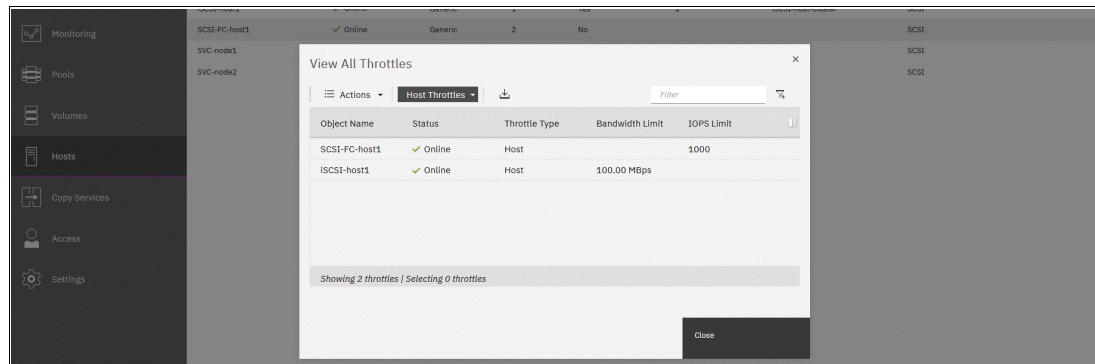


Figure 7-46 View All Throttles window

From this view, you can delete or edit any existing throttle by right-clicking it in the list and selecting the required action.

Removing private mappings from a host

A host can access only those volumes on the system that are mapped to it. You can remove access to all volumes for one host regardless of how many volumes are mapped to it. Only private mappings are removed, and shared host cluster volume mappings remain.

To remove all host private mappings, complete the following steps:

1. Right-click a host that needs its mapping to be removed, and select **Remove Private Mappings**.
2. If a host is assigned to cluster, a window opens with a warning that shared mappings will not be removed. Click **Yes** if you want to continue.

- In the next window, to confirm your action, enter the number of volume mappings to be removed, as shown in Figure 7-47, and click **Remove**.

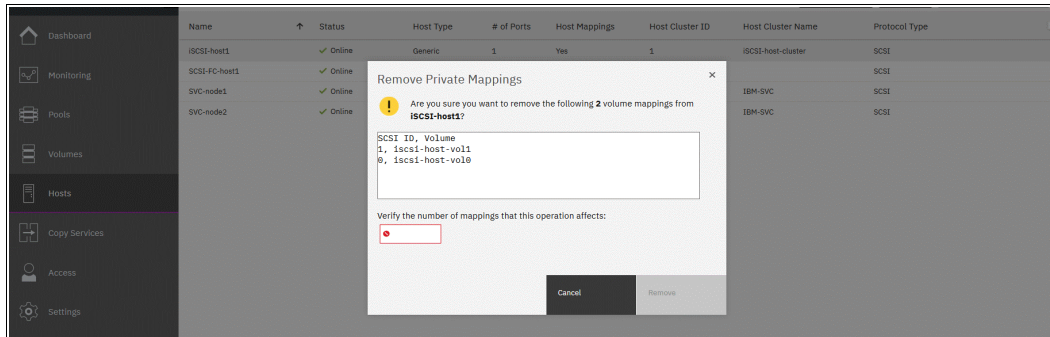


Figure 7-47 Confirming the number of mappings to be removed

Note: When you click **Remove**, the host loses access to the unmapped volumes. Ensure that you run the required procedures on your host OS, such as unmounting the file system, taking the disk offline, or disabling the volume group, before removing the volume mappings from your host object on the GUI.

Removing a host

To remove a host object, complete the following steps:

- Select the host or multiple hosts that must be removed, right-click them, and select **Remove**.
- Confirm that the window shows the correct list of hosts that you want to remove by entering the number of hosts to remove and clicking **Remove** (see Figure 7-48).

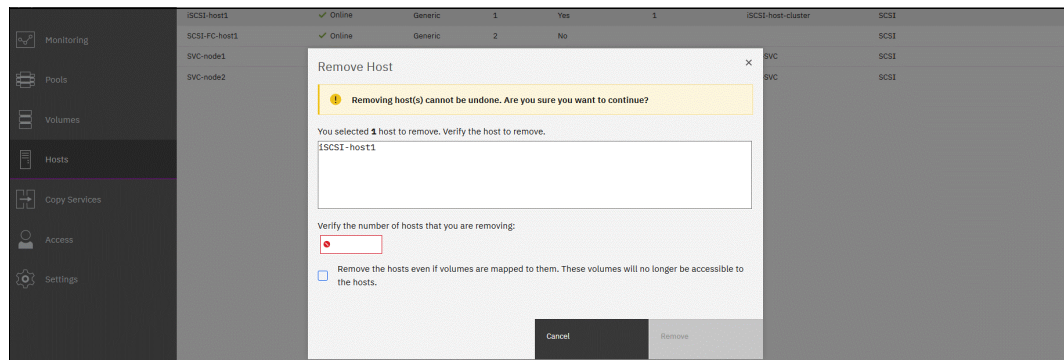


Figure 7-48 Confirming the removal of the host

- If the host that you are removing has volumes that are mapped to it, you can force the removal by selecting the **Remove the hosts even if volumes are mapped to them** checkbox in the lower part of the window. When this option is selected, all volume mappings of this host are deleted, and the host is removed.

Viewing IP logins

If you right-click an iSCSI or iSER host, the **IP Login Information** window opens, where you check the state of the host logins, as shown in Figure 7-49 on page 447. You can use the drop-down menu in the upper part of the window to switch between the IQNs of the host.

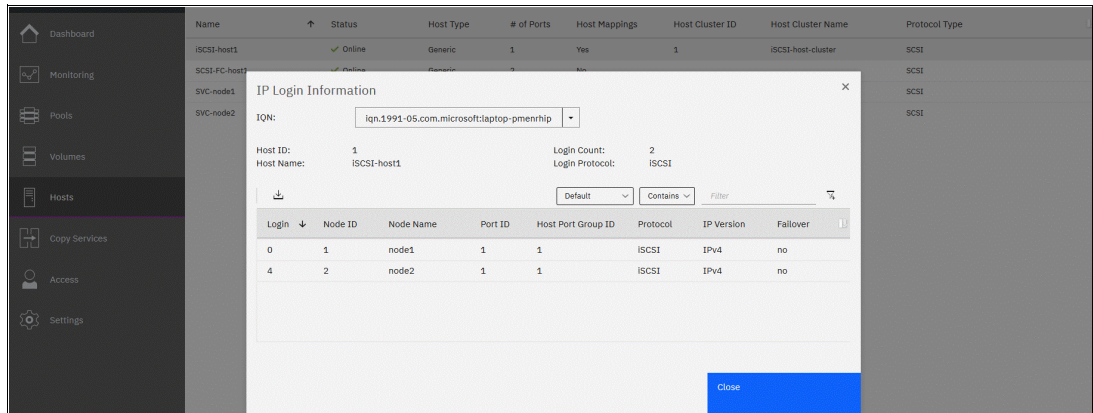


Figure 7-49 Viewing the IP login information

Viewing the host properties

To view a host object's properties, complete the following steps:

1. Right-click a host and select **Properties**.
2. The Host Details window opens (see Figure 7-50).

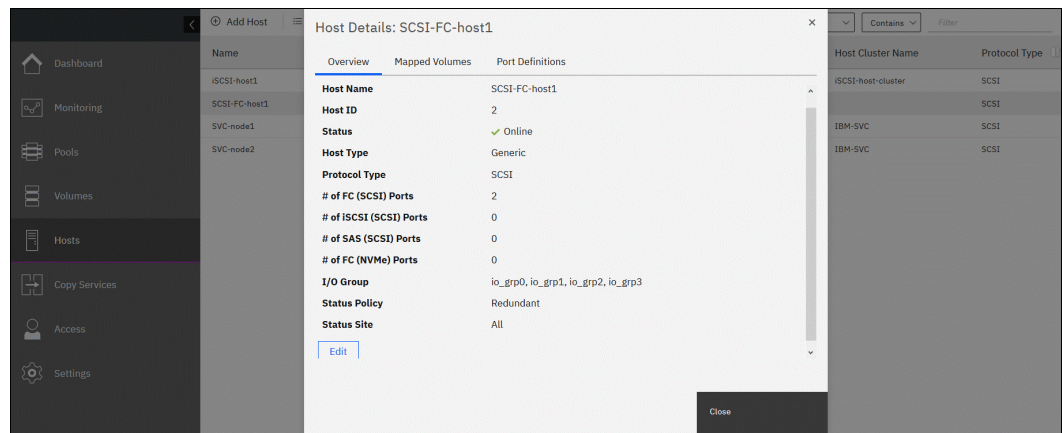


Figure 7-50 Host properties overview

The Host Details window has three tabs: **Overview**, **Mapped Volumes**, and **Port Definitions**:

- In the **Overview** window, you can click **Edit** to change hostname, host type, select and clear the associated host I/O groups, and modify the host status policy and status site.
- In the **Mapped Volumes** tab, you can list all volumes that are mapped to the host. Both private and shared mappings are shown.

- In the **Port Definitions** tab, you can see all ports that belong to the host, add ports to it, or remove any assigned ports. An example is shown in Figure 7-51. This tab is also where you can find the NQN of your NVMe host.

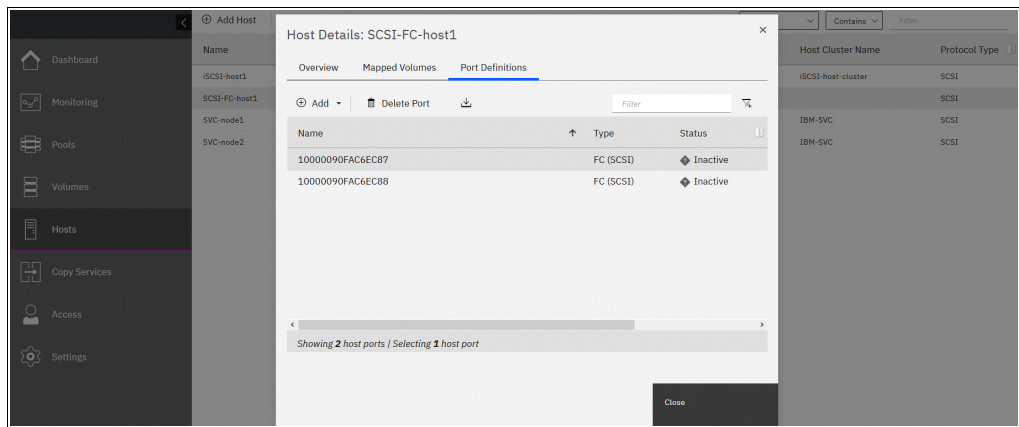


Figure 7-51 Listing port definitions

7.5.4 Actions on host clusters

An overview of the Host Cluster feature and the actions that are required to create a host cluster are in 7.5.2, “Host clusters” on page 433. This section covers actions that can be performed on a host cluster object by using the **Hosts** → **Host Clusters** menu.

Selecting **Hosts** → **Host Clusters** shows a list of configured host clusters and their major parameters, like cluster status, number of hosts in a cluster, and number of shared mappings.

Right-clicking any of the clusters or selecting one or several clusters and clicking the **Actions** drop-down menu opens the list of available actions, as shown in Figure 7-52.

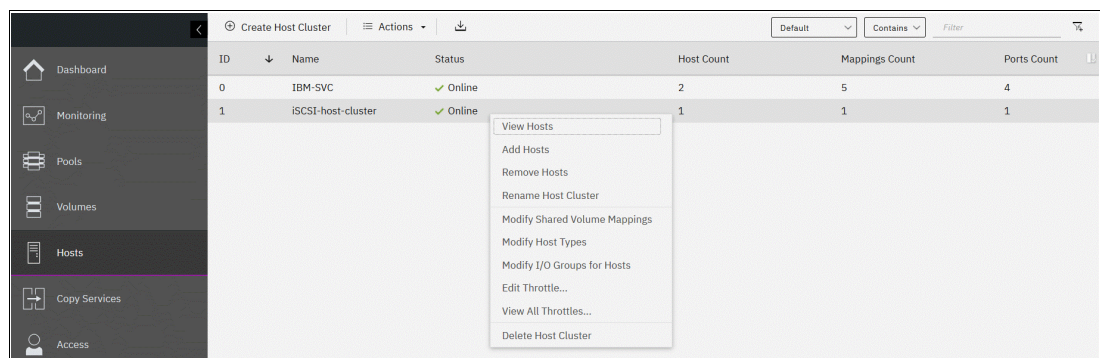


Figure 7-52 Actions that are available on a host cluster object

View Hosts action

By clicking **View Host**, you see a list of hosts that are assigned to a host cluster, as shown in Figure 7-53 on page 449. Click **Next** and **Previous** to switch to other clusters in the cluster list.

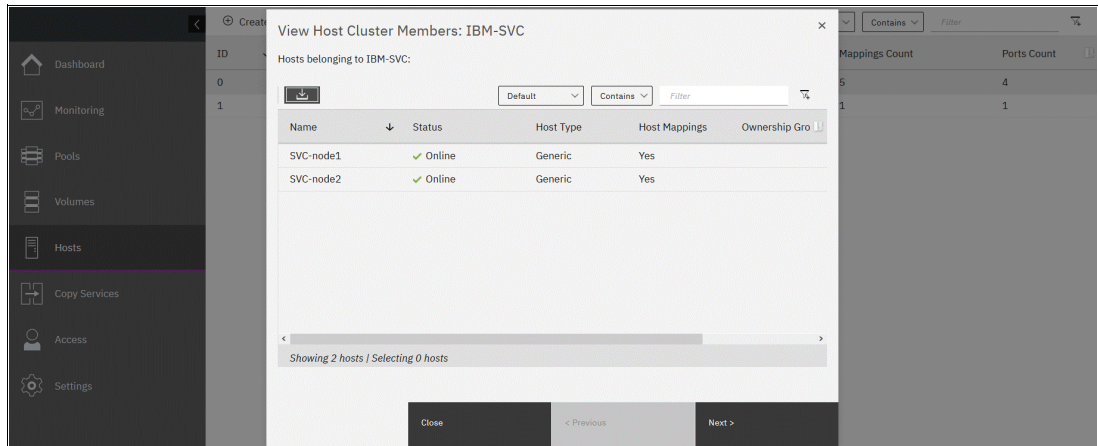


Figure 7-53 View Host Cluster Members window

Add Hosts action

Clicking **Add Hosts** opens a dialog box that shows all stand-alone hosts, that is, the hosts that are not assigned to any clusters, as shown in Figure 7-54. You can select a host to add and click **Next**.

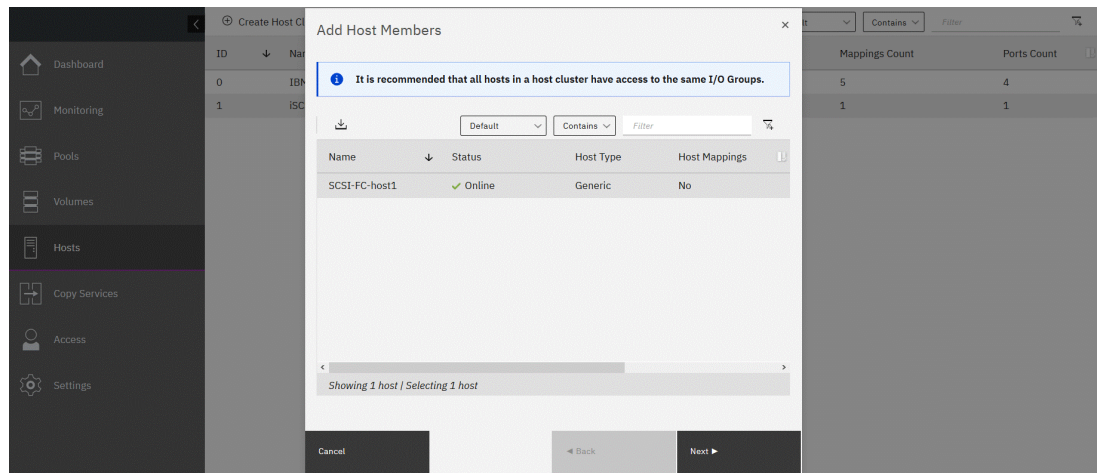


Figure 7-54 Adding a host member

You see a prompt that says that the shared host cluster mappings will be applied to the added host and that the host will gain access to all volumes that are mapped to host cluster, as shown in Figure 7-55.

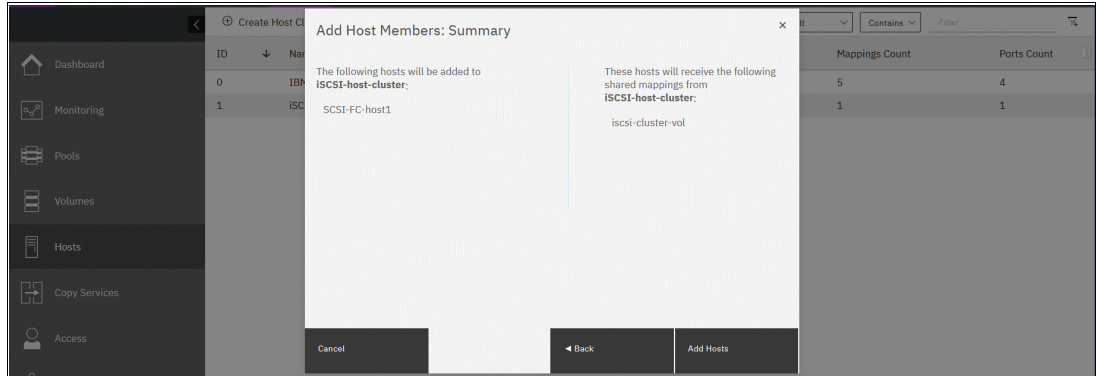


Figure 7-55 Confirming the addition of a host to a cluster

If your changes are correct, click **Add Hosts** to complete the operation.

Remove Hosts action

To remove a host or hosts from a cluster and convert them to stand-alone hosts, right-click the cluster and select **Remove Hosts**. The dialog box that is shown in Figure 7-56 opens.

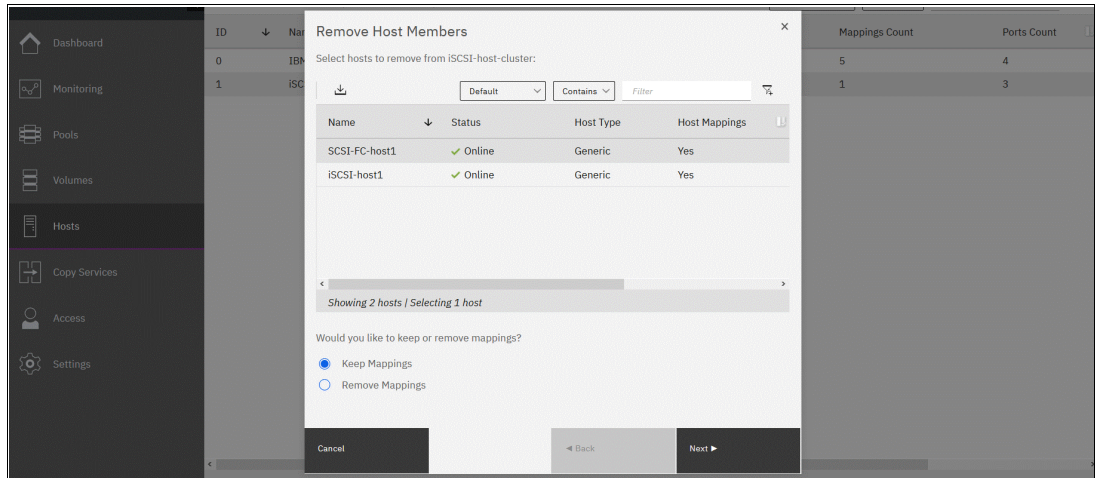


Figure 7-56 Remove Host from Cluster

There are two options:

- ▶ A removed host can keep all the shared cluster mappings as private mappings and retain access to volumes.
- ▶ All shared cluster mappings can be removed from the host.

Select the action that you want, click **Next**, and after verifying the changes, click **Remove Hosts** to complete the procedure.

Rename Host Cluster action

This action changes the host cluster object name.

Modify Shared Volume Mappings action

With this action, you can create shared mappings for a host cluster or modify an existing host cluster.

To add or remove a shared mapping, complete the following steps:

1. Right-click the host cluster and select **Modify Shared Volume Mappings**.
2. A window that shows all shared mappings that exist for the selected cluster opens, as shown in Figure 7-57.

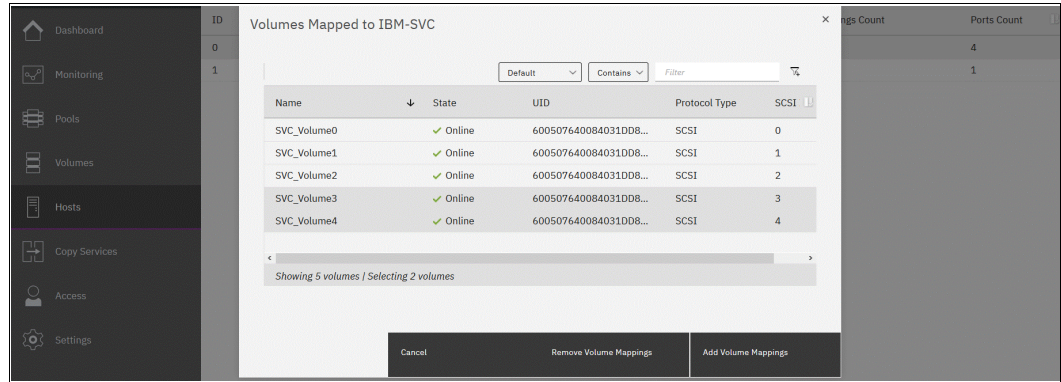


Figure 7-57 Existing shared mappings

3. With this view, you can select one or more shared mappings that must be removed, and then click **Remove Volume Mappings**.
4. If new shared mappings must be created, click **Add Volume Mappings** to open the next window, as shown in Figure 7-58. A list shows the volumes that are not yet mapped to the cluster that was selected.

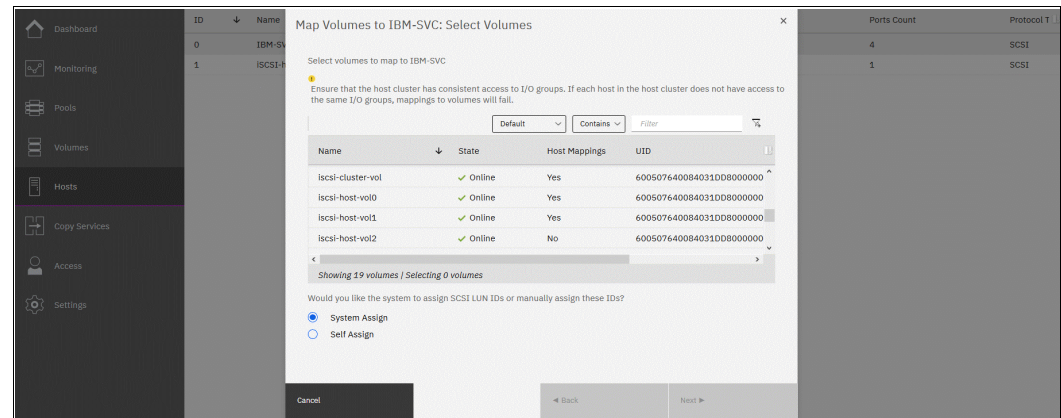


Figure 7-58 Adding shared mappings

- By default, the system assigns the next available SCSI ID for new mappings automatically. But, you can also click **Self Assign** to manually assign SCSI LUN IDs, as shown in Figure 7-59.

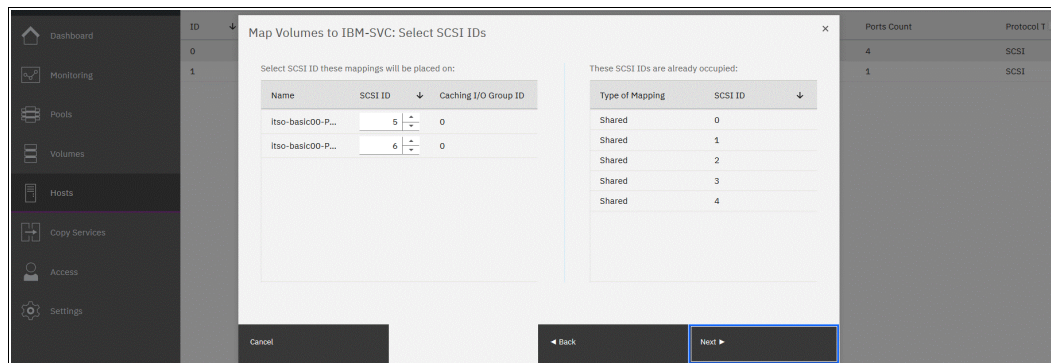


Figure 7-59 Assigning a SCSI ID to mapped volumes manually

- After clicking **Next**, the next window prompts you to verify that the changes are correct, as shown in Figure 7-60. Click **Map Volumes** to complete the operation, click **Back** to return and change the SCSI IDs or IBM-SVC that are being mapped, or click **Cancel** to stop the task.

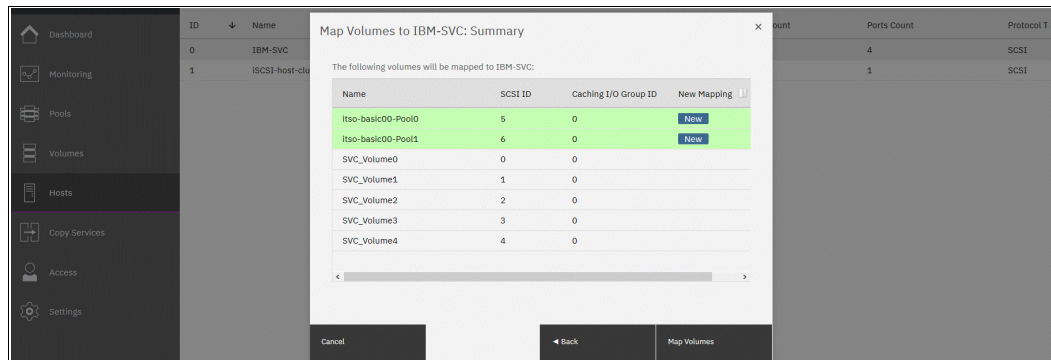


Figure 7-60 Summary of added shared mappings

Modifying host types

With this action, you can change the host type for all members of a host cluster. The procedure is similar to changing a type on a separate host, as described in “Modifying the host type” on page 443. The only difference is that the changes are applied to all hosts that are assigned to the cluster.

Modify I/O groups for hosts action

On the Host Clusters window, it is possible to change a list of I/O groups that are associated with a host. By default, all hosts are assigned to all I/O groups. If necessary, the list of associated I/O groups can be reduced, which leaves the host assigned only to one or a few of them.

The **Modify I/O groups for hosts** action for a host cluster object changes the I/O group assignment for all hosts who are members of this cluster, as shown in Figure 7-61 on page 453.

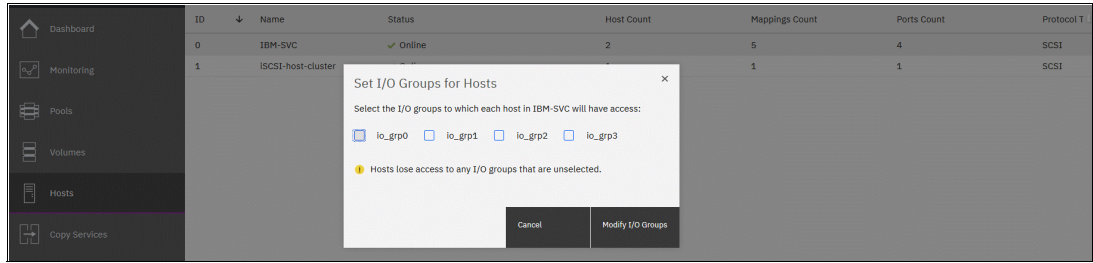


Figure 7-61 Setting the I/O groups for hosts

Edit Throttle and View All Throttles

Throttles are a mechanism to control the amount of resources that are used when the system is processing I/Os on a specific host or host cluster. You can create and edit the host cluster throttle by using the **Edit Throttle** action.

If you are creating a throttle for a host cluster, any hosts within that cluster adopt the throttle for processing.

To create or change a host cluster throttle, complete the following steps:

1. Right-click a host cluster object and select **Edit Throttle**.
2. If the hosts in the cluster already have individual throttles that are defined, the host throttles must be removed. Ensure that any throttles on the hosts that are members of the host cluster are removed before you create a cluster throttle. If host throttles exist for the cluster members, the system shows a warning, as shown in Figure 7-62. You can click **Remove Throttles**, or click **Cancel** to leave the individual throttles and stop the host throttle creation.

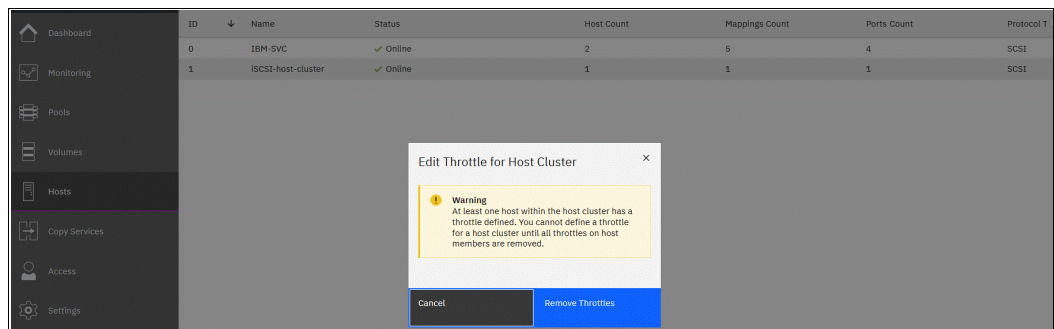


Figure 7-62 Warning that host throttles exist

3. If there are no individual throttles, a window opens where you can set or edit I/O or data rate limits, as shown in Figure 7-63. Click **Create** to create a throttle, or click **IOPS limit** and click **Save** to change the existing throttle.

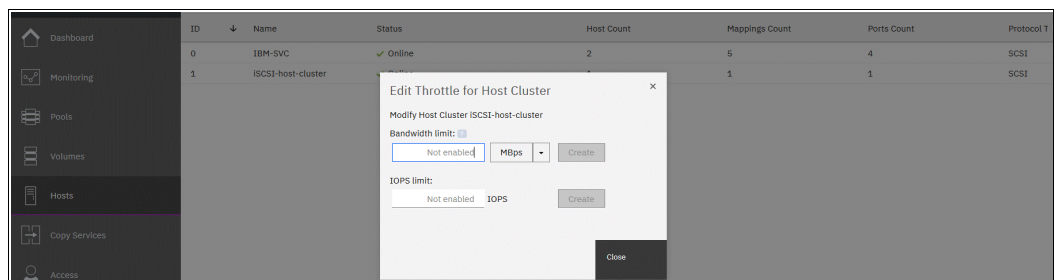


Figure 7-63 Creating a host cluster throttle

Using the **View All Throttles** action on any host cluster object opens a list of all host cluster throttles that are configured on the object. You can switch the display to other types of throttles by clicking a drop-down menu next to the **Actions** menu. You can also change the view to see all the system's throttles in one list.

From this view, you can also delete or edit any existing throttle by right-clicking it in the list and selecting the required action. An example of the **View All Throttles** window is shown in Figure 7-46 on page 445.

Delete Host Cluster

With the **Delete Host Cluster** action, a cluster object is removed and all hosts that were assigned to it become stand-alone hosts. When a cluster is removed, there are two options available:

- ▶ Keep the volume mappings by converting them from shared to private mappings for each host.
- ▶ Remove all shared mappings before deleting the host object.

An example of the Delete Host Cluster window is shown in Figure 7-64. You can hover your mouse pointer over the question marks that are next to the suggested removal options to get more details.

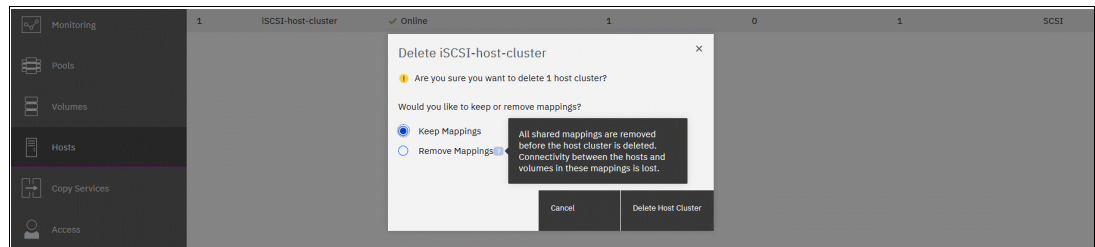


Figure 7-64 Removing a host cluster

7.5.5 Host management views

The Hosts menu provides four more management views: **Ports by Host**, **Mappings**, **Volumes by Host**, and **Volumes by Host Cluster**, as shown in Figure 7-65.

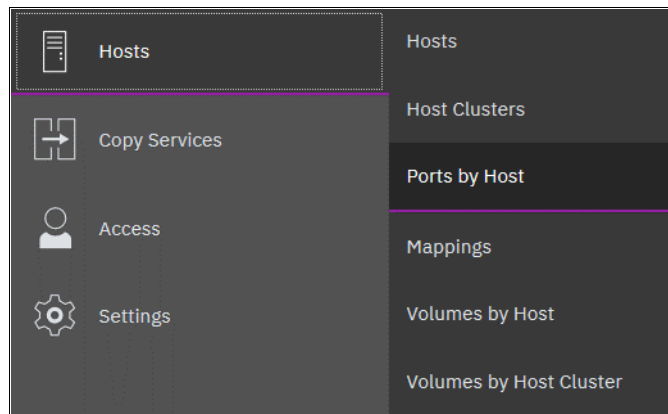


Figure 7-65 Host management views

The actions that are performed from those views are the same ones that can be done from the Hosts and Host Clusters views. However, depending on your current administration task and the size of your configuration, they can provide a better view and are more convenient.

Ports by Host view

This view provides a list of configured host objects. You can use this view to easily manage ports that are assigned to host objects. An example of this view is shown in Figure 7-66.

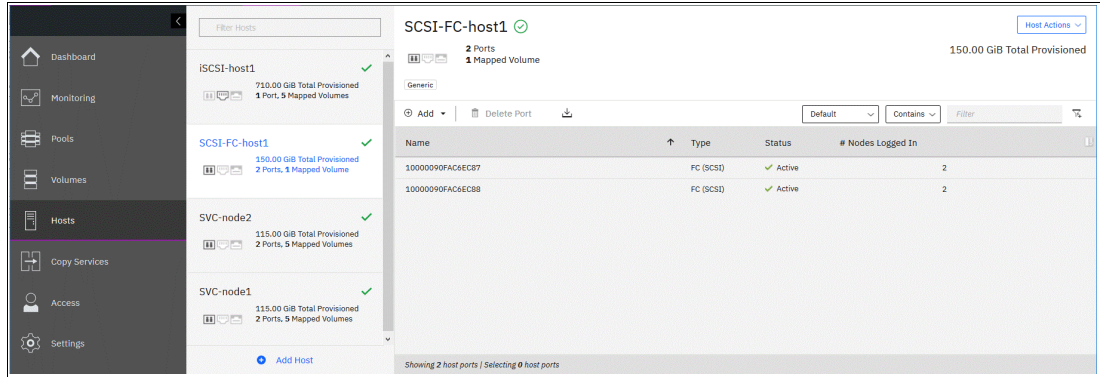


Figure 7-66 Ports by Host view

The left column shows all the configured hosts. At the top of the column is a text input string that can perform quick filtering by hostnames. Below the list of hosts there is an **Add Host** button, which opens a dialog box that is described in step 2 on page 419.

For each host in the list, the following data is shown:

- ▶ Hostname
- ▶ Host status icon (green check mark indicates that host is online)
- ▶ Host connection media icon (FC, Ethernet, or SAS)
- ▶ Total capacity provisioned to a host
- ▶ Number of host ports
- ▶ Number of volume mappings (both private and shared)

Main window shows a list of ports that are assigned to selected host. In the upper right, there is a **Host Actions** drop-down menu that provides same set of actions as it was described in 7.5.3, “Actions on hosts” on page 437.

The list of host ports shows the type and status for each port, as shown in Figure 7-67.

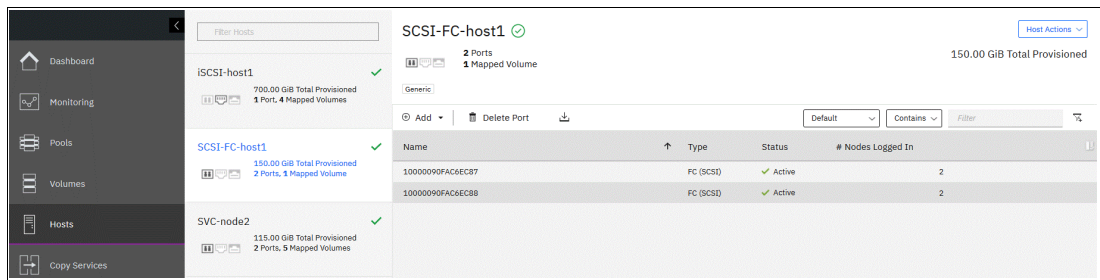


Figure 7-67 Host ports and their statuses

Adding a host port

To add a host port, complete the following steps:

1. Select the host in the left column.
2. Click **Add** (see Figure 7-68). A drop-down menu appears that when clicked shows you a list of the port types that can be assigned to the selected host.

Note: If a host record already has ports that are assigned to it, you can add only ports of the same type to it. For example, you can add an iSCSI port to a host with iSCSI ports, but you cannot add an FC-SCSI port.

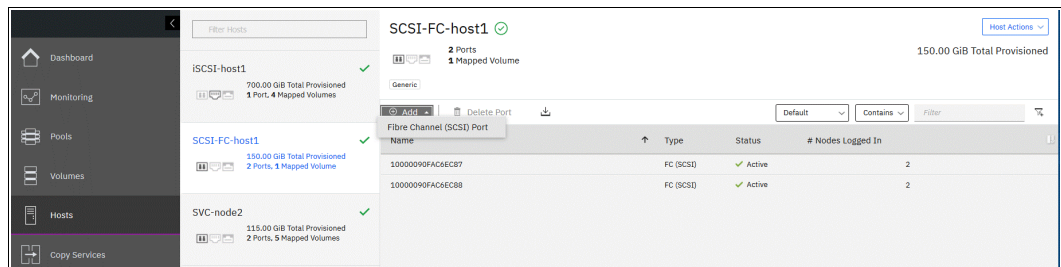


Figure 7-68 Adding host ports

3. The Add Ports dialog box opens (see Figure 7-69).

If you are adding an FC-SCSI port, you can click the drop-down menu to open a list of all discovered FC WWPNs. If the WWPN of your host is not available in the menu, check the SAN zoning to ensure that the connectivity is configured, and click **Rescan**. You can also enter the WWPN manually.

If you are adding a FC-NVMe port or iSCSI port, there is no list of discovered host ports. You must enter the host NQN or IQN manually.

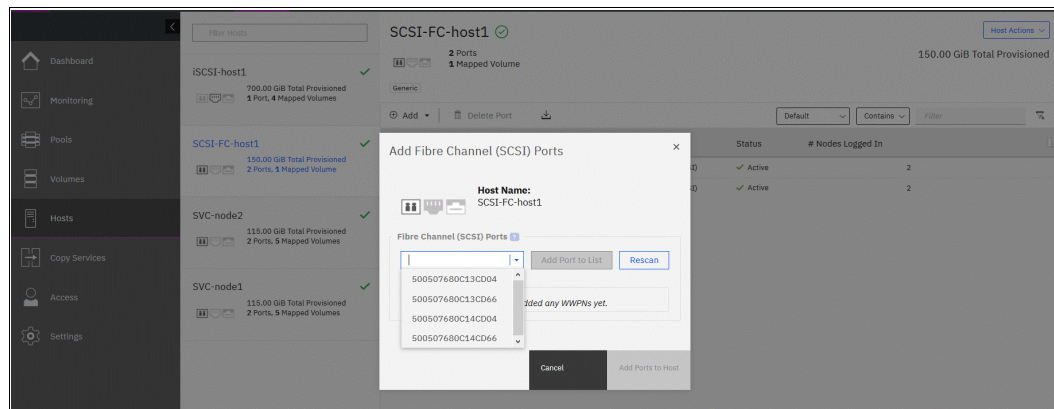


Figure 7-69 Add Fibre Channel (SCSI) Ports window

4. Select the discovered port from the list or enter the port address manually and click **Add Port to List**.

If the FC-SCSI port WWPN is not logged in to the system and its address was entered manually, it is shown as *unverified* in the list, as shown in Figure 7-70 on page 457. The first time that the port logs on, its state is automatically changed to Online.

For other host types, no automatic port verification is performed.

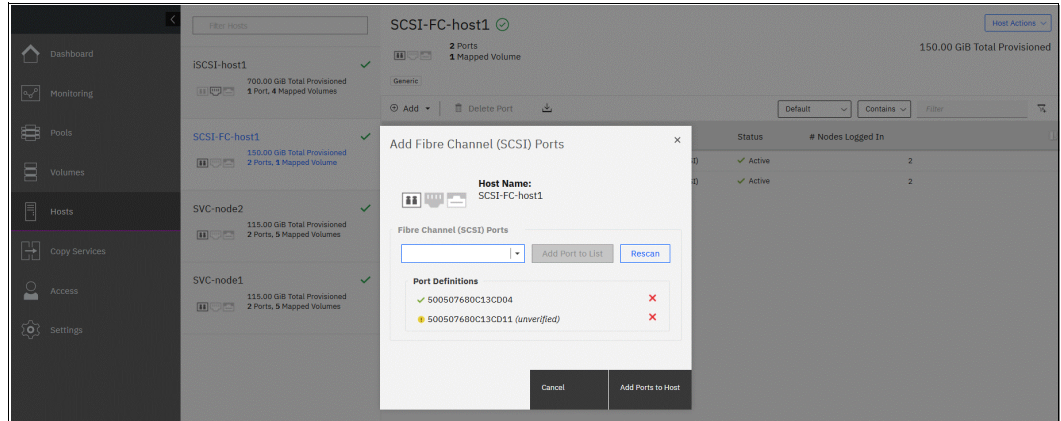


Figure 7-70 Unverified port

5. To remove a port from the list, click the red **X** next to the port.
6. After the list contains all the ports that you want to add, click **Add Ports to Host** to apply the changes.

Deleting a host port

To delete a host port or ports, complete the following steps:

1. Highlight the single host port or select multiple ports in the list, and click **Delete Port**, or right-click any of them and click **Delete Port** (see Figure 7-71). To select multiple host ports, use the Ctrl or Shift key.

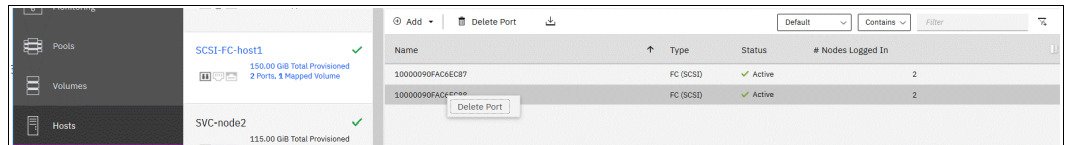


Figure 7-71 Deleting the host port

2. Click **Delete** and confirm the number of host ports that you want to remove by entering that number into the **Verify** field (see Figure 7-72).

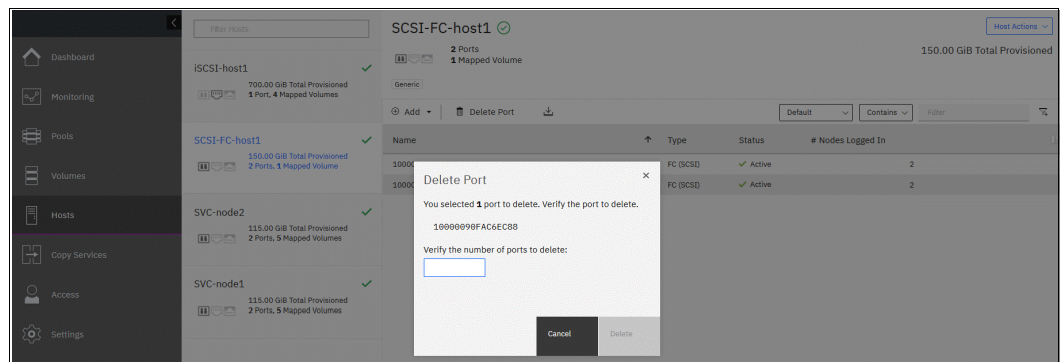


Figure 7-72 Entering the number of host ports to delete

3. Click **Delete** to apply the changes, and then click **Close**.

Mappings view

To see an overview of the host mappings, select **Hosts** → **Mappings**. This view lists all volume-to-host mappings in the system, and shows the hosts, mapped volumes and their SCSI IDs, volume unique identifiers (UIDs), and mapping types. You also see in which I/O group the mapping exists.

By using a drop-down menu in the upper left, you can switch between listing only private mappings, only shared mappings, and all host mappings. The “Private mappings” and “All Host mappings” views show the hosts, and switching to “Shared mappings” shows a list of host clusters and their mappings. Examples of these views are shown in Figure 7-73 and Figure 7-74.

Host Name	SCSI ID	Volume Name	UID	I/O Group ID	I/O Group Name
ISCSI-host1	0	iscsi-host-vol0	600507640084031D0800000000000060	0	io_grp0
ISCSI-host1	1	iscsi-host-vol1	600507640084031D0800000000000061	0	io_grp0
SCSI-FC-host1	2	iscsi-cluster-vol	600507640084031D0800000000000065	0	io_grp0

Figure 7-73 Private mappings list

Host Cluster Name	SCSI ID	Volume Name	UID	I/O Group ID	I/O Group Name
ISCSI-host-cluster	3	iscsi-host-vol3	600507640084031D0800000000000067	0	io_grp0
ISCSI-host-cluster	2	iscsi-host-vol2	600507640084031D0800000000000066	0	io_grp0
IBM-SVC	4	SVC_Volume4	600507640084031D080000000000004D	0	io_grp0
IBM-SVC	0	SVC_Volume0	600507640084031D0800000000000049	0	io_grp0
IBM-SVC	2	SVC_Volume2	600507640084031D080000000000004B	0	io_grp0
IBM-SVC	3	SVC_Volume3	600507640084031D080000000000004C	0	io_grp0
IBM-SVC	1	SVC_Volume1	600507640084031D080000000000004A	0	io_grp0

Figure 7-74 Shared mappings list

If you select a line and click **Actions**, or right-click a mapping in the list, the following tasks are available:

- ▶ Unmap Volumes
- ▶ Host Properties
- ▶ Volume Properties

Unmapping a volume

This action removes the mappings for all selected entries. An unmap action is allowed for shared mappings if you select the **Shared mappings** view, as shown in Figure 7-74. If you select **Private mappings** or **All Host mappings** view, you can remove only private mappings.

To remove a volume mapping or mappings, select the records to remove, right-click, and select **Unmap volumes**, or select **Unmap Volumes** from the **Actions** menu. You can see an example in Figure 7-75 on page 459.

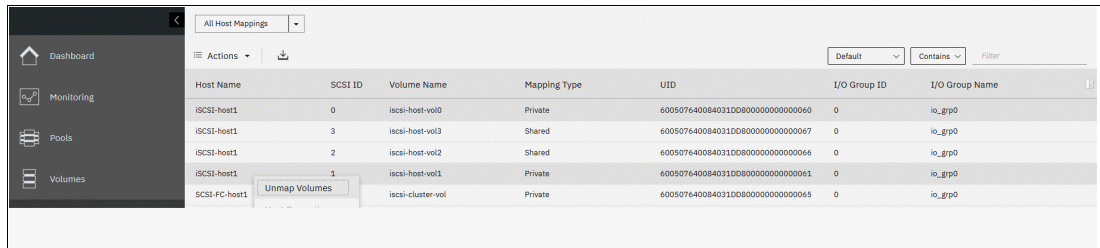


Figure 7-75 Removing two private mappings

A dialog box opens. Confirm how many volumes are to be unmapped by entering that number into the **Verify** field (see Figure 7-76), and then click **Unmap**.

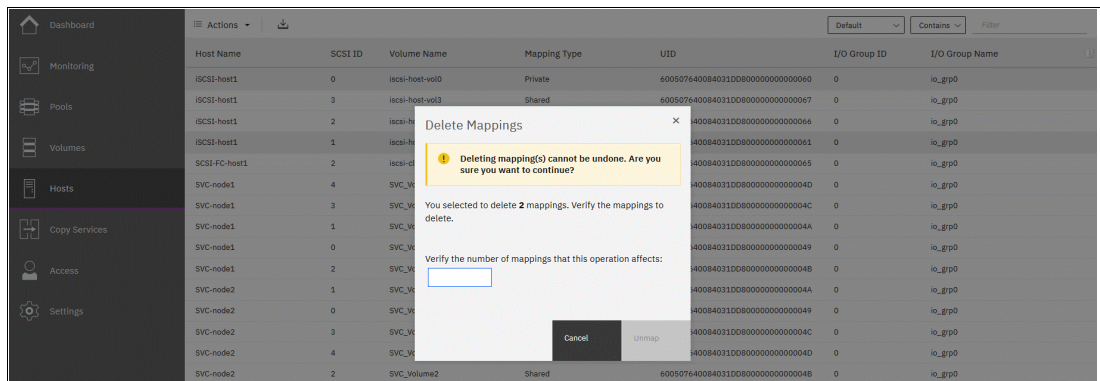


Figure 7-76 Confirming the mapping deletion

Host Properties

Select a single entry and click **Actions** → **Host Properties**. The Host Properties window opens. The contents of this window are described in “Viewing the host properties” on page 447.

Volume Properties

Select an entry and select **Actions** → **Volume Properties**. The Volume Properties view opens. The contents of this window are described in Chapter 6, “Volumes” on page 299.

Volumes by Host and Volumes by Host Cluster

If you need a convenient way to manage volumes that are mapped to a particular host or host cluster, select either **Hosts** → **Volumes by Host** or **Hosts** → **Volumes by Host Cluster**. In contrast to the **Mappings** view, these views focus on volume management.

The left column shows all configured hosts or host clusters. At top of the column is a text input string that you can use to perform quick filtering by object names. Below the list, there is an **Add Host (Create Host Cluster)** button, which opens a dialog box that is described in 7.5.1, “Creating hosts” on page 419 and in 7.5.2, “Host clusters” on page 433.

The main window shows a list of volumes and their parameters that are mapped to the selected object, as shown in Figure 7-77. The **Volumes by Host** view shows volumes that are mapped with both private and shared mappings. The **Volumes by Host Cluster** view shows only volumes with shared mappings.

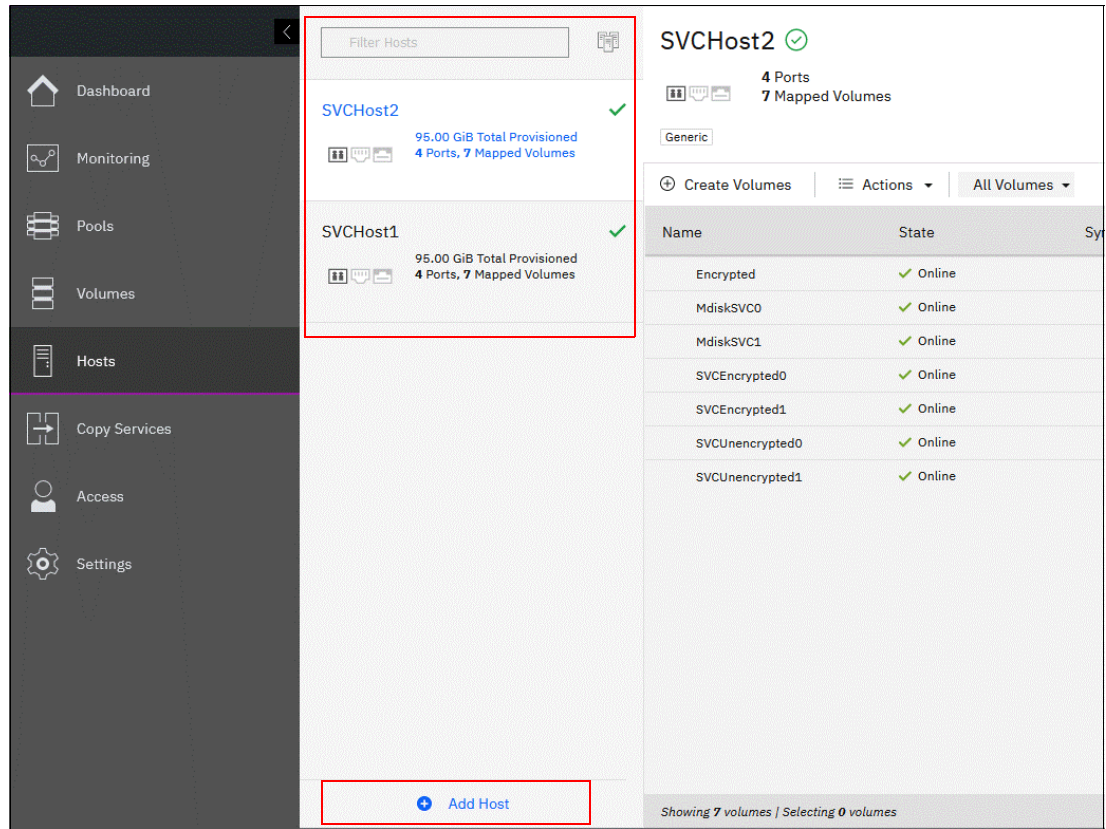


Figure 7-77 Volumes by Host menu

You can also filter by the type of volume by selecting an option from the **Volumes** menu. The options are as follows:

- ▶ **All Volumes**
- ▶ **Thin-Provisioned Volumes**
- ▶ **Compressed Volumes**
- ▶ **Deduplicated Volumes**

Right-clicking a volume in the list opens the **Volume Actions** menu, which is covered in Chapter 6, “Volumes” on page 299. Finally, you can create and map a volume by clicking **Create Volumes**.

7.6 Performing hosts operations by using CLI

This section describes some of the host-related actions that can be taken within the system by using the CLI.

7.6.1 Creating a host by using the CLI

This section describes how to create FC and iSCSI hosts by using the CLI. It is assumed that the hosts are prepared for attachment as noted in the guidelines in the “Host Attachment” section of [IBM Documentation](#).

Creating Fibre Channel hosts

To create an FC host, complete the following steps:

1. Rescan the SAN on the system by running the **detectmdisk** command (Example 7-14).

Example 7-14 Rescanning the SAN

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>detectmdisk
```

Note: The **detectmdisk** command does not return any response.

If zoning was implemented correctly, any new WWPNs are discovered by the system after running the **detectmdisk** command.

2. List the candidate WWPNs and identify the WWPNs belonging to the new host, as shown in Example 7-15.

Example 7-15 Available WWPNs

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>lsfcportcandidate
fc_WWPN
2100000E1E09E3E9
2100000E1E30E5E8
2100000E1E30E60F
2100000E1EC2E5A2
2100000E1E30E597
2100000E1E30E5EC
```

3. Run the **mkhost** command with the required parameters, as shown in Example 7-16.

Example 7-16 Host creation

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>mkhost -name FC-SCSI-HOST-03 -fcwwpn
2100000E1E30E597:2100000E1E30E5EC
Host, id [3], successfully created
```

Creating iSCSI hosts

Before you create an iSCSI host in an IBM FlashSystem system, you must find out the IQN address of the host. To find the IQN of the host, see your host OS-specific documentation.

To create a host, complete the following steps:

1. Create the iSCSI host by running the **mkhost** command (see Example 7-17).

Example 7-17 Creating an iSCSI host by running the mkhost command

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>mkhost -iscsiname
iqn.1994-05.com.redhat:e6ff477b58 -name RHEL-Host-04
Host, id [4], successfully created
```

2. The iSCSI host can be verified by running the **lshost** command, as shown in Example 7-18.

Example 7-18 Verifying the iSCSI host by running the lshost command

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>lshost 4
id 4
name RHEL-Host-04
port_count 1
type generic
....
status_site all
iscsi_name iqn.1994-05.com.redhat:e6ff477b58
node_logged_in_count 1
state active
```

Note: When the host is initially configured, the default authentication method is set to no authentication, and no CHAP secret is set. To set a CHAP secret for authenticating the iSCSI host with the system, run the **chhost** command with the **chapsecret** parameter. If you must display a CHAP secret for a defined server, run the **lsiscsiauth** command. The **lsiscsiauth** command lists the CHAP secret that is configured for authenticating an entity to the system.

FC hosts and iSCSI hosts are handled in the same way operationally after they are created.

Creating NVMe hosts

Before you create an NVMe host, you must know the NQN address of the host. See your host OS-specific documentation to find the NQN of the host.

Create a host by completing the following steps:

1. Create the NVMe host by running the **mkhost** command, as shown in Example 7-19.

Example 7-19 The mkhost command

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>mkhost -name NVMe-Host-01 -nqn
nqn.2014-08.com.redhat:nvme:nvm-nvmehost01-edf223876 -protocol nvme -type
generic
Host, id [6], successfully created
```

2. The NVMe host can be verified by running the **lshost** command, as shown in Example 7-20.

Example 7-20 The lshost command

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>lshost 6
id 6
name NVMe-Host-01
```

```
port_count 1
...
status_site all
nqn nqn.2014-08.com.redhat:nvme:nvm-nvmehost01-edf223876
node_logged_in_count 2
state active
```

Note: If you have OBAC set up, you can use the `-ownershipgroup` parameter when creating a host to add the host to a pre-configured ownership group. You can use either the ownership group name or ID. Here is an example command:

```
mkhost -name NVMe-Host-01 -nqn
nqn.2014-08.com.redhat:nvme:nvm-nvmehost01-edf223876 -protocol nvme -type
generic -ownershipgroup ownershipgroup0
```

7.6.2 Host administration by using the CLI

This section describes the following advanced host operations, which can be carried out by using the CLI:

- ▶ Mapping a volume to a host
- ▶ Mapping a volume that is already mapped to a different host
- ▶ Unmapping a volume from a host
- ▶ Renaming a host
- ▶ Removing a host
- ▶ Host properties

Mapping a volume to a host

To map a volume, complete the following steps:

1. To map a volume to a host, run the `mkvdiskhostmap` command, as shown in Example 7-21.

Example 7-21 Mapping a volume

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>mkvdiskhostmap -host RHEL-HOST-01
-scsi 0 RHEL_VOLUME
Virtual Disk to Host map, id [0], successfully created
```

2. The volume mapping can be checked by running the `lshostvdiskmap` command against that host, as shown in Example 7-22.

Example 7-22 Checking the mapped volume

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>lshostvdiskmap RHEL-HOST-01
id name          SCSI_id vdisk_id vdisk_name .. mapping_type
0 RHEL-HOST-01 0        109      RHEL_VOLUME .. private
```

Mapping a volume that is already mapped to a different host

To map a volume to another host that is mapped to a different host, complete the following steps:

1. Run the `mkvdiskhost -force` command, as shown in Example 7-23.

Example 7-23 Mapping the same volume to a second host

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>svctask mkvdiskhostmap -force -host
RHEL-Host-06 -scsi 0 RHEL_VOLUME
Virtual Disk to Host map, id [0], successfully created
```

Note: The volume RHEL_VOLUME is mapped to both of the hosts by using the same SCSI ID. Typically, that is the requirement for most host-based clustering software, such as Microsoft Clustering Service, IBM PowerHA, and VMware ESX clustering.

2. The volume RHEL_VOLUME is mapped to two hosts (RHEL-HOST-01 and RHEL-Host-06), and can be seen by running the `lsvdiskhostmap` command, as shown in Example 7-24.

Example 7-24 Ensuring that the same volume is mapped to multiple hosts

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>lsvdiskhostmap RHEL_VOLUME
id name          SCSI_id host_id host_name  .. IO_group_name mapping_type
0  RHEL_VOLUME 0          0      RHEL-HOST-01 .. io_grp0      private
0  RHEL_VOLUME 0          1      RHEL-Host-06 .. io_grp0      private
IBM_IBM FlashSystem:ITS0-FS7200:superuser>
```

Unmapping a volume from a host

To unmap a volume from the host, run the `rmvdiskhostmap` command, as shown in Example 7-25.

Example 7-25 Unmapping a volume from a host

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>rmvdiskhostmap -host RHEL-Host-06 RHEL_VOLUME
```

Important: Before unmapping a volume from a host on an IBM FlashSystem system, ensure that the host side action is completed on that volume by using the respective host OS platform commands, such as unmounting the file system or removing the volume or volume group. Otherwise, it can result in data corruption.

Renaming a host

To rename a host definition, run the `chhost -name` command, as shown in Example 7-26. In this example, the host RHEL-Host-06 is renamed to FC_RHEL_HOST.

Example 7-26 Renaming a host

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>chhost -name FC_RHEL_HOST RHEL-Host-06
```

Removing a host

To remove a host from the IBM FlashSystem system, run the `rmhost` command, as shown in Example 7-27.

Example 7-27 Removing a host

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>rmhost RHEL-Host-07
```


Note: Before removing a host from an IBM FlashSystem system, ensure that all of the volumes are unmapped from that host, as shown in Example 7-25.

Host properties

To get more information about a host, run the **lshost** command with **hostname** or **host id** as a parameter, as shown in Example 7-28.

Example 7-28 Host details

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>lshost ITS0-VMHOST-01
id 10
name ITS0-VMHOST-01
port_count 2
type generic
mask 1111111111111111111111111111111111111111111111111111111111111111
iogrp_count 4
status offline
site_id
site_name
host_cluster_id 0
host_cluster_name ITS0-ESX-Cluster-01
protocol scsi
status_policy redundant
status_site all
WWPN 2100000E1E30E597
node_logged_in_count 0
state offline
WWPN 2100000E1E30E5E8
node_logged_in_count 0
state offline
owner_id 0
IBM_IBM
FlashSystem:ITS0-FS7200:superuser>
```

Note: Starting from code release 8.3.0.0, the new **status_policy** property was added to each host. The property has two potential values:

- ▶ Complete: The default policy when a host is created. It uses the legacy algorithm. Existing hosts on systems that are upgraded to a new code level have this policy set.
- ▶ Redundant: This policy changes the meaning of **Online** and **Degraded** in the **status** property:
 - **Online** indicates redundant connectivity, that is, enough host ports are logged in to enough nodes so that the removal of a single node or a single host port still enables that host to access all its volumes.
 - **Degraded** indicates non-redundant connectivity, that is, a state in which a single point of failure (SPOF) prevents a host from accessing at least some of its volumes.

These options can be changed only by running the **chhost** command. When the host is created by running **mkhost**, the default policy of **redundant** is set.

7.6.3 Adding and deleting a host port by using the CLI

This section describes adding and deleting a host port to and from the system.

Adding ports to a defined host

To add ports to a defined host, complete the following steps:

- ▶ For FC-SCSI host ports:
 - a. If the host is connected through SAN with FC, and if the WWPN is zoned to the system, you can run the **lsfcportcandidate** command to compare it with the information that is available from the server administrator.

Example 7-29 Listing the newly available WWPN

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>lsfcportcandidate
fc_WWPN
2100000E1E09E3E9
2100000E1E30E5E8
2100000E1E30E60F
2100000E1EC2E5A2
```

- b. Use host or SAN switch utilities to verify whether the WWPN matches the information for the new WWPN. If the WWPN matches, run the **addhostport** command to add the port to the host, as shown in Example 7-30.

Example 7-30 Adding the newly discovered WWPN to the host definition

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>addhostport -hbawpn
2100000E1E09E3E9:2100000E1E30E5E8 ITS0-VMHOST-01
```

This command adds the WWPNs 2100000E1E09E3E9 and 2100000E1E30E5E8 to the ITS0-VMHOST-01 host.

- c. If the new HBA is not connected or zoned, the **lshbaportcandidate** command does not display your WWPN. In this case, you can manually enter the WWPN of your HBA or HBAs and use the **-force** flag to create the host, as shown in Example 7-31.

Example 7-31 Adding a WWPN to the host definition by using the -force option

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>addhostport -hbawpn
2100000000000001 -force ITS0-VMHOST-01
```

This command forces the addition of the WWPN 2100000000000001 to the host ITS0-VMHOST-01.

Note: WWPNs are not case-sensitive within the CLI.

- d. Verify the host port count by running the **lshost** command. Example 7-32 shows that the host ITS0-VMHOST-01 has a port count that updated from 2 to 5 after two commands in previous examples ran.

Example 7-32 Host with the updated port count

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>lshost
id name          port_count iogrp_count status  site_id site_name
0  ITS0-VMHOST-01  5          4          online
```

- ▶ For iSCSI and FC-NVMe host ports:
 - a. If the host uses iSCSI or FC-NVMe as a connection method, the host port ID (iSCSI IQN or NVMe NQN) is used to add the port. Unlike FC-attached hosts, the available candidate IDs cannot be checked. Your host administrator provides you with the IQN or NQN.
 - b. After getting the ID, run the **addhostport** command. Example 7-33 shows a command for an iSCSI port.

Example 7-33 Adding an iSCSI port to the defined host

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>addhostport -iscsiname
iqn.1994-05.com.redhat:e6ddffaab567 RHEL-Host-05
```

Example 7-34 shows the FC-NVMe port being added.

Example 7-34 The addhostport command

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>addhostport -nqn
nqn.2016-06.io.rhel:875adad3345 RHEL-Host-08
```

Deleting ports from a defined host

If a host port record must be removed from a host object, run the **rmhostport** command.

To perform the removal procedure, complete the following steps:

1. Ensure that it is the correct port being removed by running the **lshost** command. In Example 7-35, a check ID performed to verify that the WWPN to be removed belongs the host ITS0-VMHOST-01.

Example 7-35 Running the lshost command to check the WWPNs

```
IBM_2145:ITS0-SV1:superuser>lshost ITS0-VMHOST-01
id 0
name ITS0-VMHOST-01
port_count 2
...
WWPN 2100000E1E30E597
node_logged_in_count 2
state online
WWPN 2100000E1E30E5E8
node_logged_in_count 2
state online
```

2. When you discover the WWPN or iSCSI IQN that must be deleted, run the **rmhostport** command to delete the host port, as shown in Example 7-36.

Example 7-36 Running the rmhostport command to remove a WWPN

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>rmhostport -fcwwpn 2100000E1E30E597
ITS0-VMHOST-01
```

To remove the iSCSI IQN, run the **rmhostport** command with the **iscsiname** argument, as shown in Example 7-37.

Example 7-37 Removing the iSCSI port from the host

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>rmhostport -iscsiname
iqn.1994-05.com.redhat:e6ddffaab567 RHEL-Host-05
```

3. To remove the NVMe NQN, run the **rmhostport** with the **nqn** argument, as shown in Example 7-38.

Example 7-38 Removing the NQN port from the host

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>rmhostport -nqn
nqn.2016-06.io.rhel:875adad3345 RHEL-Host-08
```

Note: Multiple ports can be removed at once by using the separator or colon (:) between the port names, as shown in the following example:

```
rmhostport -hbawpn 210000E08B054CAA:210000E08B892BCD ITS0-VMHOST-02
```

7.6.4 Host cluster operations

This section describes the following host cluster operations that can be performed by using the CLI:

- ▶ Creating a host cluster (**mkhostcluster**).
- ▶ Adding a member to the host cluster (**addhostclustermember**).
- ▶ Listing a host cluster (**lshostcluster**).
- ▶ Listing a host cluster member (**lshostclustermember**).
- ▶ Assigning a volume to the host cluster (**mkvolumehostclustermap**).
- ▶ Listing a host cluster for mapped volumes (**lshostclustervolumemap**).
- ▶ Removing a volume mapping from the host cluster (**rmvolumehostclustermap**).
- ▶ Removing a host cluster member (**rmhostclustermember**).
- ▶ Removing the host cluster (**rmhostcluster**).

Creating a host cluster

To create a host cluster, run the **mkhostcluster** command, as shown in Example 7-39.

Example 7-39 Creating a host cluster

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>mkhostcluster -name ITS0-ESX-Cluster-01
Host cluster, id [0], successfully created.
```

Adding a host to a host cluster

After creating a host cluster, a host or a list of hosts can be added by running the **addhostclustermember** command, as shown in Example 7-40.

Example 7-40 Adding a host or hosts to a host cluster

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>addhostclustermember -host
ITS0-VMHOST-01:ITS0-VMHOST-02 ITS0-ESX-Cluster-01
IBM_IBM FlashSystem:ITS0-FS7200:superuser>
```

In Example 7-40, the hosts ITS0-VMHOST-01 and ITS0-VMHOST-02 were added as part of host cluster ITS0-ESX-Cluster-01.

Listing the host cluster member

To list the host members that are part of a particular host cluster, run the `lshostclustermember` command, as shown in Example 7-41.

Example 7-41 Listing host cluster members by running the `lshostclustermember` command

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>lshostclustermember ITS0-ESX-Cluster-01
host_id host_name      status type   site_id site_name
0       ITS0-VMHOST-01 offline generic
4       ITS0-VMHOST-02 offline generic
IBM_IBM FlashSystem:ITS0-FS7200:superuser>
```

Mapping a volume to a host cluster

To map a volume to a host cluster so that it automatically is mapped to member hosts, run the `mkvolumehostclustermap` command, as shown in Example 7-42.

Example 7-42 Mapping a volume to a host cluster

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>mkvolumehostclustermap -hostcluster
ITS0-ESX-Cluster-01 VMware1
Volume to Host Cluster map, id [0], successfully created
IBM_IBM FlashSystem:ITS0-FS7200:superuser>
```

Note: When a volume is mapped to a host cluster, that volume is mapped to all of the members of the host cluster with the same SCSI_ID.

Listing the volumes that are mapped to a host cluster

To list the volumes that are mapped to a host cluster, run the `lshostclustervolumemap` command, as shown in Example 7-43.

Example 7-43 Listing volumes that are mapped to a host

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>lshostclustervolumemap
ITS0-ESX-Cluster-01
id name                SCSI_id volume_id volume_name .. protocol
0 ITS0-ESX-Cluster-01 0      8      VMware1    .. scsi
0 ITS0-ESX-Cluster-01 1      9      VMware2    .. scsi
0 ITS0-ESX-Cluster-01 2      10     VMware3    .. scsi
```

Note: You can run the `lshostvdiskmap` command against each host that is part of a host cluster to ensure that the mapping type for the shared volume is shared, and that the non-shared volume is private.

Removing a volume mapping from a host cluster

To remove a volume mapping to a host cluster, run the `rmvolumehostclustermap` command, as shown in Example 7-44.

Example 7-44 Removing a volume mapping

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>rmvolumehostclustermap -hostcluster
ITS0-ESX-Cluster-01 VMware3
```

In Example 7-44, volume VMware3 is unmapped from the host cluster ITS0-ESX-Cluster-01.

Note: To specify the host or hosts that acquire private mappings from the volume that is being removed from the host cluster, use the **-makeprivate** flag.

Removing a host cluster member

To remove a host cluster member, run the **rmhostcluster** command, as shown in Example 7-45.

Example 7-45 Removing a host cluster member

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>rmhostcluster -host ITS0-VMHOST-02  
-removemappings ITS0-ESX-Cluster-01
```

In Example 7-45, the host ITS0-VMHOST-02 was removed as a member from the host cluster ITS0-ESX-Cluster-01, along with the associated volume mappings because the **-removemappings** flag was specified.

Removing a host cluster

To remove a host cluster, run the **rmhostcluster** command, as shown in Example 7-46.

Example 7-46 Removing a host cluster

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>rmhostcluster -removemappings ITS0-ESX-Cluster-01
```

Using the **-removemappings** flag also causes the system to remove any shared host mappings to volumes. The mappings are deleted before the host cluster is deleted.

Note: To keep the volumes mapped to the host objects even after the host cluster is deleted, use the **-keepmappings** flag instead of **-removemappings** for the **rmhostcluster** command. When **-keepmappings** is specified, the host cluster is deleted, but the volume mapping to the host becomes private instead of shared.

7.6.5 Adding a host or host cluster to an ownership group

To add a host or a host cluster to an ownership group, run the **chhost** or **chhostcluster** command with the **-ownershipgroup** parameter, as shown in Example 7-47.

Example 7-47 Adding a host cluster to an ownership group

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>chhostcluster -ownershipgroup 1 0
```

Note: You must specify the ID of the ownership group to which you want to add the host, and then specify the ID of the host or host cluster. So, the command in Example 7-47 adds host cluster ID 0 to ownership group ID 1.

Removing a host or host cluster from an ownership group

To remove a host or a host cluster from an ownership group, run the **chhost** or **chhostcluster** command with the **-noownershipgroup** parameter, as shown in Example 7-48.

Example 7-48 Removing a host cluster from an ownership group

```
IBM_IBM FlashSystem:ITS0-FS7200:superuser>chhostcluster -noownershipgroup 0
```

This command removes host cluster 0 from the ownership group that it is assigned to.

7.7 Host attachment practical examples

This section provides practical examples for Linux based host attachment that are implemented by using the information that is provided in the previous sections of this chapter.

7.7.1 Prerequisites

The host should be running on the supported OS, which in this example is Red Hat Enterprise Linux (RHEL), and supported HBAs.

In the case of RHEL, it is possible to check the OS level by running the command that is shown in Example 7-49.

Example 7-49 RHEL release check

```
20201028-09:50:34 root@redbookvm7-1:~ # cat /etc/redhat-release
Red Hat Enterprise Linux Server release 7.6 (Maipo)
```

7.7.2 Fibre Channel host connectivity and capacity allocation

To collect the necessary data, configure the host object in the storage system, and get access to the storage capacity, complete the following steps:

1. Obtain the necessary credentials for connectivity from the host. In this case, we need the WWPN of the host HBAs. You can obtain the WWPN in RHEL by running the command that is shown in Example 7-50. In the example, the information about the HBAs on the host with FC connectivity ability is available in the `/sys/class/fc_host` directory. The hosts' WWPNs are in the `port_name` file in each `hostN` directory. The WWPNs are in bold in the example. Use these WWPNs for the host object configuration of storage system.

Example 7-50 Discovering the hosts' WWPNs

```
20201028-10:39:29 root@redbookvm7-1:~ # cd /sys/class/fc_host

20201028-10:40:02 root@redbookvm7-1:/sys/class/fc_host # ls -la
total 0
drwxr-xr-x.  2 root root 0 Oct 26 14:19 .
drwxr-xr-x. 59 root root 0 Oct 26 14:19 ..
lrwxrwxrwx.  1 root root 0 Oct 28 10:01 host33 ->
../../../../devices/pci0000:00/0000:00:17.0/0000:13:00.0/host33/fc_host/host33
lrwxrwxrwx.  1 root root 0 Oct 28 10:01 host34 ->
../../../../devices/pci0000:00/0000:00:17.0/0000:13:00.1/host34/fc_host/host34

20201028-10:45:35 root@redbookvm7-1:/sys/class/fc_host # cat host33/port_name
0x10000090fac6ec87
20201028-10:45:50 root@redbookvm7-1:/sys/class/fc_host # cat host34/port_name
0x10000090fac6ec88
```

2. To configure the host object on the storage system, follow the instructions in “Creating Fibre Channel host objects” on page 420. If zoning is already done for the host, the host's WWPN should be available in the **Host Port (WWPN)** list. If the host is not zoned, it is possible to add ports manually into the field.

3. After the host object is defined, it is visible in the hosts view, and volumes (VDisks) can be mapped to it, as described in Chapter 2, “Planning” on page 71. Details about the host can be found by double-clicking its entry in the Hosts view. After the volumes are mapped to the host, you can check them by going to the Host Details window in the Mapped Volumes tab (see Figure 7-78).

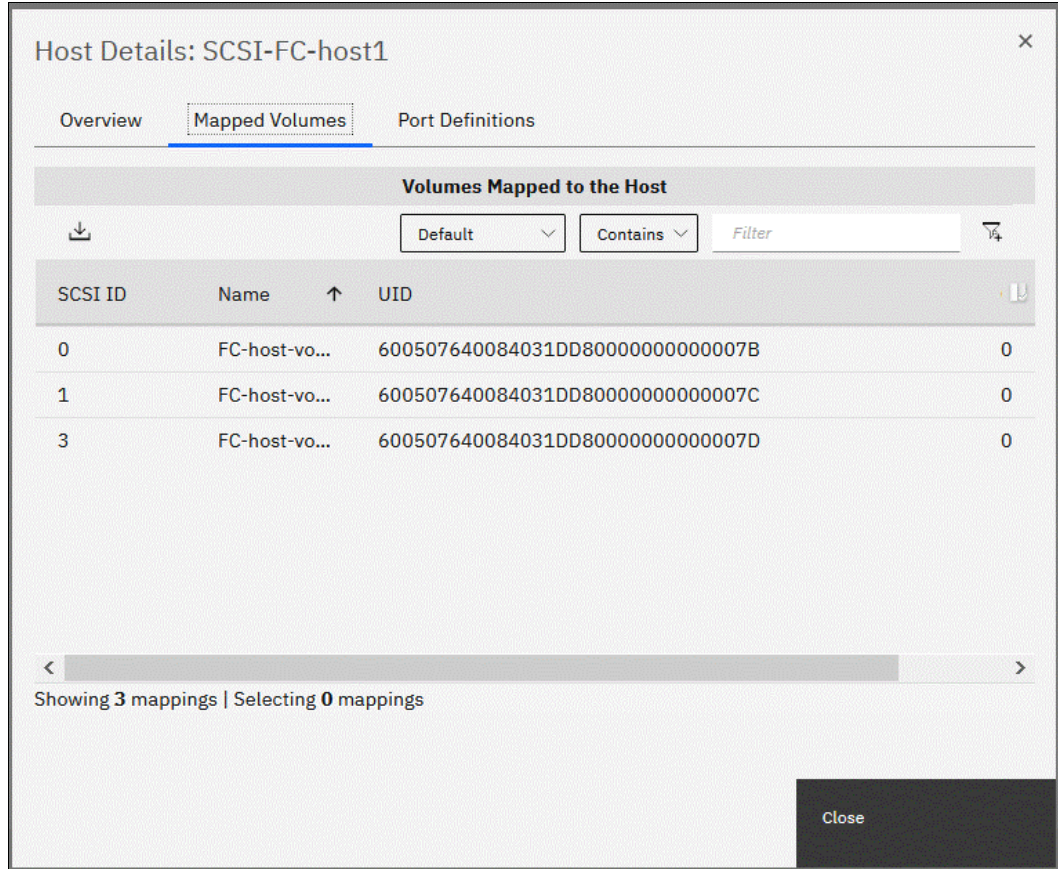


Figure 7-78 Host Details: Mapped volumes

4. Configure the host side to discover the mapped VDisks and use them:
 - a. RHEL has its own native multipath driver, which maps the discovered drives and their paths to the `mpath n` device files in `/dev/mapper`. The multipath driver must be correctly configured, which is described at [IBM Documentation](#). To check that the volumes are detected correctly by the host, run the command in Example 7-51.

Example 7-51 Scanning and rebuilding the multipath

```
20201028-14:19:53 root@redbookvm7-1:/dev # rescan-scsi-bus.sh -r
Syncing file systems
Scanning SCSI subsystem for new devices and remove devices that have
disappeared
Scanning host 0 for SCSI target IDs 0 1 2 3 4 5 6 7, all LUNs
Scanning for device 0 0 0 0 ...
. . .
20201028-14:10:25 root@redbookvm7-1:/dev # multipath -F
20201028-14:10:30 root@redbookvm7-1:/dev # multipath
20201028-14:14:42 root@redbookvm7-1:/dev # multipath -ll
mpathau (3600507640084031dd80000000000007d) dm-4 IBM ,2145
size=100G features='1 queue_if_no_path' hwhandler='0' wp=rw
```



```

|+- policy='service-time 0' prio=50 status=enabled
| | - 33:0:15:3 sdd 8:48 active ready running
| | - 33:0:27:3 sdu 65:64 active ready running
| | - 33:0:28:3 sdaa 65:160 active ready running
| | - 33:0:31:3 sdaf 65:240 active ready running
| | - 34:0:13:3 sdah 66:16 active ready running
| | - 34:0:15:3 sdak 66:64 active ready running
| | - 34:0:1:3 sdv 65:80 active ready running
| | ~- 34:0:3:3 sdac 65:192 active ready running
+- policy='service-time 0' prio=10 status=enabled
| - 33:0:19:3 sdg 8:96 active ready running
| - 33:0:24:3 sdj 8:144 active ready running
| - 33:0:25:3 sdm 8:192 active ready running
| - 33:0:26:3 sdp 8:240 active ready running
| - 34:0:20:3 sdan 66:112 active ready running
| - 34:0:26:3 sdaq 66:160 active ready running
| - 34:0:29:3 sdat 66:208 active ready running
| ~- 34:0:31:3 sdaw 67:0 active ready running
mpathat (3600507640084031dd8000000000007c) dm-3 IBM ,2145
size=100G features='1 queue_if_no_path' hwhandler='0' wp=rw
|+- policy='service-time 0' prio=50 status=enabled
| | - 33:0:19:1 sdf 8:80 active ready running
| | - 33:0:24:1 sdi 8:128 active ready running
| | - 33:0:25:1 sd1 8:176 active ready running
| | - 33:0:26:1 sdo 8:224 active ready running
| | - 34:0:20:1 sdam 66:96 active ready running
| | - 34:0:26:1 sdap 66:144 active ready running
| | - 34:0:29:1 sdas 66:192 active ready running
| | ~- 34:0:31:1 sdav 66:240 active ready running
+- policy='service-time 0' prio=10 status=enabled
| - 33:0:15:1 sdc 8:32 active ready running
| - 33:0:27:1 sds 65:32 active ready running
| - 33:0:28:1 sdy 65:128 active ready running
| - 33:0:31:1 sdad 65:208 active ready running
| - 34:0:13:1 sdag 66:0 active ready running
| - 34:0:15:1 sdaj 66:48 active ready running
| - 34:0:1:1 sdt 65:48 active ready running
| ~- 34:0:3:1 sdz 65:144 active ready running
mpathas (3600507640084031dd8000000000007b) dm-2 IBM ,2145
size=100G features='1 queue_if_no_path' hwhandler='0' wp=rw
|+- policy='service-time 0' prio=50 status=enabled
| | - 33:0:15:0 sdb 8:16 active ready running
| | - 33:0:27:0 sdq 65:0 active ready running
| | - 33:0:28:0 sdw 65:96 active ready running
| | - 33:0:31:0 sdab 65:176 active ready running
| | - 34:0:13:0 sdae 65:224 active ready running
| | - 34:0:15:0 sdai 66:32 active ready running
| | - 34:0:1:0 sdr 65:16 active ready running
| | ~- 34:0:3:0 sdx 65:112 active ready running
+- policy='service-time 0' prio=10 status=enabled
| - 33:0:19:0 sde 8:64 active ready running
| - 33:0:24:0 sdh 8:112 active ready running
| - 33:0:25:0 sdk 8:160 active ready running
| - 33:0:26:0 sdn 8:208 active ready running
| - 34:0:20:0 sda1 66:80 active ready running

```

```

|- 34:0:26:0 sdao 66:128 active ready running
|- 34:0:29:0 sdar 66:176 active ready running
^- 34:0:31:0 sdau 66:224 active ready running

```

- b. As shown in Example 7-51 on page 472, the **rescan-scsi-bus.sh -r** command rescans for new devices. In some cases, it might be necessary to run the **rescan-scsi-bus.sh -a** command because it scans all available devices. The **multipath -F** command flushes the configuration of multipath driver. Then, the **multipath** command builds a new configuration for new devices and paths. The **multipath -ll** command provides information about path states and to which the mpath n device capacity was mapped for each mapped VDisk (see the UUID without digit 3 in the beginning).
- c. To start using the capacities that are provided as logical volumes, use *only* the /dev/mapper/mpath n device for access because it is created early in the RHEL boot process. For example, to use the VDisk 600507640084031dd80000000000007c as a logical volume manager (LVM) physical volume, use the name of the RHEL device (which was mapped by the multipath driver), which you can find by running the **multipath -ll** command. Example 7-51 on page 472 shows the output. mpathat with UID 3600507640084031dd80000000000007c is marked bold in the example.
- d. This new physical volume can be added to the volume group of the host, and logical volumes can be created or extended and configured for any application on the host, as shown in Example 7-52.

Example 7-52 Creating a physical volume in LVM for further use

```

20201028-14:39:10 root@redbookvm7-1:/dev/mapper # pvcreate
/dev/mapper/mpathat
Physical volume "/dev/mapper/mpathat" successfully created.
20201028-14:39:55 root@redbookvm7-1:/dev/mapper # pvs
PV          VG      Fmt Attr PSize  PFree
/dev/mapper/mpathat    lvm2 --- 100.00g 100.00g
/dev/sda2      rhel  lvm2 a--  <15.00g    0

```

Summary

To introduce capacity to the host from the storage system, you must first deal with several abstractions:

1. On the storage system:
 - a. Define the host object definition with all the credentials of the host.
 - b. Map volumes to the defined host object to introduce capacity to the host.
2. On the host:
 - a. The multipathing driver should be configured (usually, the native multipathing driver or device mapper are configured and running in some OSs). It is used to map all paths for the specific volume (VDisk) to the one device because the specifics of the protocol system see each path as a separate device even for the one volume (VDisk). Therefore, the multipath driver is essential for correct representation and usage of the provided capacity resource.
 - b. Set up the LVM layer if you plan to use it for more flexibility.
 - c. Set the file system level, depending on the application.

7.7.3 iSCSI host connectivity and capacity allocation

The iSCSI protocol uses an initiator from host side to send SCSI commands to storage systems' target devices. Therefore, it is necessary to prepare the correct environment on the host side and configure the storage system, as described in "Creating iSCSI host objects" on page 429.

This section demonstrates an RHEL host configuration and how to obtain access to the dedicated volumes (VDisks) on the storage system.

The detailed steps to prepare an RHEL host for SCSI connectivity can be found by going to [IBM Documentation](#), selecting your specific system, and then selecting **Configuring** → **Host Attachment** → **iSCSI Ethernet host attachment**.

Complete the following steps:

1. Install the iSCSI initiator on the RHEL host by running the `yum` command, as shown in Example 7-53.

Example 7-53 Installing iscsi-initiator-utils

```
20201028-18:13:51 root@redbookvm7-1:/mnt/disc # yum install  
iscsi-initiator-utils
```

2. Now, the iSCSI initiator should be configured, and the connection credentials should be set in the `/etc/iscsi` files. Check or define IQN in `/etc/iscsi/initiatorname.iscsi`, as shown in Example 7-54.

Example 7-54 Checking the initiator's IQN

```
20201028-19:04:34 root@redbookvm7-1:/etc/iscsi # cat initiatorname.iscsi  
InitiatorName=iqn.1994-05.com.redhat:f3de6ef11811
```

3. Restart the iSCSI initiator service if the IQN was modified.

- After the host is ready and the iSCSI initiator is configured, define a host object on the storage system, as described in “Creating iSCSI host objects” on page 429. Make sure that the IQN is set correctly in the Host Details window in the Port Definitions tab, as shown in Figure 7-79.

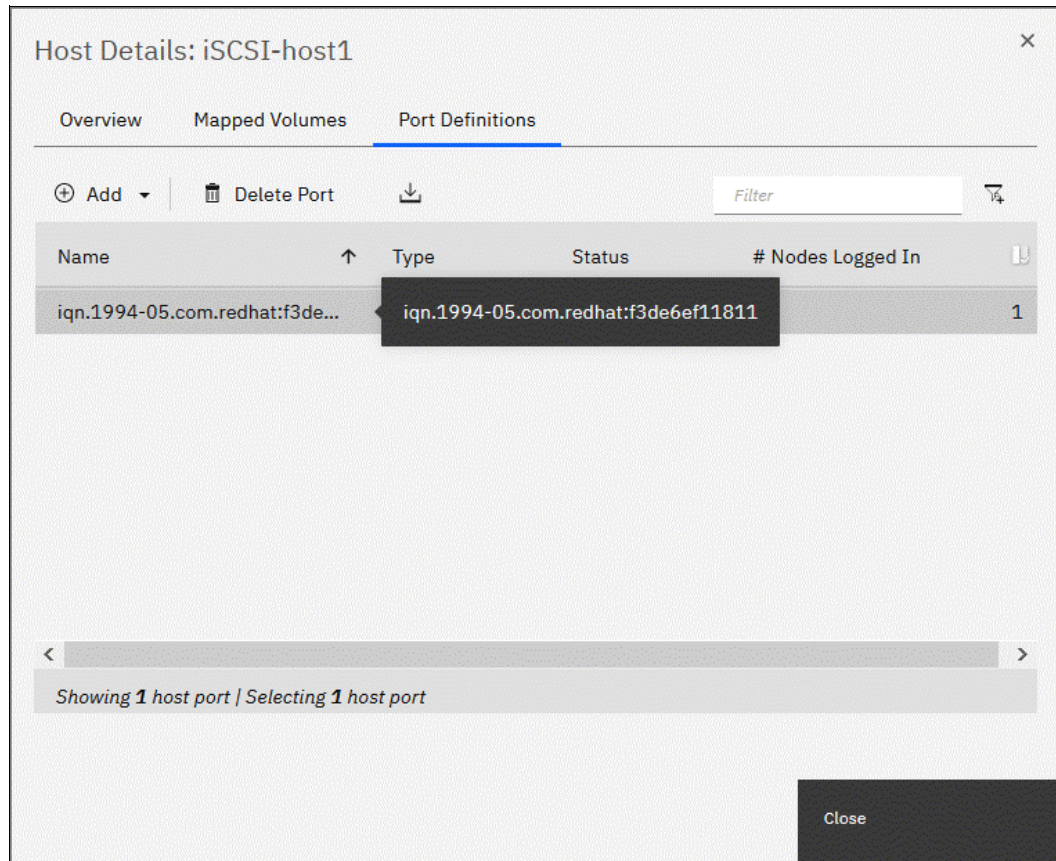


Figure 7-79 Host Details: Port Definitions tab

- Map the dedicated volumes (VDisks) to the host object.
- After the iSCSI host object is configured in the storage system and the volumes (VDisks) are mapped to it, you must discover the iSCSI targets from the host. There are two ways to do this task:
 - Using send targets
 - iSNS

Example 7-55 shows discovery by using the send targets method. Record the IP address under the IP column in Figure 7-80 on page 477, which was configured for an iSCSI connection on the storage system as described in “Creating iSCSI host objects” on page 429 because it is used to find the IQN of our target and for further logins.

Example 7-55 iSCSI targets discovery

```

20201028-19:04:40 root@redbookvm7-1:/etc/iscsi # iscsiadm --mode discovery
--type sendtargets --portal 9.71.42.61
9.71.42.61:3260,1 iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node1
20201028-19:15:53 root@redbookvm7-1:/etc/iscsi # iscsiadm --mode discovery
--type sendtargets --portal 9.71.42.67
9.71.42.67:3260,1 iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node2

```

Name	Port	State	IP	Speed	Host Attach	IPv4 Remote Copy	Storage Port IPv4
~io_grp0							
node1	1	✓ Configured	9.71.42.61	1Gb/s	Yes	Disabled	Disabled
node2	1	✓ Configured	9.71.42.67	1Gb/s	Yes	Disabled	Disabled

Figure 7-80 Ethernet Ports Configuration tab

For the iSNS discovery method, complete the following steps:

1. Update the configuration file `/etc/iscsi/iscsid.conf` by providing the connection credentials of the iSNS server and placing them into the following variable, if it is available in the environment:


```
isns.address = <iSNS server IP address>
isns.port = <iSNS server port>
```
2. Restart the iSCSI initiator service to make the configuration active.
3. Run `iscsiadm --mode discovery --type isns` to generate the list of all iSCSI targets that are registered with the iSNS server.

Finally, to access the volumes (VDisks) space, which is mapped on the storage system to the host object, log in to the discovered targets (Example 7-56).

Example 7-56 Logging in to the discovered targets/storage

```
20201028-19:16:09 root@redbookvm7-1:/etc/iscsi # iscsiadm --mode node --target
iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node1 --portal 9.71.42.61 --login
Logging in to [iface: default, target: iqn.1986-03.com.ibm:2145.ibmIBM
FlashSystem7200.node1, portal: 9.71.42.61,3260] (multiple)
Login to [iface: default, target: iqn.1986-03.com.ibm:2145.ibmIBM
FlashSystem7200.node1, portal: 9.71.42.61,3260] successful.
20201028-19:17:56 root@redbookvm7-1:/etc/iscsi # iscsiadm --mode node --target
iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node2 --portal 9.71.42.67 --login
Logging in to [iface: default, target: iqn.1986-03.com.ibm:2145.ibmIBM
FlashSystem7200.node2, portal: 9.71.42.67,3260] (multiple)
Login to [iface: default, target: iqn.1986-03.com.ibm:2145.ibmIBM
FlashSystem7200.node2, portal: 9.71.42.67,3260] successful.
```

After logging in successfully, make sure that the native multipath driver on the RHEL host is installed and configured correctly similar to the example with the Fibre Channel connection (IBM FICON) in 7.7.2, “Fibre Channel host connectivity and capacity allocation” on page 471, and check the output by running `multipath -ll`.

Example 7-57 shows an example of the output.

Example 7-57 Multipathing driver/device mapper output

```
20201028-19:39:22 root@redbookvm7-1:/etc/iscsi # multipath -ll
mpathaw (3600507640084031dd800000000000060) dm-6 IBM ,2145
size=100G features='1 queue_if_no_path' hwhandler='0' wp=rw
|+- policy='service-time 0' prio=50 status=active
| `-- 35:0:0:0 sdbn 68:16 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  `-- 36:0:0:0 sdbn 68:80 active ready running
mpathaz (3600507640084031dd800000000000067) dm-9 IBM ,2145
```

```

size=250G features='1 queue_if_no_path' hwhandler='0' wp=rw
| -+- policy='service-time 0' prio=50 status=active
|  ~- 36:0:0:3  sdbu 68:128 active ready running
^-+- policy='service-time 0' prio=10 status=enabled
  ~- 35:0:0:3  sdbq 68:64  active ready running
mpathay (3600507640084031dd800000000000066) dm-8 IBM          ,2145
size=250G features='1 queue_if_no_path' hwhandler='0' wp=rw
| -+- policy='service-time 0' prio=50 status=active
|  ~- 35:0:0:2  sdbp 68:48  active ready running
^-+- policy='service-time 0' prio=10 status=enabled
  ~- 36:0:0:2  sdbt 68:112 active ready running
mpathax (3600507640084031dd800000000000061) dm-7 IBM          ,2145
size=100G features='1 queue_if_no_path' hwhandler='0' wp=rw
| -+- policy='service-time 0' prio=50 status=active
|  ~- 36:0:0:1  sdfs 68:96  active ready running
^-+- policy='service-time 0' prio=10 status=enabled
  ~- 35:0:0:1  sdbo 68:32  active ready running

```

Record the names of the devices that are marked bold in Example 7-57 on page 477 because they will be used in further configuration, such as LVM physical volume creation or file system creation and mounting, which are in `/dev/mapper/`.

Record the UID number after devices names without the first digit (3) because they correspond to the UID of the volume (VDisk) on the storage system.

Summary

Although the example in this section is specifically for RHEL host connectivity, the main principals can be followed when configuring connectivity through iSCSI for other OSs.

In summary, the actions that are necessary for host to storage iSCSI connectivity are:

1. Install the iSCSI initiator software on the host.
2. Configure the iSCSI initiator software according to the requirements for the storage system target and the host's OS.
3. Get the host IQNs.
4. Define the host object with iSCSI connectivity by using host IQNs.
5. Record and check the Ethernet ports IP addresses on the storage system, which are configured for iSCSI connectivity.
6. Discover the iSCSI targets by using the storage system IP address that was obtained in step 5.
7. Log in to the storage system iSCSI targets.
8. Check and configure the native multipath driver to confirm the volumes on the host.

7.7.4 NVMe over Fabric host connectivity example

NVMe-oF uses different fabrics for transport by using the NVMe protocol. In this example, we use an FC fabric for our NVMe connectivity from the RHEL host to an IBM FlashSystem system.

Start by defining the necessary connectivity information and configuring the host and system.

The concept of NVMe-oF is in much like iSCSI connectivity because an initiator and target must be defined and configured so that the connection works.

Collect the information for connectivity from the host to the System by completing the following steps:

1. You must discover the WWPNs of the host because FC-NVMe connectivity is achieved through FC. To do so, run the command that is show in Example 7-58.

Example 7-58 Obtaining host WWPNs

```
[root@flashlnx4 fc_host]# cat /sys/class/fc_host/host*/port_name
0x10000090faf20bc0
0x10000090faf20bc1
```

2. Discover the NVMe FC ports for the system by running the command that is shown in Example 7-59. Decide which ones that you will use. The FC-NVMe connectivity dedicated port is a virtualized port, so you must have NPIV enabled.

Example 7-59 Discovering the NVMe FC ports

```
IBM_IBM FlashSystem:FS9100-1:redbook>lstargetportfc|grep -i nvme
```

id	WWPN	WWNN	port_id	owning_node_id	current_node_id	nportid	host_io_permitted	virtualized	protocol
3	50050768101901E5	50050768100001E5	1	1	1	080E02	yes	yes	nvme
6	50050768101A01E5	50050768100001E5	2	1	1	020102	yes	yes	nvme
9	50050768101B01E5	50050768100001E5	3	1	1	020102	yes	yes	nvme
12	50050768101C01E5	50050768100001E5	4	1	1	000000	yes	yes	nvme
15	50050768102901E5	50050768100001E5	5	1	1	330242	yes	yes	nvme
18	50050768102A01E5	50050768100001E5	6	1	1	340242	yes	yes	nvme
21	50050768102B01E5	50050768100001E5	7	1	1	000000	yes	yes	nvme
24	50050768102C01E5	50050768100001E5	8	1	1	000000	yes	yes	nvme
51	50050768101901DF	50050768100001DF	1	2	2	080F02	yes	yes	nvme
54	50050768101A01DF	50050768100001DF	2	2	2	021002	yes	yes	nvme
57	50050768101B01DF	50050768100001DF	3	2	2	021002	yes	yes	nvme
60	50050768101C01DF	50050768100001DF	4	2	2	000000	yes	yes	nvme
63	50050768102901DF	50050768100001DF	5	2	2	330342	yes	yes	nvme
66	50050768102A01DF	50050768100001DF	6	2	2	340342	yes	yes	nvme

3. Zone the host with at least one NVMe dedicated port. In the example, the host is zoned to the ports that are marked in bold in Example 7-59.
4. On the host, make sure that the driver is ready to provide NVMe connectivity. In this example, we use an Emulex HBA, as shown in Example 7-60.

Example 7-60 Checking NVMe support for the lpfc driver

```
[root@flashlnx4 fc_host]# cat /etc/modprobe.d/lpfc.conf
options lpfc lpfc_enable_fc4_type=3
```

If the lpfc.conf is absent or does not contain the string that is marked in bold in the example, create it and populate it with the string. Then, restart the lpfc driver by running **modprob** commands (First, remove the driver, and then add it back).

Note: Reinitiating the lpfc driver by running the **modprob** command changes the NQN of the host.

5. Check that `nvme-cli` and `nvme-fc-connect` are installed on the host, as shown in Example 7-61.

Example 7-61 Checking the nvme-cli and nvme-connect availability

```
root@flashlnx4 nvme]# rpm -qa|grep nvme
nvme-cli-1.6-1.el7.x86_64
nvme-fc-connect-12.6.61.0-1.noarch
```

If the packages are not installed, install them.

6. Obtain the NQN (see Example 7-62) from the host because it is used to define the host objects on the system.

Example 7-62 Obtaining the NQN

```
[root@flashlnx4 nvme]# cat /etc/nvme/hostnqn
nqn.2014-08.org.nvmexpress:uuid:0c3f53f4-8161-49c6-aaeb-a98d8e5f572c
```

7. Create a host object on the system by using the host NQN, as described in “Creating NVMe host objects” on page 430. Check that the host object has the correct NQN set. (see Figure 7-81).

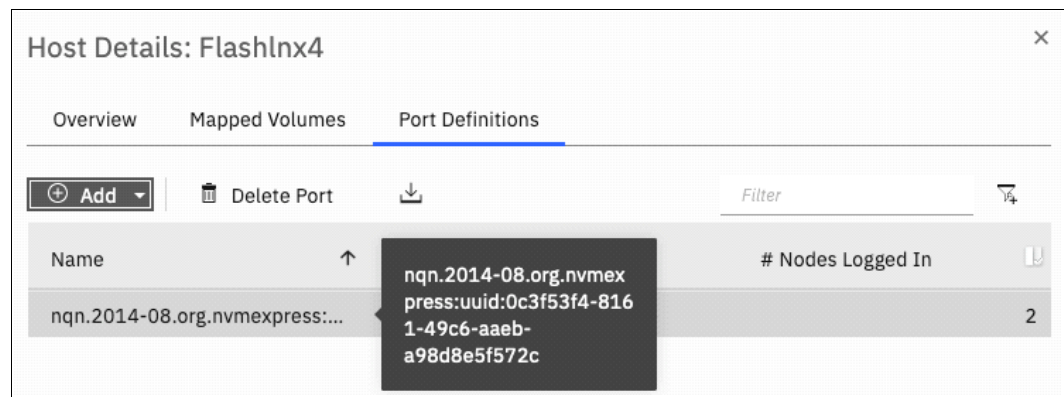


Figure 7-81 Checking the host object NQN on the system

8. On the system, map the volumes to the host object.
9. If the zoning is done correctly, the host object is created on the system, and on the host, the necessary utilities and drivers are configured. Verify the target ports that the host can see, as shown in Example 7-63. This information is used in the discovery and connection process.

Example 7-63 Verifying the remote/target ports and information about the FC-NVMe connection

```
[root@flashlnx4 nvme]# cat /sys/class/scsi_host/*/nvme_info

NVMe Initiator Enabled
XRI Dist lpfc0 Total 6144 I/O 5894 ELS 250
NVMe LPORT lpfc0 WWPN x10000090faf20bc0 WWNN x20000090faf20bc0 DID x330040 ONLINE
NVMe RPORT      WWPN x500507605e8c3443 WWNN x500507605e8c3440 DID x333e40 TARGET DISCSRV ONLINE
NVMe RPORT      WWPN x500507605e8c3463 WWNN x500507605e8c3440 DID x333f40 TARGET DISCSRV ONLINE
NVMe RPORT      WWPN x50050768102901e5 WWNN x50050768100001e5 DID x330242 TARGET DISCSRV ONLINE

NVMe Statistics
LS: Xmt 0000000031 Cmpl 0000000031 Abort 00000000
LS XMIT: Err 00000000 CMPL: xb 00000000 Err 00000000
Total FCP Cmpl 00000000035d907 Issue 00000000035d90a OutI/O 0000000000000003
```



```
abort 00000001 noxri 00000000 nondlp 00000000 qdepth 00000000 wqerr 00000000 err 00000000
FCP Cmpl: xb 00000001 Err 00000005
```

NVMe Initiator Enabled

```
XRI Dist lpfc1 Total 6144 I/O 5894 ELS 250
NVMe LPORT lpfc1 WWPn x10000090faf20bc1 WWNN x20000090faf20bc1 DID x340040 ONLINE
NVMe RPORT      WWPn x500507605e8c3453 WWNN x500507605e8c3440 DID x343e40 TARGET DISCSRV ONLINE
NVMe RPORT      WWPn x500507605e8c3473 WWNN x500507605e8c3440 DID x343f40 TARGET DISCSRV ONLINE
NVMe RPORT      WWPn x50050768102a01df WWNN x50050768100001df DID x340342 TARGET DISCSRV ONLINE
```

NVMe Statistics

```
LS: Xmt 0000000030 Cmpl 0000000030 Abort 00000000
LS XMIT: Err 00000000 Cmpl: xb 00000000 Err 00000000
Total FCP Cmpl 000000000035d6c3 Issue 000000000035d6c6 OutI/O 0000000000000003
abort 00000001 noxri 00000000 nondlp 00000000 qdepth 00000000 wqerr 00000000 err 00000000
FCP Cmpl: xb 00000001 Err 00000005
```

Tip: If the remote ports (RPORTs), which are presented from the system are not visible, check whether zoning is done correctly for the virtualized NVMe ports on the system.

10. Discover and connect to the storage resources, which requires using information from the `nvme_info` file, such as the WWNN and WWPn of the local port (host port) and RPORT (storage port). This information can be cumbersome to collect and put into the discovery and connect command manually, so you can use the script that is shown in Example 7-64 to automate the process. The commands for `nvme-c1i` are in bold.

Example 7-64 Script for FC-NVMe discovery and connection

```
[root@flashlnx4 tmp]# cat /tmp/disco_connect.bash
#!/bin/bash
#gather list of valid FC adapters by listing /sys/class/fc_host
for HOST in `ls -l /sys/class/fc_host`;do
  host_wwpn=`grep LPORT /sys/class/scsi_host/${HOST}/nvme_info |awk '{print $5}' |sed -e 's/x//'^
  host_wwnn=`grep LPORT /sys/class/scsi_host/${HOST}/nvme_info |awk '{print $7}' |sed -e 's/x//'^
  #iterate through the list of available targets on each FC adapter
  for LINE in `grep RPORT /sys/class/scsi_host/${HOST}/nvme_info|awk '{print $4":"$6}'|sed -e's/x//g'^;do
    target_wwpn=`echo ${LINE}|cut -d: -f1`
    target_wwnn=`echo ${LINE}|cut -d: -f2`
    echo "Performing Discovery and Connection with hostwwpn: ${host_wwpn} hostwwnn: ${host_wwnn}
targetwwpn: ${target_wwpn} targetwwnn: ${target_wwnn}"
    nvme discover --transport=fc --traddr=nn-0x${target_wwnn}:pn-0x${target_wwpn}
--host-traddr=nn-0x${host_wwnn}:pn-0x${host_wwpn}
    #grab the host nqn from /etc/nvme/hostnqn
    NQN=`cat /etc/nvme/hostnqn`
    nvme connect --transport=fc --traddr=nn-0x${target_wwnn}:pn-0x${target_wwpn}
--host-traddr=nn-0x${host_wwnn}:pn-0x${host_wwpn} -n ${NQN}
  done
done
```

Example 7-65 shows an example of the possible output.

Example 7-65 Discovery and connection script output

```
[root@flashlnx4 tmp]# ./tmp/disco_connect.bash
Performing Discovery and Connection with hostwwpn: 10000090faf20bc0 hostwwnn:
20000090faf20bc0 targetwwpn: 500507605e8c3443 targetwwnn: 500507605e8c3440

Discovery Log Number of Records 1, Generation counter 0
```

```
====Discovery Log Entry 0====  
trtype: fibre-channel  
adrfam: fibre-channel  
subtype: nvme subsystem  
treq: not required  
portid: 2  
trsvcid: none  
subnqn:  
nqn.2017-12.com.ibm:nvme:mt:9840:guid:5005076061D30D60:cid:0000020061D16202  
traddr: nn-0x500507605e8c3440:pn-0x500507605e8c3443  
Performing Discovery and Connection with hostwwpn: 1000090faf20bc0 hostwwnn:  
2000090faf20bc0 targetwwpn: 0x500507605e8c3463 targetwwnn: 500507605e8c3440
```

Discovery Log Number of Records 1, Generation counter 0

```
====Discovery Log Entry 0====  
trtype: fibre-channel  
adrfam: fibre-channel  
subtype: nvme subsystem  
treq: not required  
portid: 10  
trsvcid: none  
subnqn:  
nqn.2017-12.com.ibm:nvme:mt:9840:guid:5005076061D30D60:cid:0000020061D16202  
traddr: nn-0x500507605e8c3440:pn-0x500507605e8c3463  
Performing Discovery and Connection with hostwwpn: 1000090faf20bc0 hostwwnn:  
2000090faf20bc0 targetwwpn: 0x50050768102901e5 targetwwnn: 50050768100001e5
```

Discovery Log Number of Records 1, Generation counter 0

```
====Discovery Log Entry 0====  
trtype: fibre-channel  
adrfam: fibre-channel  
subtype: nvme subsystem  
treq: unrecognized  
portid: 4  
trsvcid: none  
subnqn: nqn.1986-03.com.ibm:nvme:2145.00000204228003CA  
traddr: nn-0x50050768100001e5:pn-0x50050768102901e5  
Performing Discovery and Connection with hostwwpn: 1000090faf20bc1 hostwwnn:  
2000090faf20bc1 targetwwpn: 500507605e8c3453 targetwwnn: 500507605e8c3440
```

Discovery Log Number of Records 1, Generation counter 0

```
====Discovery Log Entry 0====  
trtype: fibre-channel  
adrfam: fibre-channel  
subtype: nvme subsystem  
treq: not required  
portid: 6  
trsvcid: none  
subnqn:  
nqn.2017-12.com.ibm:nvme:mt:9840:guid:5005076061D30D60:cid:0000020061D16202  
traddr: nn-0x500507605e8c3440:pn-0x500507605e8c3453  
Performing Discovery and Connection with hostwwpn: 1000090faf20bc1 hostwwnn:  
2000090faf20bc1 targetwwpn: 500507605e8c3473 targetwwnn: 500507605e8c3440
```

Discovery Log Number of Records 1, Generation counter 0

```

=====Discovery Log Entry 0=====
trtype: fibre-channel
adrfam: fibre-channel
subtype: nvme subsystem
treq: not required
portid: 14
trsvcid: none
subnqn:
nqn.2017-12.com.ibm:nvme:mt:9840:guid:5005076061D30D60:cid:0000020061D16202
traddr: nn-0x500507605e8c3440:pn-0x500507605e8c3473
Performing Discovery and Connection with hostwwpn: 1000090faf20bc1 hostwwnn:
2000090faf20bc1 targetwwpn: 50050768102a01df targetwwnn: 50050768100001df

```

Discovery Log Number of Records 1, Generation counter 0
 =====Discovery Log Entry 0=====

```

trtype: fibre-channel
adrfam: fibre-channel
subtype: nvme subsystem
treq: unrecognized
portid: 4101
trsvcid: none
subnqn: nqn.1986-03.com.ibm:nvme:2145.00000204228003CA
traddr: nn-0x50050768100001df:pn-0x50050768102a01df

```

After the discovery and connection is successful, record the ports that marked in bold in Example 7-65 on page 481, and check the list of NVMe devices that are visible from the host, as shown in Example 7-66.

Example 7-66 NVMe devices list that is visible from the host

```

[root@flashlnx4 tmp]# nvme list
Node          S/N              Model
Namespace Usage          Format           FW Rev
-----
/dev/nvme6n1  204228003c      IBM           2145
78          132.07 GB / 137.44 GB  512 B + 0 B  8.4.0.0
/dev/nvme7n1  204228003c      IBM           2145
78          132.07 GB / 137.44 GB  512 B + 0 B  8.4.0.0

```

11. Run the **multipath** command to find newly connected volumes.
12. Run **multipath -ll** to see the paths and information about the volumes, as shown in Example 7-67.

Example 7-67 Output of the multipath -ll command

```

multipath -ll
...
eui.28000000000005300507608108a000f dm-4 NVMe,IBM 2145
size=128G features='1 queue_if_no_path' hwhandler='0' wp=rw
|+- policy='service-time 0' prio=1 status=active
|  ~- 6:0:1:0 nvme6n1 259:0 active ready running
~+- policy='service-time 0' prio=1 status=enabled
  ~- 7:0:1:0 nvme7n1 259:1 active ready running

```

- Record the name of the volume and UID in bold so that you can use for further actions, such as adding a partition or implementing a volume as part of an LVM. In Figure 7-82, the same volume is shown on the system.

Name	State	Synchronized	Pool	Protocol Type	UID
FlashInx4_0	✓ Online		Master	NVMe	28000000000005300507608108A000F

Figure 7-82 Volume as it is seen on the system

Summary

In this section, we provided an example of using an RHEL host and an IBM FlashSystem 9100 system. Although other OS distributions might have specific steps for configuration, the main idea and principles are the same. If it is necessary to connect to the storage through FC-NVMe, the following considerations and actions are usually performed:

- Ensure that the host is ready and meets the requirements for FC-NVMe connectivity, such as:
 - HBA supports FC-NVMe.
 - The drivers are configured for NVMe connectivity.
- Make sure that the system supports the host HBA for FC-NVMe connectivity.
- Obtain the connectivity information from the host.
- Create a host object on the system by using connectivity information from the host.
- Map volumes to the host object.
- Do discovery and connection from the host, although some hosts OS can do it automatically.
- Use the obtained storage resources.



Storage migration

This chapter describes the steps that are involved in migrating data from an external storage controller to an IBM FlashSystem system by using the storage migration wizard. Migrating data from other storage systems to the IBM FlashSystem system consolidates storage and enables IBM Spectrum Virtualize features, such as Easy Tier, thin provisioning, compression, encryption, storage replication, and the GUI, which can be used across all volumes.

Storage migration uses the volume mirroring function to enable reads and writes during the migration, which minimizes disruption and downtime. After the migration completes, the existing controller can be retired.

The system supports migration through Fibre Channel (FC) and internet Small Computer Systems Interface (iSCSI) connections.

In addition to migrating data through external virtualization and volume mirroring that is used by the storage migration wizard, there are also scenarios in which host-based mirroring is a best practice. In environments where operating system (OS) administrators can perform the migration by using host-side tools, host-based mirroring can potentially reduce or eliminate downtime if the new volumes that are presented from the IBM Spectrum Virtualize system and the legacy storage system are visible to the host concurrently.

Note: For a “real-life” demonstration of the storage migration capabilities that are offered with IBM Spectrum Virtualize, see [this web page](#) (login required).

The demonstration includes three different step-by-step scenarios showing the integration of an IBM SAN Volume Controller (SVC) cluster into an environment with one Microsoft Windows Server (image mode), one IBM AIX server (logical volume manager (LVM) mirroring), and one VMware Elastic Sky X Integrated (ESXi) server (storage vMotion).

Finally, one last scenario worth mentioning is enclosure upgrade migration, which is a fairly specialized case that is specifically for environments with IBM Storwize systems upgrading to an IBM FlashSystem system. These scenarios can use three capabilities in IBM Spectrum Virtualize to provide a seamless transition to the new hardware:

- ▶ Clustering of the new enclosure with the existing storage control enclosure (see Table 8-1 for information about the compatible systems).
- ▶ Modifying an I/O Group or performing a Non-disruptive Volume Move (NDVM) to change the caching I/O group for the volumes.
- ▶ Volume mirroring to move the data.

Table 8-1 New hardware clustering options for Storwize control enclosures

Storwize enclosure	Clustering options		
IBM Storwize V7000	IBM FlashSystem 7200	IBM FlashSystem 9100	IBM FlashSystem 9200
IBM Storwize 5100	IBM FlashSystem 5100		
IBM Storwize 5030	IBM FlashSystem 5030		

This chapter includes the following topics:

- ▶ 8.1, “Storage migration overview” on page 486
- ▶ 8.2, “Storage migration wizard” on page 489
- ▶ 8.3, “Enclosure Upgrade Migration” on page 507

Note: This chapter covers the storage migration wizard in detail, along with a less detailed description of the enclosure upgrade scenario. However, this chapter does not describe other migration methodologies such as ones that use replication or host-based migrations. This chapter also does not cover virtualization of external storage. For more information about these topics, see Chapter 5, “Storage pools” on page 237.

8.1 Storage migration overview

To migrate data from a storage controller to the system, you must use the built-in external virtualization capability. This capability places externally connected logical units (LUs) under the control of the IBM Spectrum Virtualize system, which acts as a proxy while hosts continue to access them. The volumes are then fully virtualized in the system.

Attention: The system does not require a license for its own control and expansion enclosures. However, a license is required for any external systems that are being virtualized, either based on storage capacity units (SCU) or based on the number of enclosures. Data can be migrated from storage systems to your system by using the external virtualization function within 90 days of purchase of the system without the purchase of a license. After 90 days, any ongoing use of the external virtualization function requires a license.

Set the license temporarily during the migration process to prevent messages that indicate that you are in violation of the license agreement from being sent. When the migration is complete, or after 45 days, reset the license to its original limit or purchase a new license.

Consider the following points about the storage migration process:

- ▶ Typically, storage controllers divide storage into many Small Computer System Interface (SCSI) LUs that are presented to hosts.
- ▶ I/O to the LUs must be stopped and changes made to the mapping of the external storage controller LUs and to the fabric or iSCSI configuration so that the original LUs are presented directly to the system and not to the hosts anymore. The system discovers the external LUs as *unmanaged* managed disks (MDisks).
- ▶ The unmanaged MDisks are *imported* to the system as *image mode volumes* and placed into a temporary storage pool. This storage pool is now a logical container for the LUs.
- ▶ Each MDisk has a one-to-one mapping with an image mode volume. From a data perspective, the image mode volumes represent the LUs exactly as they were before the import operation. The image mode volumes are on the same physical drives of the external storage controller and the data remains unchanged. The system is presenting active images of the LUs and acting as a proxy.
- ▶ You might need to remove the storage system multipath device driver from the host and reconfigure host attachment with this system. However, most current OSs might not require vendor-specific multipathing drivers and can access both the legacy and the new IBM Spectrum Virtualize systems through native multipathing drivers, such as AIX AIXPCM, Linux device mapper, or Microsoft Device Specific Module (MSDSM). The hosts are defined with worldwide port names (WWPNs) or iSCSI Qualified Names (IQNs), and the volumes are mapped to the hosts. After the volumes are mapped, the hosts discover the system's volumes through a host rescan or restart operation.
- ▶ After IBM Spectrum Virtualize volume mirroring operations are initiated, the image-mode volumes are mirrored to standard striped volumes. Volume mirroring is an online migration task, which means a host can still access and use the volumes during the mirror synchronization process.
- ▶ After the mirror operations are complete, the image mode volumes are removed. The external storage system LUs are now migrated and the now redundant storage can be decommissioned or reused elsewhere.

Important: If you are migrating volumes from another Storwize or IBM FlashSystem family product through external virtualization instead of clustering or replication, the target system *must* be configured in the *replication* layer, and the source system must be configured in the *storage* layer. Otherwise, the source system does not discover the target as a host, and the target does not discover the source as a back-end controller.

The default layer setting for Storwize and IBM FlashSystem family systems is storage:

```
chsystem -layer replication
chsystem -layer storage
```

Similarly, the layer setting might need to be changed if you cluster a Storwize system with an IBM FlashSystem enclosure.

8.1.1 Interoperability and compatibility

Interoperability is an important consideration when a new storage system is set up in an environment that contains a storage infrastructure. Before attaching any external storage systems to the system, see the [IBM System Storage Interoperation Center \(SSIC\)](#).

At the SSIC site, select **IBM System Storage Enterprise Flash** for IBM FlashSystem 9200 or **IBM System Storage Midrange Disk** for other hardware platforms like the Storwize family or IBM FlashSystem 7200, and then select the appropriate **Storage Controller Support** entry for your system as the Storage Model. You can refine your search by selecting the external storage controller that you want to use from the **Storage Controller** menu.

The matrix results indicate the external storage that you want to attach to the system, such as validated firmware levels or support for disks greater than 2 TB.

8.1.2 Prerequisites

Before the storage migration wizard can be started, the external storage controller must be visible to the system. You also must confirm that the restrictions, limits, and prerequisites are met.

Data from the external storage system to the IBM Spectrum Virtualize system is sent through an iSCSI or Fibre Channel connection (IBM FICON).

Common prerequisites

It is unlikely that VMware environments will use the Storage Migration wizard to move data because it requires downtime for path cutover. It is much more likely that Storage V-motion will be used to move guest data transparently to newly provisioned data stores from the IBM Spectrum Virtualize system. However, if you have VMware Elastic Sky X (ESX) server hosts and want to migrate by using *image mode*, you must change the settings on the VMware host so that copies of the volumes can be recognized by the system after the migration completes. To ensure that volume copies can be recognized by the system for VMware ESX hosts, you must complete one of the following actions:

- ▶ Enable the `EnableResignature` setting.
- ▶ Disable the `DisallowSnapshotLUN` setting.

To learn more about these settings, see the documentation for the VMware ESX host.

Note: Test the setting changes on a non-production server. The logical unit number (LUN) has a different unique identifier (UID) after it is imported. It resembles a mirrored volume to the VMware server.

Prerequisites for a Fibre Channel connection

The following prerequisites for an FC connection must be met:

- ▶ Make sure that an FC host interface card/host bus adapter (HIC/HBA) is installed in the node canisters.
- ▶ Cable this system into the storage area network (SAN) of the external storage that you want to migrate. Ensure that your system is cabled into the same SAN as the external storage controller that you are migrating.
- ▶ If you are using FC, connect the FC cables to the FC ports in *both* canisters of your system, and then to the FC network.

For more information and considerations, see Chapter 2, “Planning” on page 71.

Alternatively, directly attach the external storage controller to the nodes instead of using a switched fabric.

Prerequisites for iSCSI connections

The following prerequisites for iSCSI connections must be met:

- ▶ Cable this system to the external storage system with a redundant switched fabric. Migrating iSCSI external storage requires that the system and the storage system are connected through an Ethernet switch. Symmetric ports on *all* nodes of the system must be connected to the same switch and must be configured on the same subnet.
- ▶ In addition, modify the Ethernet port attributes to enable the external storage on the Ethernet port to enable external storage connectivity. To modify the Ethernet port for external storage, click **Network** → **Ethernet Ports** and right-click a configured port. Select **Modify Storage Ports** to enable the port for external storage connections.
- ▶ Cable the Ethernet ports on the storage system to the fabric in the same way as the system and ensure that they are configured in the same subnet. Optionally, you can use a virtual local area network (VLAN) to define network traffic for the system ports.
- ▶ For full redundancy, configure two Ethernet fabrics with separate Ethernet switches. If the source system nodes and the external storage system both have more than two Ethernet ports, an extra redundant iSCSI connection can be established for increased throughput.

8.2 Storage migration wizard

The storage migration wizard simplifies the migration task. The wizard features easy-to-follow windows that guide users through the entire process. The wizard shows you which commands are being run so that you can see exactly what is being performed throughout the process.

Attention: The risk of losing data when using the storage migration wizard correctly is low. However, it is prudent to avoid potential data loss by creating a backup of all the data that is stored on the hosts, the storage controllers, and the system before the wizard is used.

Complete the following steps to complete the migration by using the storage migration wizard:

1. Select **Pools** → **System Migration**, as shown in Figure 8-1. The System Migration window provides access to the storage migration wizard and displays information about the migration progress.

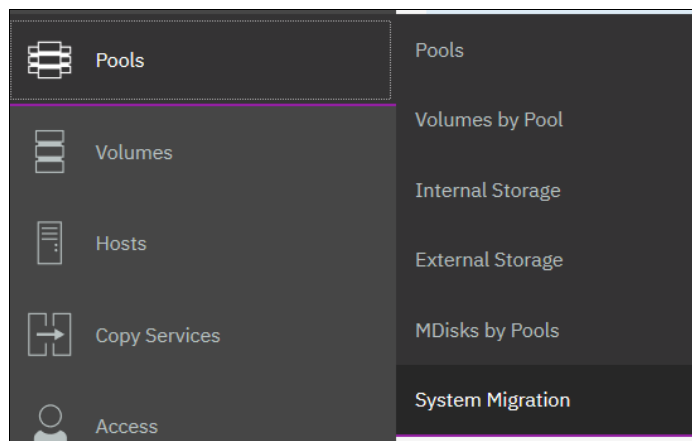


Figure 8-1 Browsing to Storage Migration

2. Click **Start New Migration** to begin the storage migration wizard, as shown in Figure 8-2.

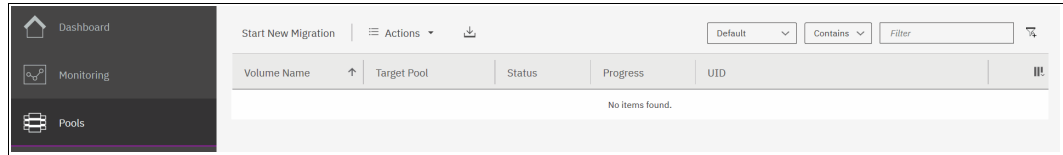


Figure 8-2 Starting a migration

Note: Starting a new migration adds the volume to be migrated to the list that is shown in Figure 8-2. After a volume is migrated, it remains in the list until you finalize the migration.

3. If both FC and iSCSI external systems are detected, a dialog box opens and prompts you about which protocol should be used. Select the type of attachment between the system and the external controller from which you want to migrate volumes and click **Next**. If only one type of attachment is detected, this dialog box does not open.

If the external storage system is not detected, the warning message that is shown in Figure 8-3 is displayed when you attempt to start the migration wizard. Click **Close** and correct the problem before you try to start the migration wizard again.

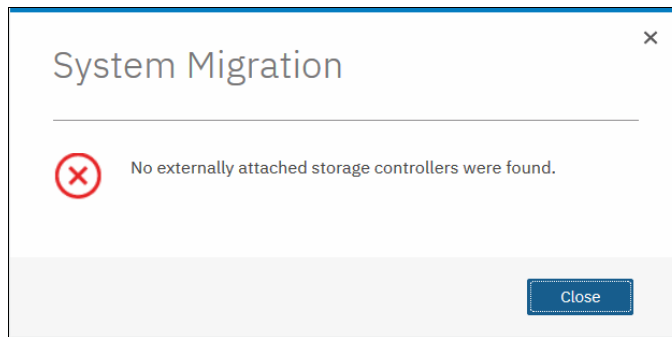


Figure 8-3 Error message if no external storage is detected

4. When the wizard starts, you are prompted to verify the restrictions and prerequisites that are listed in Figure 8-4 on page 491. Address the following restrictions and prerequisites:

– Restrictions:

- You are not using the storage migration wizard to migrate clustered hosts, including clusters of VMware hosts and Virtual I/O Servers (VIOSs).
- You are not using the storage migration wizard to migrate SAN boot images.

If you have either of these two environments, the migration must be performed outside of the wizard because more steps are required.

The VMware vSphere Storage vMotion feature might be an alternative for migrating VMware clusters. For information, see this [web page](#).

– Prerequisites:

- The system and the external storage controller are connected to the same SAN fabric.
- If there are VMware ESX hosts involved in the data migration, the VMware ESX hosts are set to allow volume copies to be recognized.

For more information about the Storage Migration prerequisites, see 8.1.2, “Prerequisites” on page 488.

If all restrictions are satisfied and prerequisites are met, select all of the options and click **Next**, as shown in Figure 8-4.

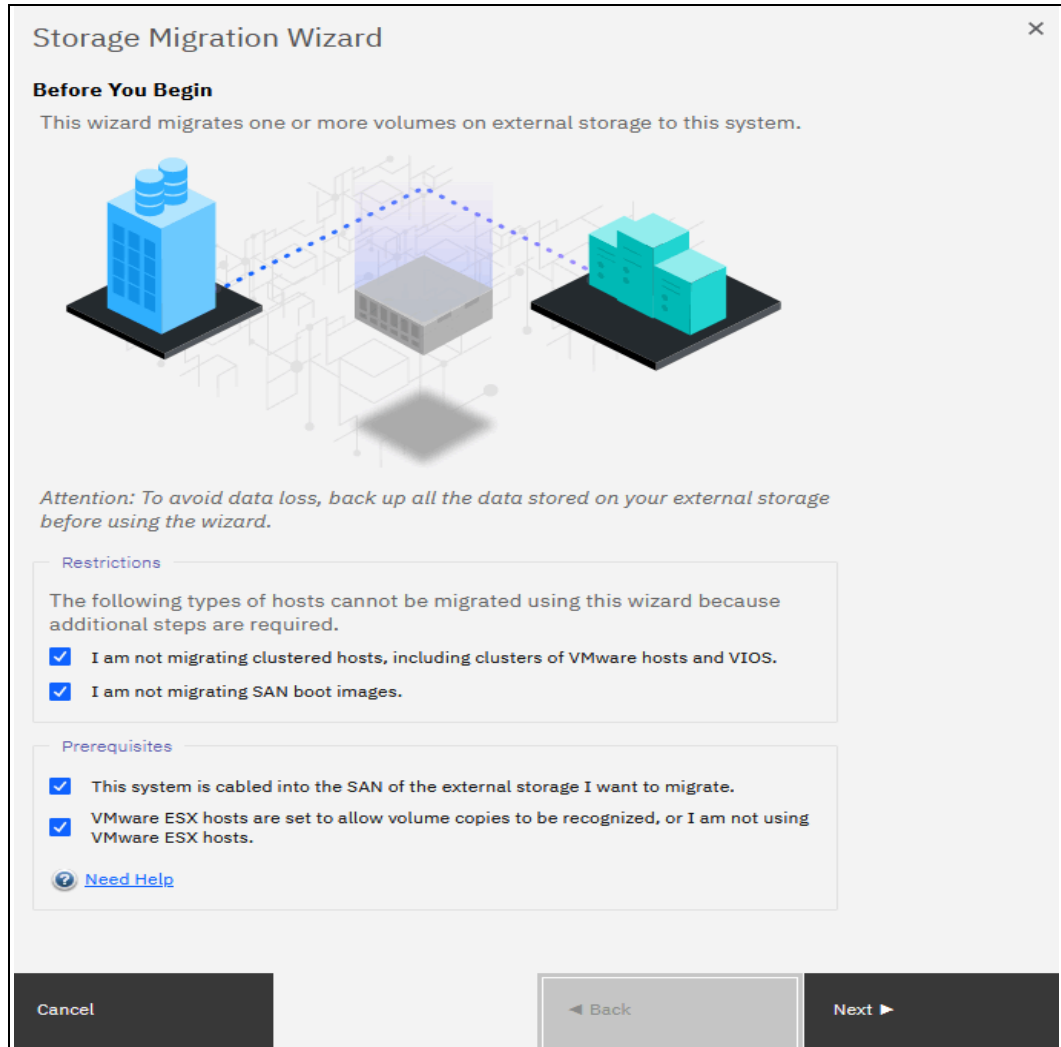


Figure 8-4 Restrictions and prerequisites confirmation

5. Prepare the environment migration by following the instructions that are shown in Figure 8-5.

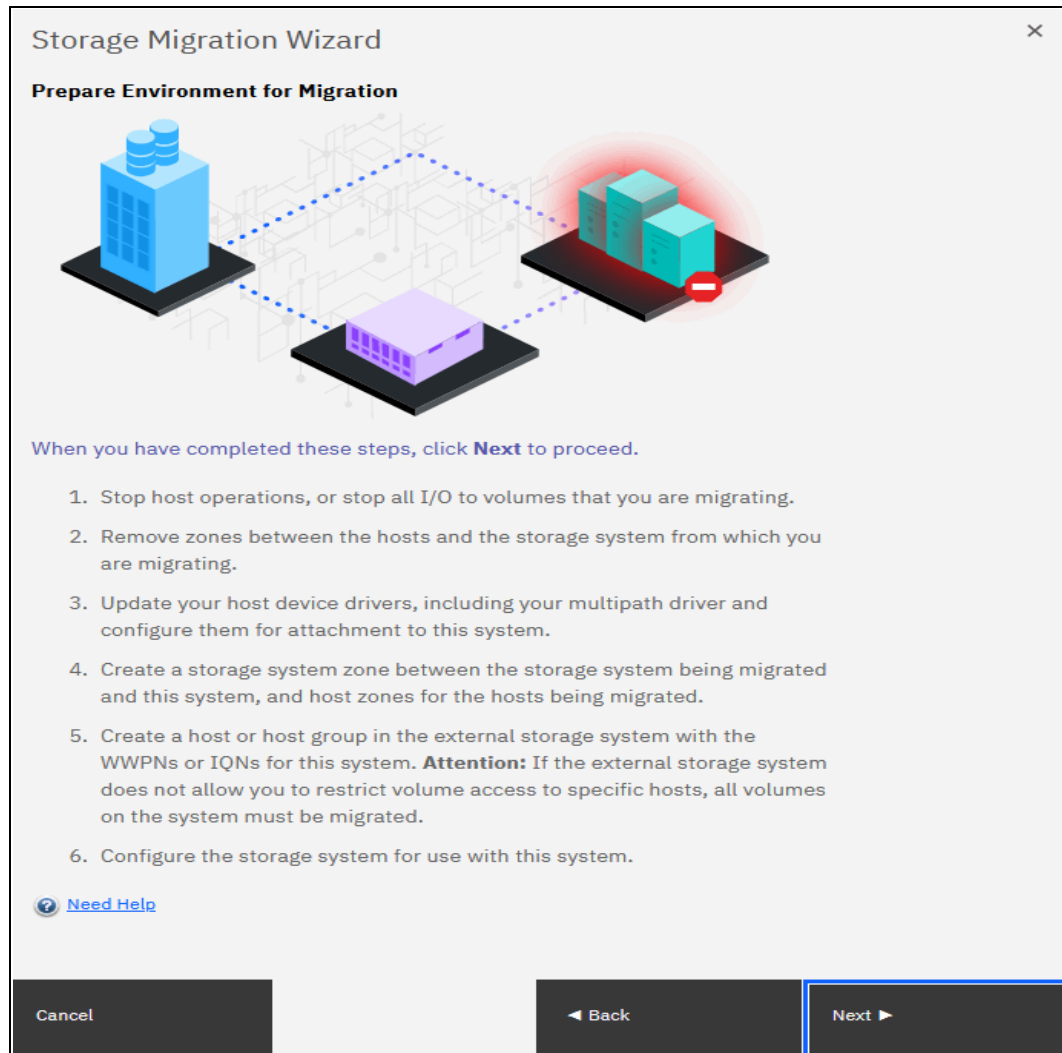


Figure 8-5 Preparing your environment for storage migration

The preparation phase includes the following steps:

- a. Before migrating the storage, ensure that all host operations are stopped to prevent applications from generating I/Os to the migrated system.
- b. Remove all zones between the hosts and the controller that you are migrating.
- c. Hosts usually do not support concurrent multipath drivers. You might need to remove drivers that are not compatible with the system from the hosts and use the recommended device drivers. For more information about supported drivers, see the [SSIC](#).
- d. If you are migrating external storage controllers that connect to the system that uses FC, ensure that you complete the appropriate zoning changes to simplify migration. In fact, an excellent preparatory step is to present a test LU from the external storage to the IBM Spectrum Virtualize system before the migration.

Use the following guidelines to ensure that zones are configured correctly for migration:

- Zoning rules

For every storage controller, create one zone that contains this system's ports from every node and all external storage controller ports, unless otherwise stated by the zoning guidelines for that storage controller.

This system requires single-initiator zoning for all large configurations that contain more than 64 host objects. Each server FC port must be in its own zone, which contains the FC port and this system's ports. In configurations of fewer than 64 hosts, you can have up to 40 FC ports in a host zone if the zone contains similar HBAs and OSs.

- Storage system zones

In a storage system zone, this system's nodes identify the storage systems. Generally, create one zone for each storage system. Host systems cannot operate on the storage systems directly. All data transfer occurs through this system's nodes.

- Host zones

In the host zone, the host systems can identify and address this system's nodes. You can have more than one host zone and more than one storage system zone. Create one host zone for each host FC port.

Because the system should now be seen as a host from the external controller to be migrated, you must define the system as a host or host group by using the WWPNs or IQNs on the system to be migrated. Some controllers do not support LUN-to-host mapping, so they present all the LUs to the system. In that case, all the LUs should be migrated.

6. If the previous preparation steps were followed, the system is now seen as a host from the controller to be migrated. LUs can then be mapped to the system. Map the external storage controller by following the instructions that are shown in Figure 8-6.

Storage Migration Wizard

Map Storage



The diagram illustrates the 'Map Storage' step of the Storage Migration Wizard. It shows three server racks on a grid background. On the left, a blue rack represents the source system. In the center, a purple rack represents the target system. On the right, a teal rack represents the external storage system, with a red prohibition sign over it. Dotted blue lines connect the blue rack to the purple rack, and the purple rack to the teal rack, indicating the migration path. A large, faint 'X' is overlaid on the background.

When you have completed these steps, click **Next** to map storage to your system.

1. Create a list of all external storage system volumes that are being migrated.
2. Record the hosts that use each volume.
3. Record the WWPNs or the IQNs associated with each host.
4. Unmap all volumes being migrated from the hosts in the storage system, and map them to the host or host group you created when preparing your environment. **Attention:** If the external storage system does not allow you to restrict volume access to specific hosts, all volumes on the system must be migrated.
5. Record the storage system LUN used to map each volume to this system.

[Need Help](#)

Cancel ◀ Back Next ▶

Figure 8-6 Steps to map the LUs to be migrated

Before you migrate storage, record the hosts and their WWPNs or IQNs for each volume that is being migrated and the SCSI LUN when it is mapped to the system.

Table 8-2 on page 495 shows an example of a table that is used to capture information that relates to the external storage system LUs.

Table 8-2 Example table for capturing external LU information

Volume Name or ID	Hosts accessing this LUN	Host WWPNs or IQNs	SCSI LUN when mapped
1 DB2logs	DB2server	21000024FF2...	0
2 Db2data	DB2Server	21000024FF2...	1
3 file system	FileServer1	21000024FF2...	2

Note: Make sure to record the SCSI ID of the LUs to which the host is originally mapped. Some OSs do not support changing the SCSI ID during the migration.

Click **Next** and wait for the system to discover external devices. The wizard runs a **detectmdisk** command, as shown in Figure 8-7.

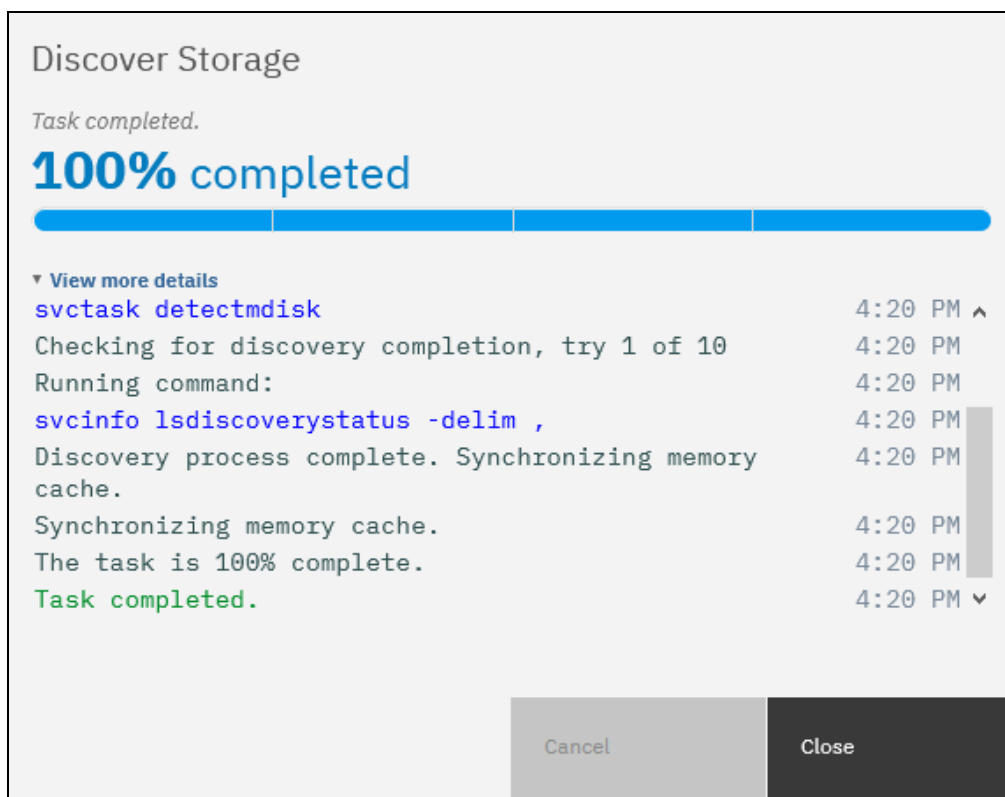


Figure 8-7 Storage Migration external storage discovery detectmdisk command detail

- The next window shows all the MDisks that were found. If the MDisks to be migrated are not in the list, check your zoning or IP configuration, as applicable, and your LUN mappings. Repeat step 6 on page 494 to trigger the discovery procedure again.

Select the MDisks that you want to migrate, as shown in Figure 8-8.

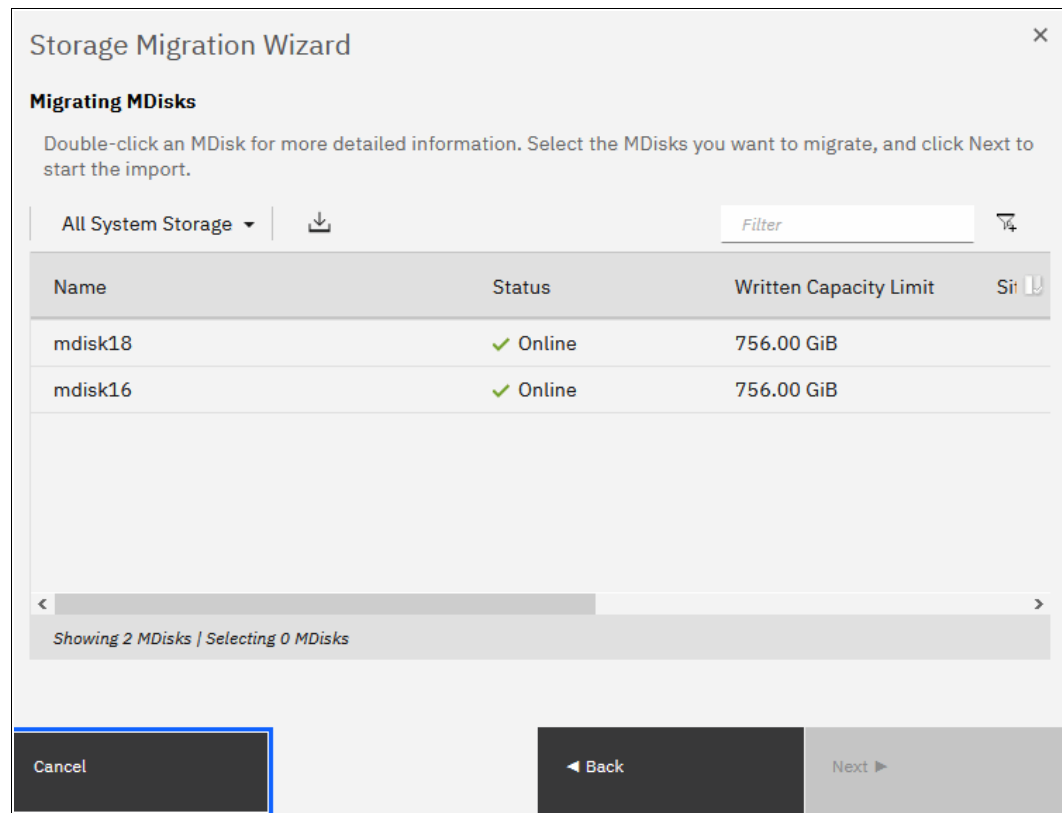



Figure 8-8 Discovering mapped LUs from external storage

In this example, two MDisks (mdisk18 and mdisk16) were found for migration. Detailed information about an MDisk is available by double-clicking it. To select multiple elements from the table, press Shift and then click or Ctrl and then click. Optionally, you can export the discovered MDisks list to a comma-separated value (CSV) file for further use by clicking the download icon () to **Export to CSV**.

Note: Select only the MDisks that are applicable to the current migration plan. After step 15 on page 505 of the current migration completes, another migration can be started to migrate any remaining MDisks.

8. Click **Next** and wait for the MDisk to be imported. During this task, the system creates a new storage pool that is called MigrationPool_XXXX and adds the imported MDisk to the storage pool as image mode volumes with the default naming of {controller}_16digitSequenceNumber(controller2_0000000000000005)..., as shown in Figure 8-9.

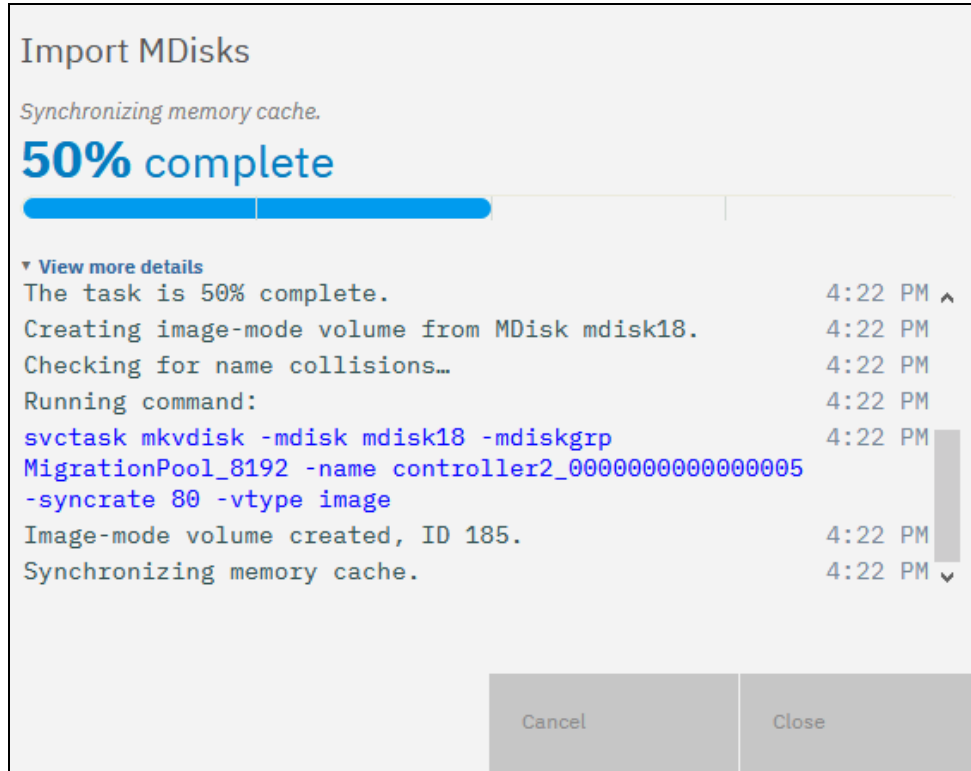


Figure 8-9 Storage Migration image mode volume creation detail

9. The next window lists the host that is configured on the system to which you can assign the volumes or configure new hosts. This step is optional and can be bypassed by clicking **Next**. In this example, the host X366-SLES-12SP2 is already configured, as shown in Figure 8-10. If no host is selected, you can create a host after the migration completes and map the imported volumes to it.

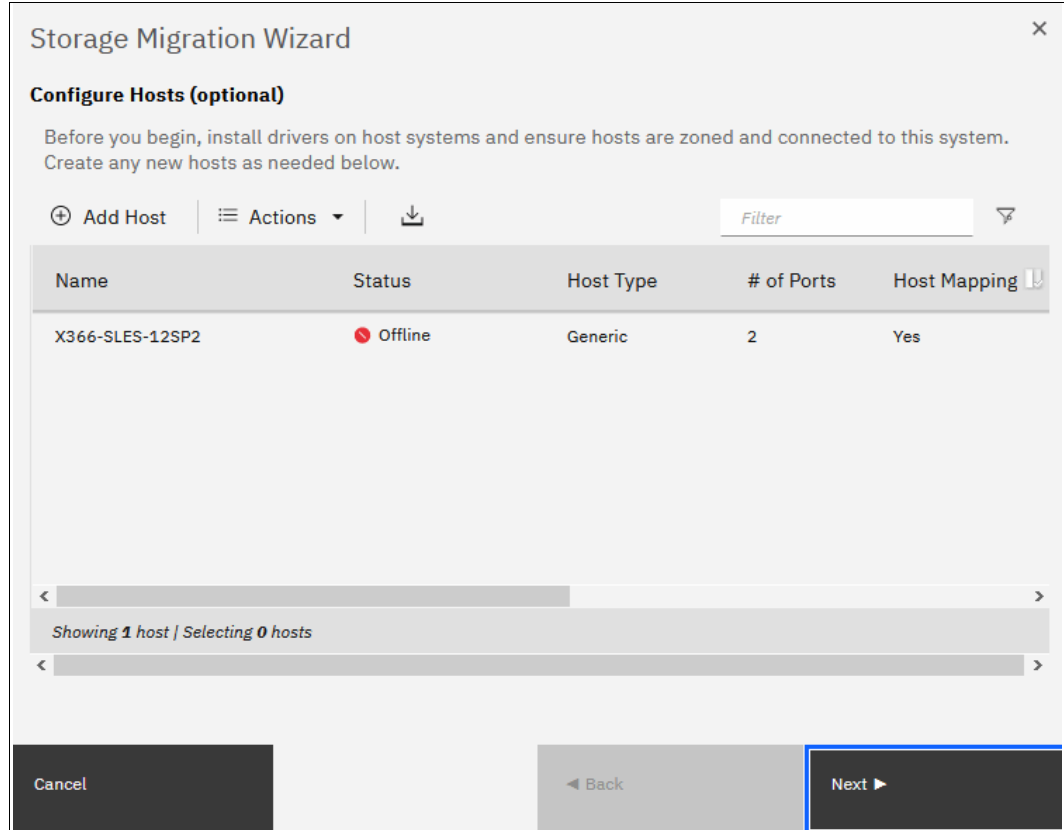


Figure 8-10 List of configured hosts to which to map the imported volume

10. If the host that needs access to the migrated data is not configured, select **Add Host** to begin the Add Host wizard. Enter the host connection type, name, and connection details. Optionally, click **Advanced** to modify the host type and I/O group assignment. Figure 8-11 shows the Add Host wizard with the details completed.

For more information about the Add Host wizard, see Chapter 7, “Hosts” on page 405.

Add Host

Required Fields

Name: ISCSI HOST

Host connections: iSCSI (SCSI)

Host IQN: iqn.1994-05.com.redhat:72dcb719

Optional Fields

CHAP authentication:

CHAP secret: Enter 1 to 79 characters

CHAP username: Enter 1 to 31 characters

Host type: Generic

I/O groups: All

Host cluster: No Host Cluster Selected

Cancel Add

Figure 8-11 Creating a host during the migration process

11. Click **Add**. The host is created and now listed in the Configure Hosts window, as shown in Figure 8-10 on page 498. Click **Next** to proceed.

12. The next window lists the new volumes, where you can map them to hosts, as shown in Figure 8-12. The volumes are listed with names that were automatically assigned by the system. The names can be changed to reflect something more meaningful to the user by selecting the volume and clicking **Rename** in the **Actions** menu.

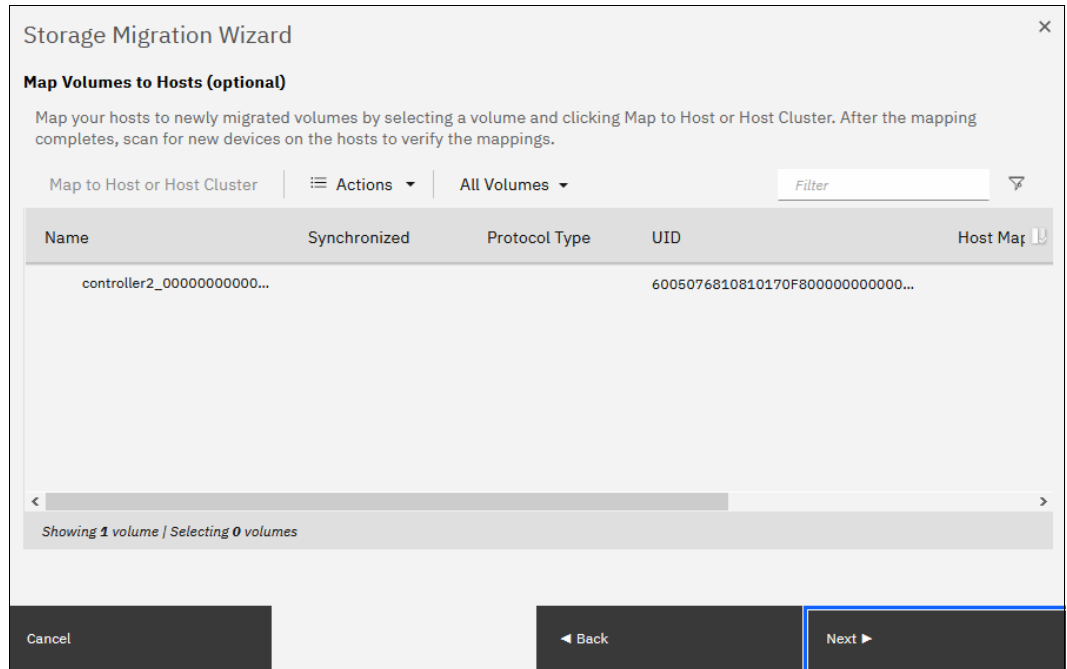


Figure 8-12 Mapping volumes to hosts

13. Map the volumes to the hosts by selecting the volumes and clicking **Map to Host or Host Cluster**, as shown in Figure 8-13. This step is optional and can be bypassed by clicking **Next**.

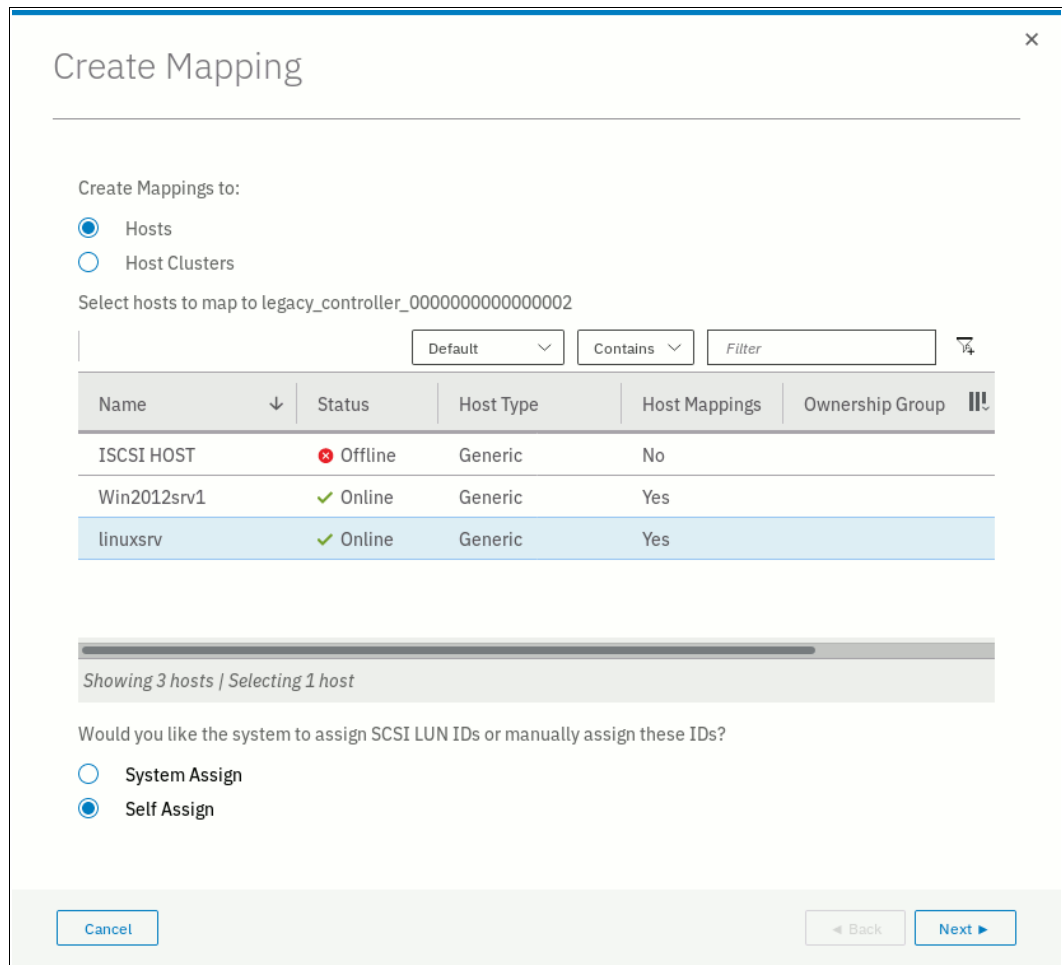


Figure 8-13 Selecting the host to which to map the new volume

You can manually assign a SCSI ID to the LUNs you are mapping. This technique is useful when the host needs to have the same LUN ID for a LUN before and after it is migrated. To assign the SCSI ID manually, select the **Self Assign** option and follow the instructions as shown in Figure 8-14.

The screenshot shows a dialog box titled "Map Volumes to linuxsrv: Select SCSI IDs". It is divided into two main sections. The left section, titled "Select SCSI ID these mappings will be placed on:", contains a table with three columns: "Name", "SCSI ID", and "Caching I/O Group ID". The "Name" column has a text input field containing "legacy_controll...". The "SCSI ID" column has a numeric input field with "12" and up/down arrow buttons. The "Caching I/O Group ID" column has a numeric input field with "0". The right section, titled "These SCSI IDs are already occupied:", contains a table with two columns: "Type of Mapping" and "SCSI ID". The "Type of Mapping" column lists "Private" for each row, and the "SCSI ID" column lists the numbers 0 through 7. At the bottom of the dialog, there are three buttons: "Cancel", "Back", and "Next".

Name	SCSI ID	Caching I/O Group ID
legacy_controll...	12	0

Type of Mapping	SCSI ID
Private	0
Private	1
Private	2
Private	3
Private	4
Private	5
Private	6
Private	7
Private	8

Figure 8-14 Manually assign a LUN SCSI ID to a mapped volume

When your LUN mapping is ready, click **Next**. A new dialog box opens with a summary of the new and existing mappings, as shown in Figure 8-15.

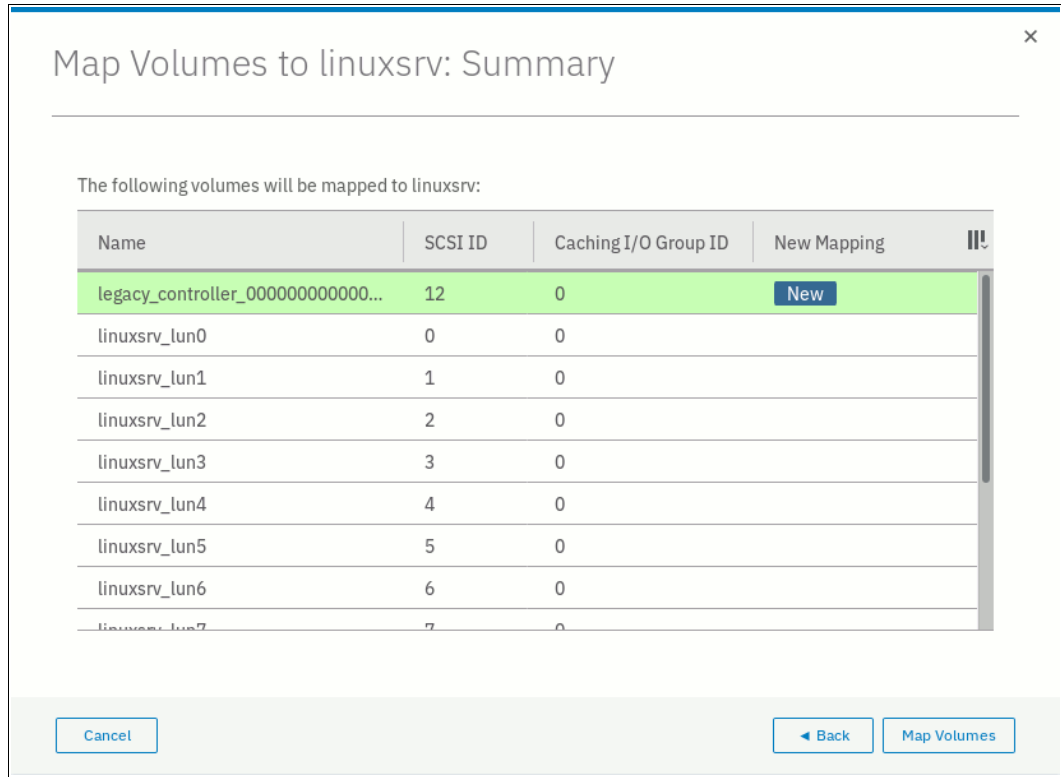


Figure 8-15 Volumes mapping summary before migration

Click **Map Volumes** and wait for the mappings to be created. Continue to map volumes to hosts until all mappings are created. Click **Next** to continue with the next migration step.

14. Select the storage pool into which you want to migrate the imported volumes. Ensure that the selected storage pool has enough space to accommodate the migrated volumes before you continue. This step is optional. You can decide not to migrate to a storage pool and to leave the imported MDisk as an image mode volume.

However, this technique is not recommended because no volume mirroring is created. Therefore, no protection is available for the imported MDisk, and no data transfer occurs from the controller to be migrated to the system. So, although it is acceptable to delay the mirroring at some point, it should be done.

Click **Next**, as shown in Figure 8-16.

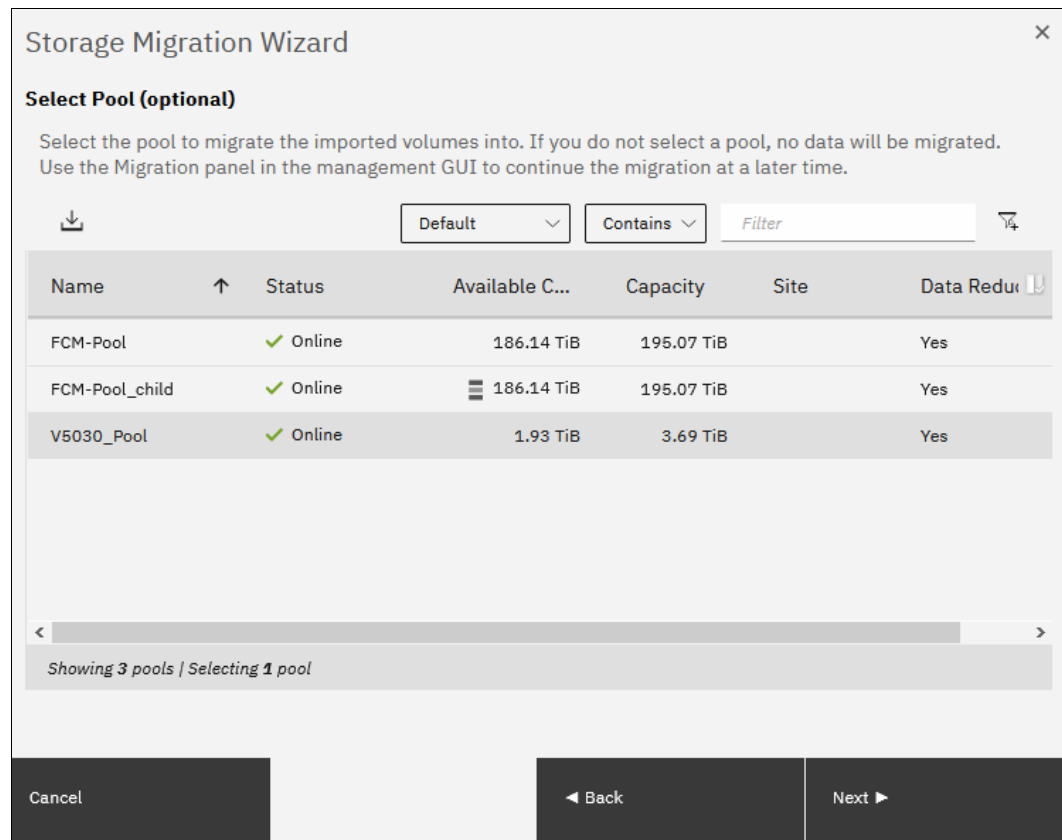


Figure 8-16 Selecting the target pool for the migration of the image mode MDisk

The migration starts. This task continues running in the background and uses the volume mirroring function to place a generic copy of the image mode volumes in the selected storage pool.

Note: With volume mirroring, the system creates two copies (Copy0 and Copy1) of a volume. Typically, Copy0 is located in the migration pool, and Copy1 is created in the target pool of the migration. When the host generates a write I/O on the volume, data is written concurrently on both copies. Read I/Os are performed on the primary copy only.

In the background, a mirror synchronization of the two copies is performed and runs until the two copies are synchronized. The speed of this background synchronization can be changed in the volume properties.

15. Click **Finish** to end the storage migration wizard, as shown in Figure 8-17.

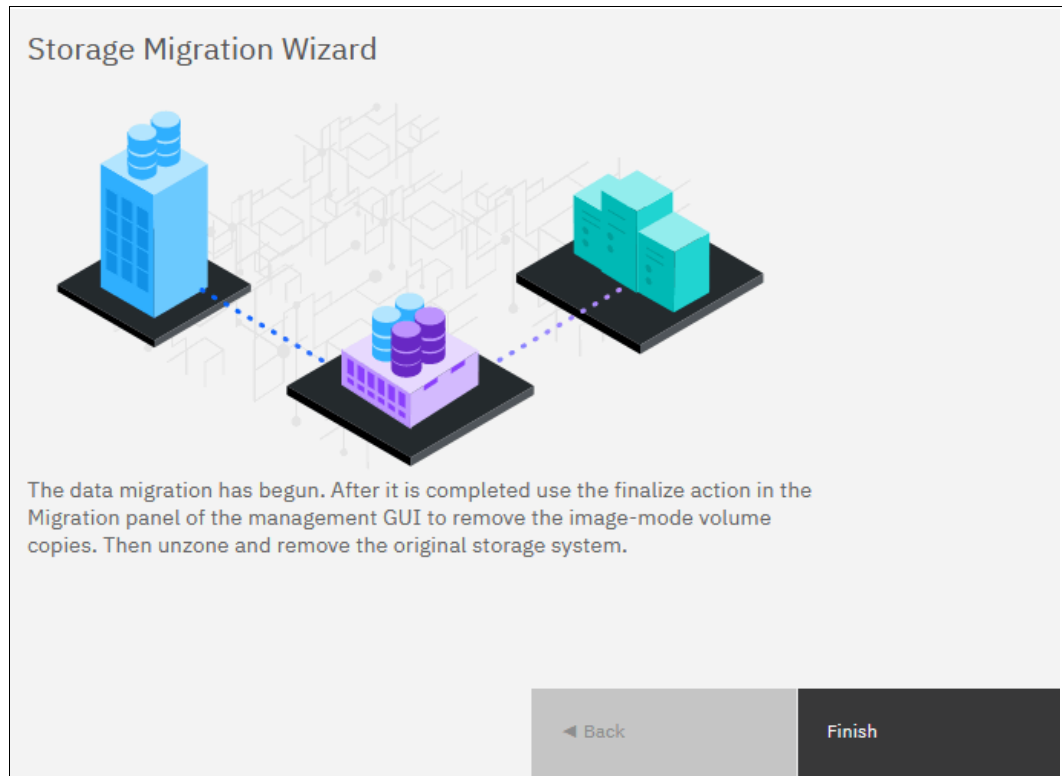


Figure 8-17 Migration is started

The end of the wizard is not the end of the migration task. You can find the progress of the migration in the Storage Migration window, as shown in Figure 8-18. The target storage pool and the progress of the volume copy synchronization is also displayed there.

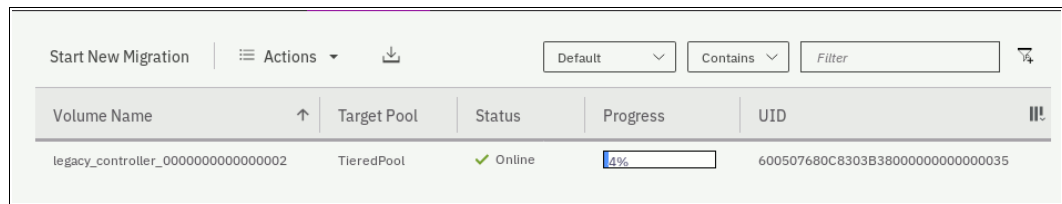


Figure 8-18 The ongoing migration is listed in the Storage Migration window

16. If you want to check the progress by using the command-line interface (CLI), run the **lsvdisksyncprogress** command because the process is essentially a volume copy, as shown in Example 8-1.

Example 8-1 Migration progress on the command-line interface

```
IBM_Storwize:ITS0-V7k:superuser>lsvdisksyncprogress
vdisk_id vdisk_name                progress estimated_completion_time
20      legacy_controller_0000000000000002 1      191021123932
```

17. When the migration completes, select all of the migrations that you want to finalize, right-click the selection, and click **Finalize**, as shown in Figure 8-19.

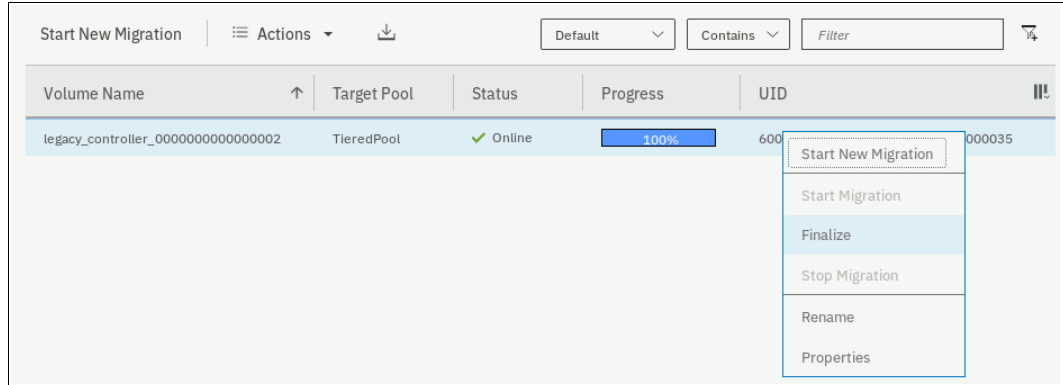


Figure 8-19 Finalizing a migration

You are asked to confirm the Finalize action because this process removes the MDisk from the Migration Pool and deletes the primary copy of the mirrored volume. The secondary copy remains in the destination pool and becomes the primary. Figure 8-20 shows the confirmation message.

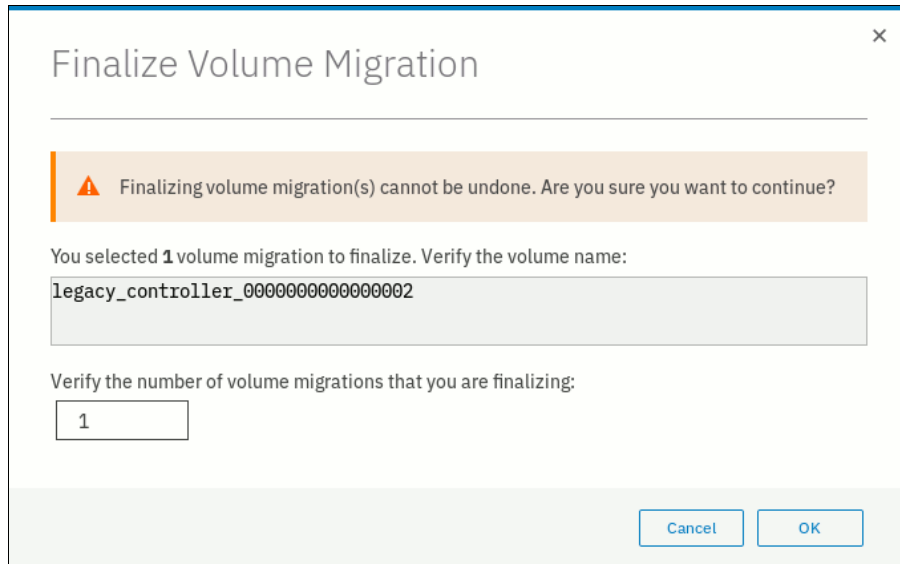


Figure 8-20 Migration finalization confirmation

18. When finalized, the image mode copies of the volumes are deleted and the associated MDisks are removed from the migration pool. The status of those MDisks returns to unmanaged. You can verify the status of the MDisks by selecting **Pools** → **External Storage**, as shown in Figure 8-21 on page 507. In the example, `mdisk3` was migrated and finalized. It appears as `unmanaged` in the external storage window.

Name	State	Written Capacity Limit	Mode	Site	Pool	Storage System	LUN
> flashsystem	Online	IBM 2145	Serial Number: 2076		Site: Unassigned	WWNN: 500507680B009479	
flashsystem	Online	IBM 2145	Serial Number: 2076		Site: Unassigned	WWNN: 5005076810000F62	
∨ legacy_cont	Online	IBM 2145	Serial Number: 2076		Site: Unassigned	WWNN: 5005076810000F88	
mdisk7	Online	1.00 TiB	Managed		TieredPool	legacy_controller	0000000000000001
mdisk5	Online	1.00 TiB	Managed		TieredPool	legacy_controller	0000000000000000
mdisk8	Online	100.00 GiB	Unmanaged			legacy_controller	0000000000000002
flashsystem	Online	IBM 2145	Serial Number: 2076		Site: Unassigned	WWNN: 500507680B009478	

Figure 8-21 External Storage MDisks window

All the steps that are described in the Storage Migration wizard can be performed manually with the GUI and the CLI, but you should use the wizard as a guide.

8.3 Enclosure Upgrade Migration

IBM FlashSystem enclosures can be clustered like Storwize enclosures, and extra options for migrating the data are available. This action assumes that the Storwize enclosure is a generation that can support the code that is required for the new hardware. For example, a Storwize V7000 system must be a Gen2, Gen2+, or Gen3 to support the Version 8.3.1 code that is needed to cluster with an IBM FlashSystem 7200 or IBM FlashSystem 9200.

With the clustering capability, you may concurrently migrate the access to volumes from the Storwize enclosure to the IBM FlashSystem enclosure and migrate the data from the Storwize internal storage pool to the IBM FlashSystem internal storage pool.

The I/O group access change can be performed at any time, but ideally should be done during a period of low production activity, and it must be coordinated with the OS administrator to ensure that path discovery occurs, as shown in the “Modify I/O Group...” wizard.

Note: There is a limitation in the NDVM process that prevents you from changing I/O groups if a volume is in a FlashCopy map or replication relationship. In those instances, the maps and relationships must be deleted and re-created. If an outage can be tolerated, use the `-sync` flag for relationship re-creation to avoid a resync. Otherwise, if no downtime is tolerable and a resync is acceptable, then the process can be concurrent and transparent to the host.

For more information about volume mirroring, see 6.5, “Operations on volumes” on page 321. Volume mirroring can be performed with either the CLI or GUI and be moderated to lessen or eliminate the impact on performance by using the sync rate volume property.

There is not any particular order that is required for the enclosure upgrade. The access change can be done before mirroring and vice versa. However, you should do the second process without too much delay, and you should consider doing the mirroring first to minimize the added impact of accessing volumes through the IBM FlashSystem enclosure while the data still is on the Storwize system, which might lead to performance impact.



Advanced features for storage efficiency

IBM Spectrum Virtualize running inside a storage system offers several functions for storage optimization and efficiency. This chapter introduces the basic concepts of those functions, and also provides a short technical overview and implementation recommendations.

For more information about the planning and configuration of storage efficiency features, see the following publications:

- ▶ *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521
- ▶ *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430

This chapter includes the following topics:

- ▶ 9.1, “IBM Easy Tier” on page 510
- ▶ 9.2, “Thin-provisioned volumes” on page 528
- ▶ 9.3, “UNMAP” on page 530
- ▶ 9.4, “Data Reduction Pools” on page 533
- ▶ 9.5, “Saving estimations for compression and deduplication” on page 545
- ▶ 9.6, “Overprovisioning and data reduction on external storage” on page 548

9.1 IBM Easy Tier

IBM Spectrum Virtualize includes the IBM System Storage Easy Tier function, which enables automated subvolume data placement throughout different storage tiers, and automatically moves extents within the same storage tier to intelligently align the system with workload requirements. Easy Tier works with all available storage tiers and drive modules: storage-class memory (SCM), flash drives, and hard disk drives (HDDs).

Many applications exhibit a significant skew in the distribution of I/O workload: A small fraction of the storage is responsible for a disproportionately large fraction of the total I/O workload of an environment.

Easy Tier acts to identify this skew and automatically place data to take advantage of it. By moving the “hottest” data onto the fastest tier of storage, the workload on the remainder of the storage is reduced. By servicing most of the application workload from the fastest storage, Easy Tier accelerates application performance and increases overall server utilization, which can reduce costs regarding servers and application licenses.

Easy Tier also reduces storage cost because the system always places the data with the highest I/O workload on the fastest tier of storage. Depending on the workload pattern, a large portion of the capacity can be provided by a lower and less expensive tier without impacting application performance.

Note: Easy Tier is a licensed function. On IBM FlashSystem 9200 and IBM FlashSystem 7200, it is included in the base code. No actions are required to activate the Easy Tier license on these systems.

On IBM FlashSystem 5100, you must have the appropriate number of licenses to run Easy Tier.

The IBM FlashSystem 5000 entry systems also require a license for Easy Tier, which is a one time charge per system.

Without a license, Easy Tier balances I/O workload only between managed disks (MDisks) in the same tier.

In HyperSwap environments, all member controllers must be licensed with Easy Tier to enable this function. For example, when clustering two IBM FlashSystem 5030 system, you need two licenses.

9.1.1 Easy Tier concepts

Easy Tier is a performance optimization function that automatically migrates (or moves) extents that belong to a volume between different storage tiers based on their I/O load. Movement of the extents is online and unnoticed from the host point of view. As a result of extent movement, the volume no longer has all its data in one tier, but rather in two or more tiers, and each tier provides optimal performance for the extent, as shown in Figure 9-1 on page 511.

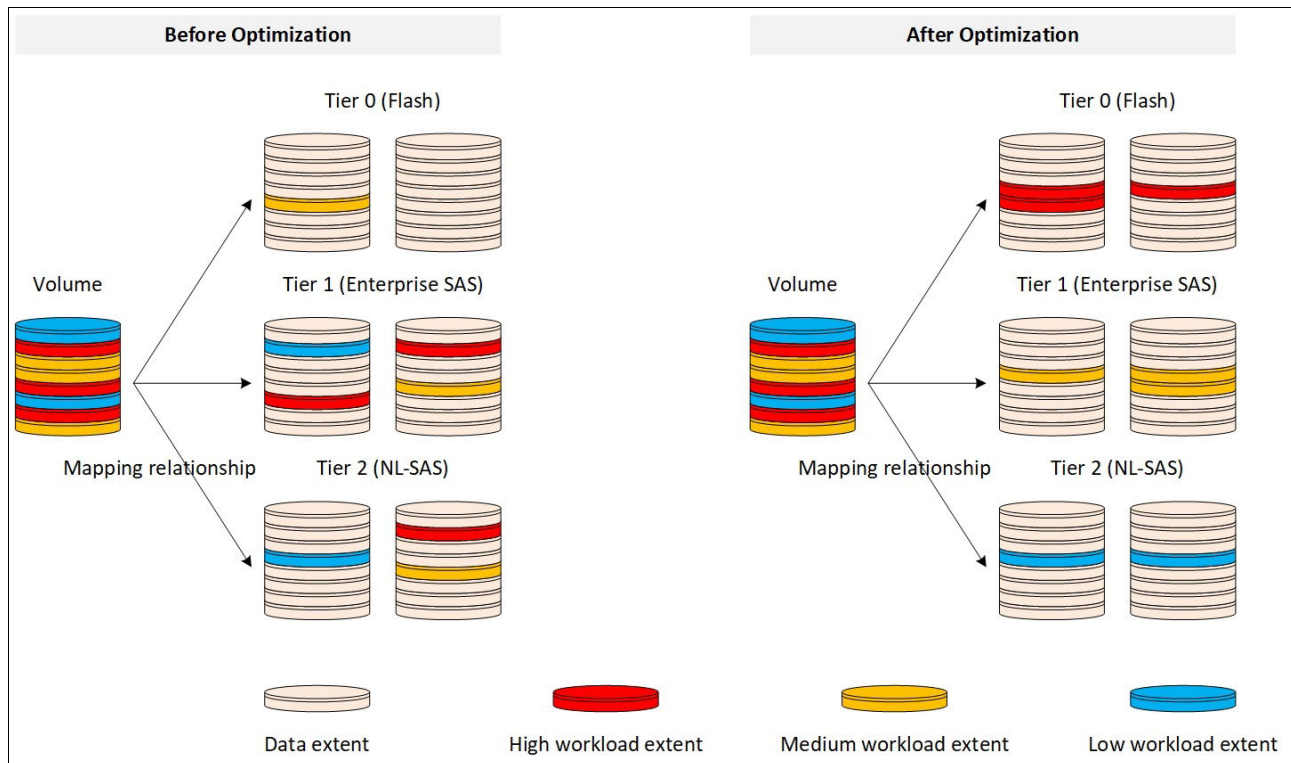


Figure 9-1 Easy Tier

Easy Tier monitors the I/O activity and latency of the extents on all Easy Tier enabled storage pools. Based on the performance log, it creates an extent migration plan and *promotes* (moves) high activity or hot extents to a higher disk tier within the same storage pool. It also *demotes* extents whose activity dropped off, or cooled, by moving them from a higher disk tier MDisk back to a lower tier MDisk.

If a pool contains only MDisks of a single tier, Easy Tier operates only in balancing mode. Extents are moved between MDisks in the same tier to balance I/O workload within that tier.

Tiers of storage

The MDisks (external logical units (LUs) or redundant array of independent disks (RAID) arrays) that are presented to the system might have different performance attributes because of their technology type, such as flash drives or HDDs and other characteristics.

The system divides available storage into the following tiers:

- ▶ SCM

The SCM tier is used when the pool contains drives that use persistent memory technologies that improve the endurance and speed of current flash storage device technologies. SCM drives are available only in Non-Volatile Memory Express (NVMe) based controller systems.

- ▶ Tier 0 flash

Tier 0 flash drives are high-performance flash drives that use enterprise flash technology.

- ▶ Tier 1 flash

Tier 1 flash drives represent the read-intensive flash drive technology. Tier 1 flash drives are lower-cost flash drives that typically offer capacities larger than enterprise-class flash, but have lower performance and write endurance characteristics.

► Enterprise tier

The enterprise tier is used when the pool contains MDisks on enterprise-class hard disk drives (HDDs), which are disk drives that are optimized for performance.

► Nearline (NL) tier

The NL tier is used when the pool has MDisks on NL-class disks drives that are optimized for capacity.

The system automatically sets the tier for internal array mode MDisks because it knows the capabilities of array members, physical drives, and modules. External MDisks need manual tier assignment when they are added to a storage pool.

Note: The tier of MDisks that is mapped from certain types of IBM System Storage Enterprise Flash is fixed to tier0_flash, and cannot be changed.

Although the system can distinguish between five tiers, Easy Tier manages only a three-tier storage architecture within each storage pool. MDisk tiers are mapped to Easy Tier tiers depending on the pool configuration, as shown in Table 9-1.

Table 9-1 Storage tier to Easy Tier mapping

Configuration	Easy Tier top tier	Easy Tier middle tier	Easy Tier bottom tier
SCM (+ Tier0_Flash)	SCM	(Tier0_Flash)	
SCM + Tier0_Flash (+ Tier1_Flash)	SCM	Tier0_Flash	(Tier1_Flash)
SCM + Tier0_Flash (+ Tier1_Flash) + Enterprise + NL (unsupported)	SCM	Tier0_Flash (+ Tier1_Flash)	Enterprise + NL
SCM + Tier0_Flash + Enterprise/NL	SCM	Tier0_Flash	Enterprise/NL
SCM + Tier0_Flash + Tier1_Flash + Enterprise/NL (unsupported)	SCM	Tier0_Flash + Tier1_Flash	Enterprise/NL
SCM + Tier1_Flash (+ Enterprise/NL)	SCM	Tier1_Flash	(Enterprise/NL)
SCM + Tier1_Flash + Enterprise + NL	SCM	Tier1_Flash + Enterprise	NL
SCM + Enterprise/NL	SCM	Enterprise/NL	
SCM + Enterprise + NL	SCM	Enterprise	NL
Tier0_Flash (+ Tier1_Flash)	Tier0_Flash	(Tier1_Flash)	
Tier0_Flash + Tier1_Flash + Enterprise/NL	Tier0_Flash	Tier1_Flash	Enterprise/NL

Configuration	Easy Tier top tier	Easy Tier middle tier	Easy Tier bottom tier
Tier0_Flash + Tier1_Flash + Enterprise + NL	Tier0_Flash	Tier1_Flash + Enterprise	NL
Tier0_Flash + Enterprise (+ NL)	Tier0_Flash	Enterprise	(NL)
Tier0_Flash + NL	Tier0_Flash	NL	
Tier1_Flash (+ Enterprise/NL)		Tier1_Flash	(Enterprise/NL)
Tier1_Flash + Enterprise + NL	Tier1_Flash	Enterprise	NL
Enterprise (+ NL)		Enterprise	(NL)
NL			NL

The table represents all the possible pool configurations. Some entries in the table contain *optional* tiers (shown in *italic* font), but the configurations without the optional tiers are also valid.

Sometimes, a single Easy Tier tier contains MDisk from more than one storage tier. For example, consider a pool with SCM, Tier1_Flash, Enterprise, and NL. SCM is the top tier, and Tier1_Flash and Enterprise share the middle tier. NL is represented by the bottom tier.

Note: Some storage pool configurations with four or more different tiers are not supported. If such a configuration is detected, an error is logged and Easy Tier enters measure mode, which means no extent migrations are performed.

For more information about planning and configuration considerations or best practices, see *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521.

Easy Tier automatic data placement

Easy Tier continuously monitors volumes for host I/O activity. It collects performance statistics for each extent, and derives exponential moving averages for a rolling 24-hour period of I/O activity. Random and sequential I/O rate, I/O block size and bandwidth for reads and writes, and I/O response time are collected.

A set of algorithms is used to decide where the extents should be and whether extent relocation is required. Once per day, Easy Tier analyzes the statistics to determine which data should be sent to a higher performing tier or a lower tier. Four times per day, it analyzes the statistics to identify whether any data must be rebalanced between MDisk in the same tier. Once every 5 minutes, Easy Tier checks the statistics to identify whether any of the MDisk are overloaded.

Based on this information, Easy Tier generates a migration plan that must be run for optimal data placement. The system spends the necessary time running the migration plan. The migration rate is limited to make sure host I/O performance is not affected while data is relocated.

The migration plan can consist of the data movement actions on volume extents, as shown in Figure 9-2. Although each action is shown once, all movement actions can be performed between any pair of adjacent tiers.

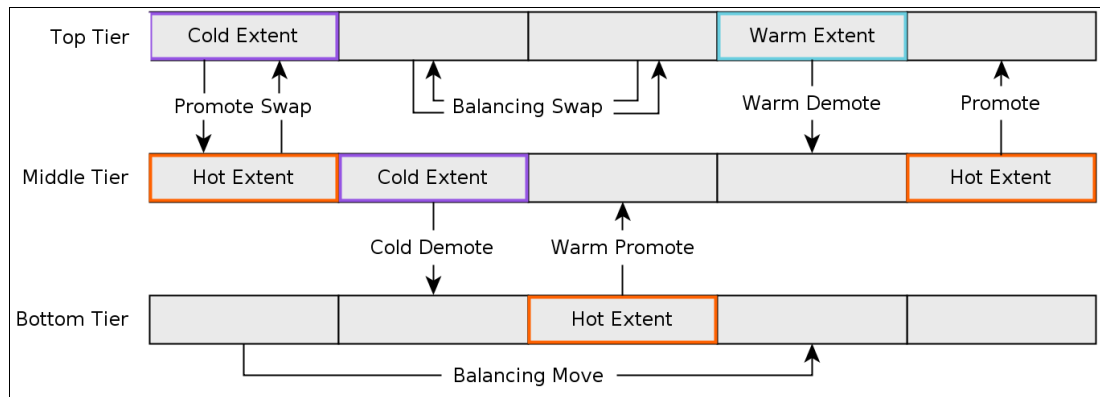


Figure 9-2 Actions on extents

Here are the possible actions:

- ▶ Promote

Active data is moved from a lower tier of storage to a higher tier to improve the overall system performance.

- ▶ Promote Swap

Active data is moved from a lower tier of storage to a higher tier to improve overall system performance. Less active data is moved first from the higher tier to the lower tier to make space.

- ▶ Warm Promote

When an MDisk becomes overloaded, active data is moved from a lower tier to a higher tier to reduce the workload on the MDisk, which addresses the situation where a lower tier suddenly becomes very active. Instead of waiting for the next migration plan, Easy Tier can react immediately. Warm promote acts in a similar way to warm demote. If the 5-minute average performance shows that a layer is overloaded, Easy Tier immediately starts to promote extents until the condition is relieved.

- ▶ Cold Demote

Inactive or less active data is moved from a higher tier of storage to a lower tier to free space on the higher tier. Easy Tier automatically frees extents on the higher storage tier before the extents on the lower tier become hot, which helps the system to be more responsive to new hot data.

- ▶ Warm Demote

When an MDisk becomes overloaded, active data is moved from a higher tier to a lower tier to reduce the workload on the MDisk. Easy Tier continuously ensures that the higher performance tier does not suffer from saturation or overload conditions that might affect the overall performance in the pool. This action is triggered when bandwidth or input/output operations per second (IOPS) exceeds a predefined threshold of an MDisk and causes the movement of selected extents from the higher-performance tier to the lower-performance tier to prevent MDisk overload.

- ▶ **Balancing Move**

Data is moved within the same tier from an MDisk with a higher workload to one with a lower workload to balance the workload within the tier, which automatically populates new MDisks that were added to the pool.

- ▶ **Balancing Swap**

Data is moved within the same tier from an MDisk with higher workload to one with a lower workload to balance the workload within the tier. Other less active data is moved first to make space.

Extent migration occurs at a maximum rate of 12 GB every 5 minutes for the entire system. It prioritizes the following actions:

- ▶ Promote and rebalance get equal priority.
- ▶ Demote is 1 GB every 5 minutes, and then gets whatever is left.

Note: Extent promotion or demotion occurs only between adjacent tiers. In a three-tier storage pool, Easy Tier does not move extents from the top directly to the bottom tier or vice versa without moving to the middle tier first.

The Easy Tier overload protection is designed to avoid overloading any type of MDisk with too much work. To achieve this task, Easy Tier must have an indication of the maximum capability of a MDisk.

For an array made of locally attached drives, the system can calculate the performance of the MDisk because it is pre-programmed with performance characteristics for different drives and array configurations. For a storage area network (SAN)-attached MDisk, the system cannot calculate the performance capabilities. Therefore, follow the best practice guidelines when configuring external storage, particularly the ratio between physical disks and MDisks that is presented to the system.

Each MDisk has an Easy Tier load parameter (low, medium, high, or very_high) that can be fine-tuned manually. If you analyze the statistics and find that the system does not appear to be sending enough IOPS to your external MDisk, you can increase the load parameter.

Easy Tier operating modes

Easy Tier includes the following main operating modes:

- ▶ **Off**

When off, no statistics are recorded and no cross-tier extent migration occurs. Also, with Easy Tier turned off, no storage pool balancing across MDisks in the same tier is performed, even in single-tier pools.

- ▶ **Evaluation or measurement only**

When in this mode, Easy Tier collects only usage statistics for each extent in a storage pool if it is enabled on both the volume and the pool. No extents are moved. This collection is typically done for a single-tier pool that contains only HDDs so that the benefits of adding flash drives to the pool can be evaluated before any major hardware acquisition.

- ▶ **Automatic data placement and storage pool balancing**

In this mode, usage statistics are collected and extent migration is performed between tiers (if there is more than one pool in a tier). Also, auto-balancing among MDisks in each tier is performed.

The default operation mode is *Enabled*. Therefore, the system balances storage pools. If the required licenses are installed, they also optimize performance.

Note: The auto-balance process automatically balances existing data when MDisks are added to a pool. However, the process does not migrate extents from existing MDisks to achieve even extent distribution among all old and new MDisks in the storage pool. The Easy Tier migration plan is based on performance. It is *not* based on the capacity of the underlying MDisks or on the number of extents on them.

Implementation considerations

Consider the following implementation and operational rules when you use the IBM System Storage Easy Tier function on the storage system:

- ▶ If the system contains self-compressing drives (IBM FlashCore Module (FCM) drives) in the top tier of storage in a pool with multiple tiers and Easy Tier is in use, consider setting an overallocation limit within these pools, as described in “Overallocation limit” on page 521.
- ▶ Volumes that are added to storage pools use extents from the “middle” tier of three-tier model, if available. Easy Tier then collects usage statistics to determine which extents to move to “faster” or “slower” tiers. If there are no free extents in the middle tier, extents from the other tiers are used (bottom tier if possible, otherwise top tier).
- ▶ When an MDisk with allocated extents is deleted from a storage pool, extents in use are migrated to MDisks in the same tier as the MDisk that is being removed, if possible. If insufficient extents exist in that tier, extents from another tier are used.
- ▶ Easy Tier monitors the extent I/O activity of each copy of a mirrored volume. Easy Tier works with each copy independently of the other copy. This situation applies to volume mirroring and IBM HyperSwap and Remote Copy (RC).

Note: Volume mirroring can have different workload characteristics on each copy of the data because reads are normally directed to the primary copy and writes occur to both copies. Therefore, the number of extents that Easy Tier migrates between the tiers might differ for each copy.

- ▶ Easy Tier automatic data placement is not supported on image mode or sequential volumes. However, it supports evaluation mode for such volumes. I/O monitoring is supported and statistics are accumulated.
- ▶ When a volume is migrated out of a storage pool that is managed with Easy Tier, Easy Tier automatic data placement mode is no longer active on that volume. Automatic data placement is also turned off while a volume is being migrated, even when it is between pools that both have Easy Tier automatic data placement enabled. Automatic data placement for the volume is reenabled when the migration is complete.

When the system migrates a volume from one storage pool to another, it attempts to migrate each extent to an extent in the new storage pool from the same tier as the original extent, if possible.

- ▶ When Easy Tier automatic data placement is enabled for a volume, you cannot use the `svctask migrateexts` command on that volume.

9.1.2 Implementing and tuning Easy Tier

The Easy Tier function is enabled by default. It starts monitoring I/O activity immediately after the storage pool and volumes are created.

Without the proper licenses installed, the system only rebalances storage pools.

A few parameters can be adjusted. Also, Easy Tier can be turned off on selected volumes in storage pools.

MDisk settings

The tier for internal (array) MDisks is detected automatically and depends on the type of drives, which are its members. No adjustments are needed.

For an external MDisk, the tier is assigned when it is added to a storage pool. To assign the MDisk, select **Pools** → **External Storage**, select the MDisk (or MDisks) to add, and click **Assign**.

Note: The tier of MDisks that is mapped from certain types of IBM System Storage Enterprise Flash is fixed to tier0_flash and cannot be changed.

You can choose the target storage pool and storage tier that is assigned, as shown in Figure 9-3.

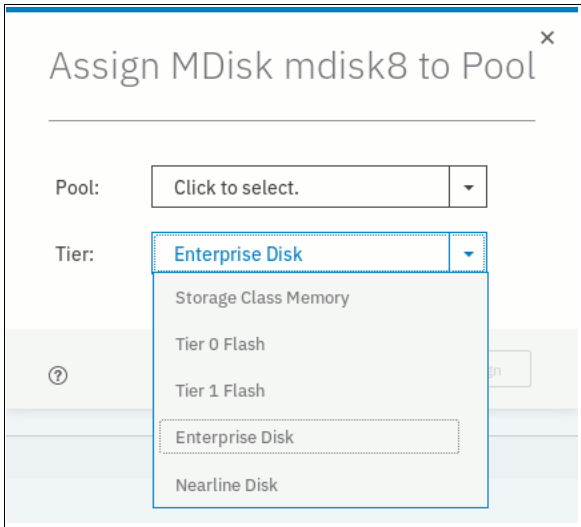


Figure 9-3 Choosing the tier when assigning MDisks

To change the storage tier for an MDisk that is assigned, select **Pools** → **External Storage**, right-click one or more selected MDisks, and choose **Modify Tier**, as shown in Figure 9-4.

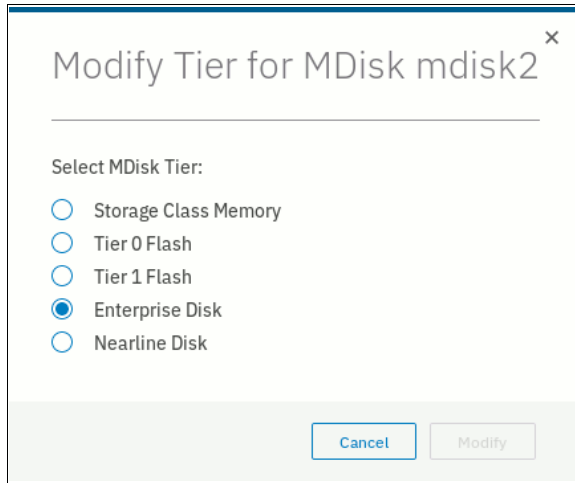


Figure 9-4 Changing the MDisk tier

Note: Assigning a tier to an external MDisk that does not match the physical back-end storage type is not supported by IBM and can lead to unpredictable consequences.

To determine what tier is assigned to an MDisk, select **Pools** → **External Storage**, select **Actions** → **Customize columns**, and select **Tier**. This action includes the current tier setting into a list of MDisk parameters that are shown in the External Storage window. You can also find this information in MDisk properties. To show this information, right-click **MDisk**, select **Properties**, and click **View more details**, as shown in Figure 9-5.

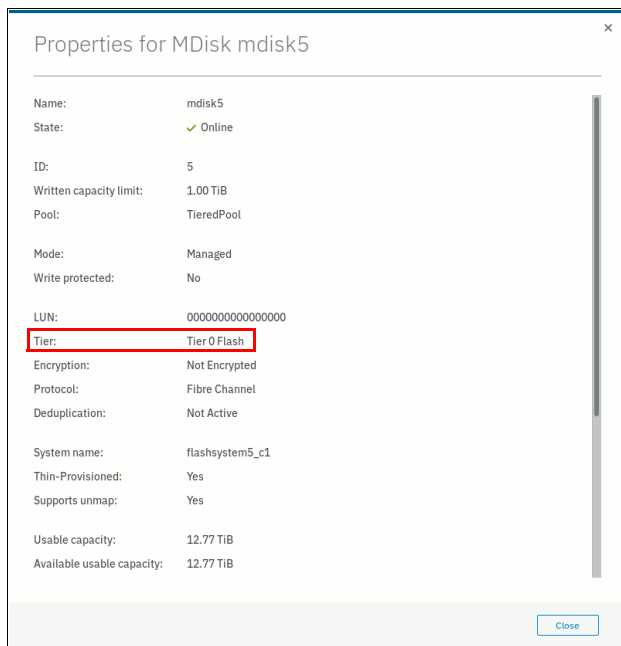


Figure 9-5 MDisk properties

To list MDisk parameters with the command-line interface (CLI), run the `lsmdisk` command. The current tier for each MDisk is shown. To change the external MDisk tier, run the `chmdisk` command with the `-tier` parameter, as shown in Example 9-1.

Example 9-1 Listing and changing tiers for MDisks (partially shown)

```
IBM FlashSystem 7200:ITS0FS7K:superuser>lsmdisk
id name  status mode      mdisk_grp_id ... tier          encrypt
1 mdisk1 online unmanaged ... tier0_flash no
2 mdisk2 online managed  0          ... tier_enterprise no
3 mdisk3 online managed  0          ... tier_enterprise no
<...>
IBM FlashSystem 7200:ITS0FS7K:superuser>chmdisk -tier tier1_flash mdisk2
IBM FlashSystem 7200:ITS0FS7K:superuser>
```

For an external MDisk, the system cannot calculate its exact performance capabilities, so it has several predefined levels. In rare cases, statistics analysis might show that Easy Tier is overusing or underusing an MDisk. If so, levels can be adjusted only by using the CLI. Run `chmdisk` with the `-easytierload` parameter. To reset the Easy Tier load to the system default for the chosen MDisk, use `-easytier default`, as shown in Example 9-2.

Example 9-2 Changing the Easy Tier load

```
IBM FlashSystem 7200:ITS0FS7K:superuser>chmdisk -easytierload default mdisk2
IBM FlashSystem 7200:ITS0FS7K:superuser>
IBM FlashSystem 7200:ITS0FS7K:superuser>lsmdisk mdisk2 | grep tier
tier tier_enterprise
easy_tier_load high
IBM FlashSystem 7200:ITS0FS7K:superuser>
```

Note: Adjust the Easy Tier load settings only if instructed to do so by IBM Technical Support or your solution architect.

To list the current Easy Tier load setting of an MDisk, run `lsmdisk` with the MDisk name or ID as a parameter.

Storage pool settings

When a storage pool (either standard pool or Data Reduction Pool (DRP)) is created, Easy Tier is turned on by default. The system automatically enables Easy Tier functions when the storage pool contains an MDisk from more than one tier. It also enables automatic rebalancing when the storage pool contains an MDisk from only one tier.

You can disable Easy Tier or switch it to measure-only mode when creating a pool or any other time. This task cannot be done by using the GUI, but can be done by using the CLI.

To check the current Easy Tier function state on a pool, select **Pools** → **Pools**, right-click the selected pool, click **Properties**, and expand **View more details**, as shown in Figure 9-6. This window also shows the amount of data that is stored on each tier.

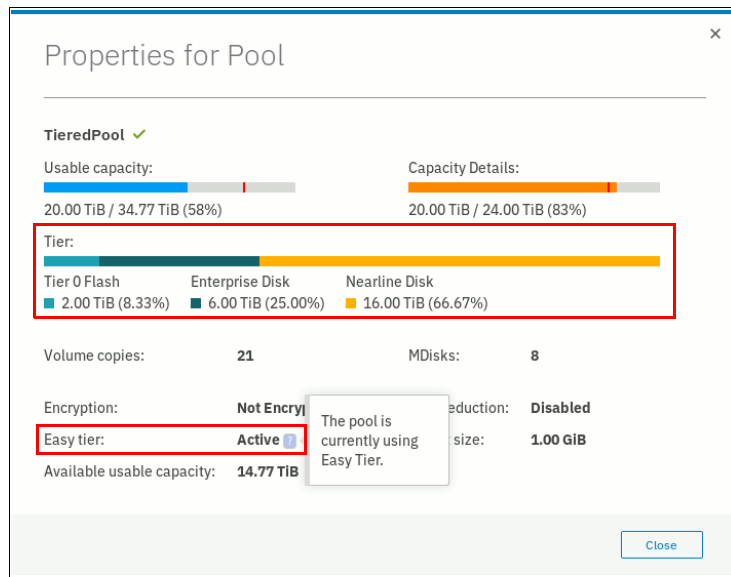


Figure 9-6 Pool properties

Easy Tier can be in one of the following states:

- ▶ **Active**
Indicates that a pool is being managed by Easy Tier, and extent migrations between tiers can be performed. Performance-based pool balancing of MDisks in the same tier is also enabled. This state is the expected one for a pool with two or more tiers of storage.
- ▶ **Balanced**
Indicates that a pool is being managed by Easy Tier to provide performance-based pool balancing of MDisks in the same tier. This state is the expected one for a pool with a single tier of storage.
- ▶ **Inactive**
Indicates that Easy Tier is inactive (disabled).
- ▶ **Measured**
Shows that Easy Tier statistics are being collected but no extent movement can be performed.

To find the state of the Easy Tier function on the pools by using the CLI, run the **lsmdiskgrp** command without any parameters. To turn off or on Easy Tier, run the **chmdiskgrp** command, as shown in Example 9-3. By running **lsmdiskgrp** with pool name/ID as a parameter, you can also determine how much storage of each tier is available within the pool.

Example 9-3 Listing and changing the Easy Tier status on pools

```

IBM FlashSystem 7200:ITS0FS7K:superuser>lsmdiskgrp
id name      status mdisk_count ... easy_tier easy_tier_status
0 TieredPool online 1           ... auto    balanced
IBM FlashSystem 7200:ITS0FS7K:superuser>chmdiskgrp -easytier measure TieredPool
IBM FlashSystem 7200:ITS0FS7K:superuser>chmdiskgrp -easytier auto TieredPool
IBM FlashSystem 7200:ITS0FS7K:superuser>

```


Overallocation limit

If the system contains self-compressing drives (FCM drives) in the top tier of storage in a pool with multiple tiers and Easy Tier is in use, consider setting an *overallocation limit* within these pools. The overallocation limit has no effect in pools with a different configuration.

Arrays that are created from self-compressing drives have a written capacity limit (virtual capacity before compression) that is higher than the array's usable capacity (physical capacity). Writing highly compressible data to the array means that the written capacity limit can be reached without running out of usable capacity. However, if data is not compressible or the compression ratio is low, it is possible to run out of usable capacity before reaching the written capacity limit of the array, which means the amount of data that is written to a self-compressing array must be controlled to prevent the array from running out of space.

Without a maximum overallocation limit, Easy Tier scales the usable capacity of the array based on the actual compression ratio of the data that is stored on the array at a point in time (PiT). Easy Tier migrates data to the array and might use a large percentage of the usable capacity in doing so, but it stops migrating to the array when the array comes close to running out of usable capacity. Then, it might start migrating data away from the array again to free space.

However, Easy Tier migrates storage only at a slow rate, which might not keep up with *changes* to the compression ratio within the tier. When Easy Tier swaps extents or data is overwritten by hosts, compressible data might be replaced with data that is less compressible, which increases the amount of usable capacity that is consumed by extents and might result in self-compressing arrays running out of space, which can cause a loss of access to data until the condition is resolved.

So, the user might specify the maximum overallocation ratio for pools that contain self-compressing arrays to prevent out-of-space scenarios. The value acts as a multiplier of the physically available space in self-compressing arrays. The allowed values are a percentage in the range of 100% (default) to 400% or off. The default setting allows no overallocation on new pools. Setting the value to off disables this feature.

When enabled, Easy Tier scales the available usable capacity of self-compressing arrays by using the specified overallocation limit and adjusts the migration plan to make sure the fullness of these arrays stays below the maximum overallocation. Specify the maximum overallocation limit based on the estimated lowest compression ratio of the data that is written to the pool.

For example, for an estimated compression ratio of 1.2:1, specify an overallocation limit of 120% to put a limit on the overallocation. Easy Tier stops migrating data to self-compressing arrays in the pool after the written capacity reaches 120% of the physical (usable) capacity of the array, which is the case even if the written capacity limit of the array is not reached yet or the current compression ratio of the data that is stored on the array is higher than 1.2:1 (and thus more usable capacity would be available). This setting prevents changes to the compression ratio within the specified limits from causing the array to run out of space.

To modify the maximum overallocation limit of a pool by using the GUI, select **Pools** → **Pools**, right-click a pool, and select **Modify Overallocation Limit**, as shown in Figure 9-7.

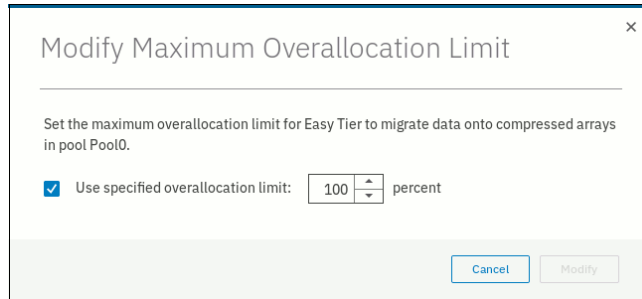


Figure 9-7 Modifying the pool overallocation limit

On the CLI, run the `chmdiskgrp` command with the `-etfcmoverallocationmax` parameter to set a percentage or use `off` to disable the limit.

Volume settings

By default, each striped-type volume enables Easy Tier to manage its extents. If you need to fix the volume extent location (for example, to prevent extent demotes and to keep the volume in the higher-performing tier), you can turn off Easy Tier management for a particular volume copy.

Note: Thin-provisioned and compressed volumes in a DRP cannot have Easy Tier turned off. You can turn off Easy Tier only at a pool level.

You can do this task only by using the CLI. Run the `lsvdisk` command to check and the `chvdisk` command to modify the Easy Tier function status on a volume copy, as shown in Example 9-4.

Example 9-4 Checking and modifying the Easy Tier settings on a volume

```
IBM_Storwize:ITS0-V7k:superuser>lsvdisk vdisk0 |grep easy_tier
easy_tier on
easy_tier_status balanced
IBM_Storwize:ITS0-V7k:superuser>chvdisk -easytier off vdisk0
IBM_FlashSystem 7200:ITS0FS7K:superuser>
```

System-wide settings

There is a system-wide setting that is called *Easy Tier acceleration* that is disabled by default. Turning it on makes Easy Tier move extents up to four times faster than the default setting. In acceleration mode, Easy Tier can move up to 48 GiB per 5 minutes, but in normal mode it moves up to 12 GiB. The following use cases are the most probable use cases for acceleration:

- ▶ When adding capacity to the pool either by adding to an existing tier or by adding a tier to the pool, accelerating Easy Tier can quickly spread volumes onto the new MDisks.
- ▶ Migrating the volumes between the storage pools when the target storage pool has more tiers than the source storage pool, so Easy Tier can quickly promote or demote extents in the target pool.

Note: Enabling Easy Tier acceleration is advised only during periods of low system activity only after migrations or storage reconfiguration occurred. It is a best practice to keep off the Easy Tier acceleration mode during normal system operation to avoid performance impacts that are caused by accelerated data migrations.

This setting can be changed non-disruptively, but only by using the CLI. To turn on or off Easy Tier acceleration mode, run the `chsystem` command. Run the `lssystem` command to check its current state, as shown in Example 9-5.

Example 9-5 The chsystem command

```
IBM FlashSystem 7200:ITS0FS7K:superuser>lssystem |grep easy_tier
easy_tier_acceleration off
IBM FlashSystem 7200:ITS0FS7K:superuser>chsystem -easytieracceleration on
IBM FlashSystem 7200:ITS0FS7K:superuser>
```

9.1.3 Monitoring Easy Tier activity

When Easy Tier is active, it constantly monitors and records I/O activity, collecting extent heat data. Heat data files are produced approximately once a day and summarize the activity per volume since the last heat data file was produced. Easy Tier activity can be monitored by using the GUI or the external IBM Storage Tier Advisor Tool (IBM STAT) application.

Monitoring Easy Tier by using the GUI

To view the most recent Easy Tier statistics, select **Monitoring** → **Easy Tier Reports**. Select the storage pool that you want to see reports for in the filter section on the left, as shown in Figure 9-8. It takes approximately 24 hours for reports to be available after turning on Easy Tier or after a configuration node failover occurred. If no reports are available, the error message in the figure is shown. In this case, wait until new reports were generated and then revisit the GUI.

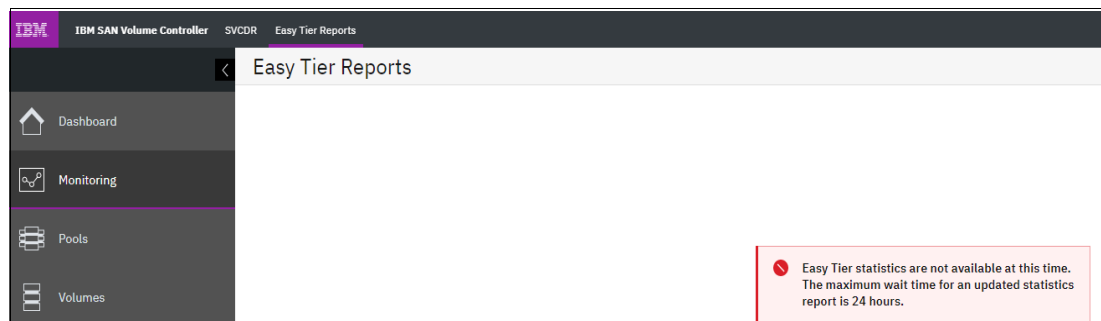


Figure 9-8 Easy Tier reports not available

Three types of reports are available per storage pool: Data Movement, Tier Composition, and Workload Skew Comparison. Select the corresponding tabs in the GUI to view the charts. Alternatively, click **Export** or **Export All** to download the reports in comma-separated value (CSV) format.

Data Movement report

The Data Movement report shows the extent migrations that Easy Tier performed to relocate data between different tiers of storage and within the same tier for optimal performance, as shown in Figure 9-9. The chart displays the data for the previous 24-hour period, in one-hour increments. You can change the time span and the increments for a more detailed view.

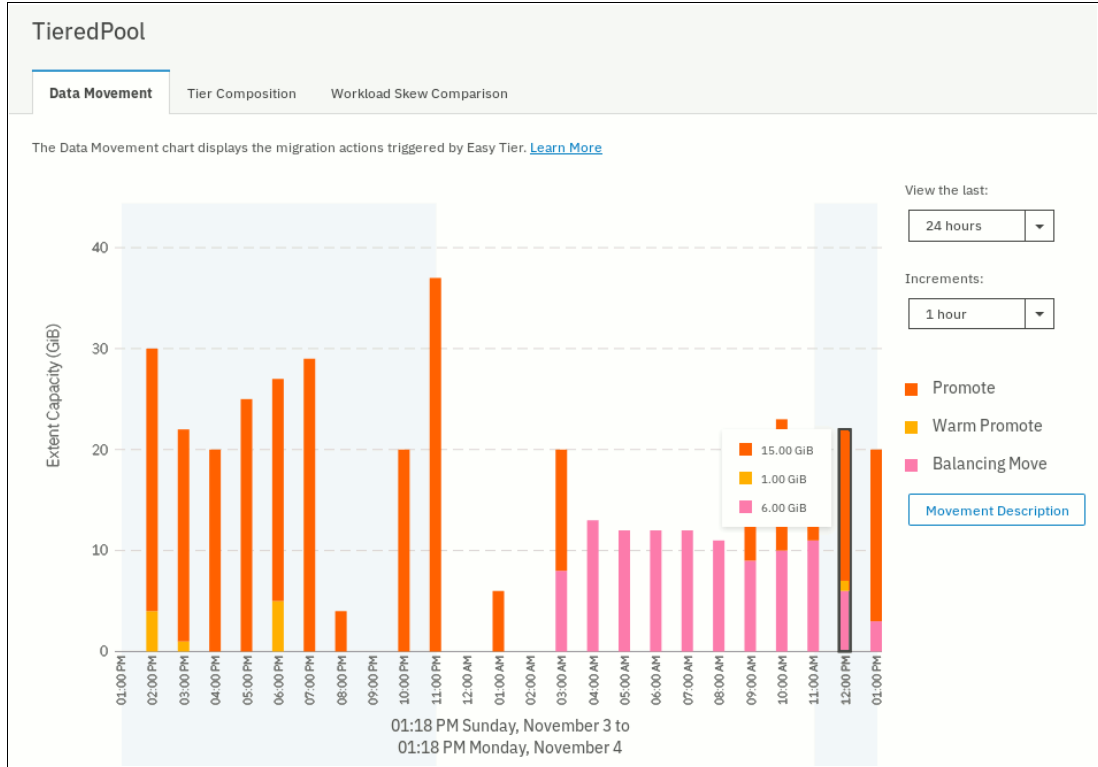


Figure 9-9 Easy Tier Data Movement report

The X-axis shows a timeline for the selected period by using the selected increments. The Y-axis indicates the amount of extent capacity that is moved. For each time increment, a color-coded bar displays the amount of data that is moved by each Easy Tier data movement action, such as promote or cold demote. For more information about the different movement actions, see “Easy Tier automatic data placement” on page 513 or click **Movement Description** next to the chart to see an explanation in the GUI.

Tier Composition chart

The Tier Composition chart shows how different types of workloads are distributed between top, middle, and bottom tiers of storage in the selected pool, as shown in Figure 9-10 on page 525.

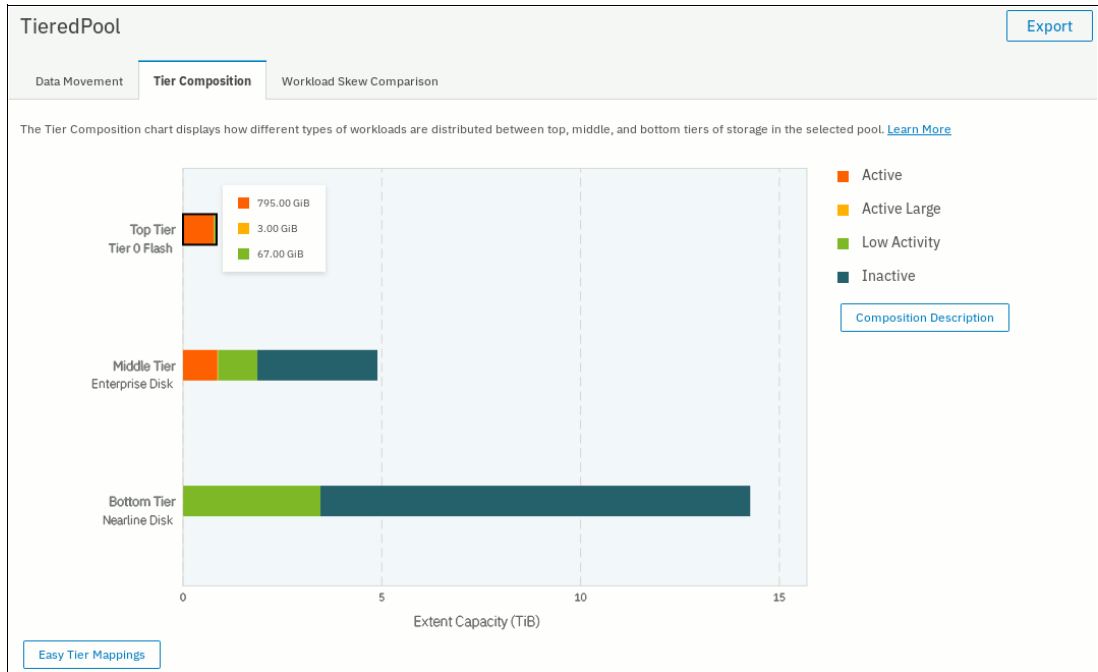


Figure 9-10 Tier Composition chart

A color-coded bar for each tier shows which workload types are present in that tier and how much of the extent capacity in that tier to which they can be attributed. Easy Tier distinguishes between the following workload types. Click **Composition Description** to show a short explanation for each workload type in the GUI.

- ▶ **Active**
Data with more than 0.1 IOPS / Extent access density for small IOPS (< 64 KB block size)
- ▶ **Active Large**
All data that is not classified above (> 64 KB block size)
- ▶ **Low Activity**
Data with less than 0.1 IOPS / Extent access density
- ▶ **Inactive**
Data with zero IOPS / Extent access density (no recent activity)

Click **Easy Tier Mappings** to show which MDisks are assigned to which of the three tiers of Easy Tier, as shown in Figure 9-11. How storage tiers in the system are mapped to Easy Tier tiers depends on the available storage tiers in the pool. For a list of all possible mappings, see Table 9-1 on page 512.

ID	MDisk Name	Tier	Easy Tier Group
0	mdisk0	Enterprise Disk	Middle Tier
1	mdisk1	Enterprise Disk	Middle Tier
2	mdisk2	Enterprise Disk	Middle Tier
3	mdisk3	Enterprise Disk	Middle Tier
4	mdisk4	Nearline Disk	Bottom Tier
5	mdisk5	Tier 0 Flash	Top Tier
6	mdisk6	Nearline Disk	Bottom Tier
7	mdisk7	Tier 0 Flash	Top Tier

Showing 8 Easy Tier Mappings / Selecting 0 Easy Tier Mappings

Figure 9-11 Easy Tier mappings

Workload Skew Comparison

The Workload Skew Comparison chart displays the percentage of I/O workload that is attributed to a percentage of the total capacity, as shown in Figure 9-12.

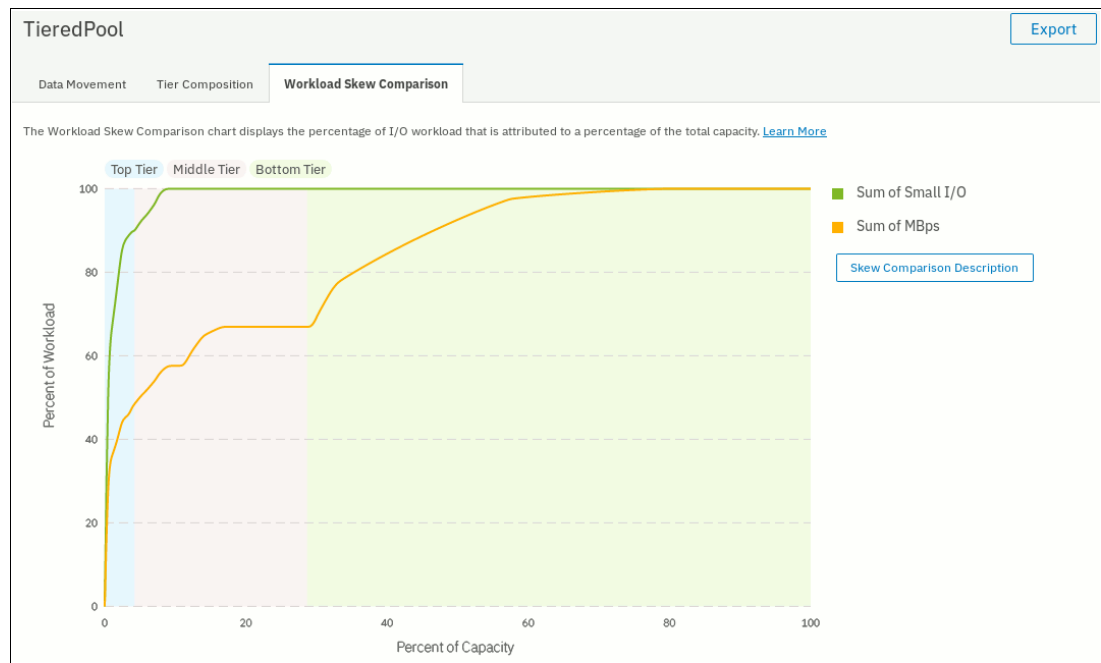


Figure 9-12 Easy Tier Workload Skew Comparison

The X-axis shows the percentage of capacity and the Y-axis shows the corresponding percentage of workload on that capacity. Workload is classified in small I/O (sum of small reads and writes) and megabytes per second (MBps) (sum of small and large bandwidth). The portion of capacity and workload that is attributed to a tier is color-coded in the chart with a legend above the chart.

Figure 9-12 on page 526 shows that the top tier (Tier1 Flash) contributes only a tiny percentage of capacity to the pool, but handles around 85% of the IOPS and more than 40% of the bandwidth in that pool. The middle tier (enterprise disk) handles almost all the remaining IOPS and an extra 20% of the bandwidth. The bottom tier (NL disk) provides most of the capacity to the pool but does almost no small I/O workload.

Use this chart to estimate how much storage capacity in the high tiers must be available to handle most of the workload.

Monitoring Easy Tier by using the IBM Storage Tier Advisor Tool

The IBM STAT is a Windows console application that can analyze heat data files that are generated by Easy Tier and produce a graphical display of the amount of “hot” data per volume and predictions of the performance benefits of adding more capacity to a tier in a storage pool.

Using this method of monitoring, Easy Tier can provide more insights on top of the information that is available in the GUI.

IBM STAT can be downloaded from this IBM Support [web page](#).

You can download the IBM STAT and install it on your Windows-based computer. The tool is packaged as an ISO file that must be extracted to a temporary location.

The tool installer is at `temporary_location\IMAGES\STAT\Disk1\InstData\NoVM\`. By default, the IBM STAT is installed in `C:\Program Files\IBM\STAT\`.

On the system, the heat data files are found in the `/dumps/easytier` directory on the configuration node and are named `dpa_heat.node_panel_name.time_stamp.data`. Any heat data file is erased when it exists for longer than 7 days.

Heat files must be offloaded and IBM STAT started from a Windows command prompt console with the file specified as a parameter, as shown in Example 9-6.

Example 9-6 Running IBM STAT by using the Windows command prompt

```
C:\Program Files (x86)\IBM\STAT>stat dpa_heat.78DXRY0.191021.075420.data
```

The IBM STAT creates a set of `.html` and `.csv` files that can be used for Easy Tier analysis.

To download a heat data file, select **Settings** → **Support** → **Support Package** → **Download Support Package** → **Download Existing Package**, as shown in Figure 9-13.

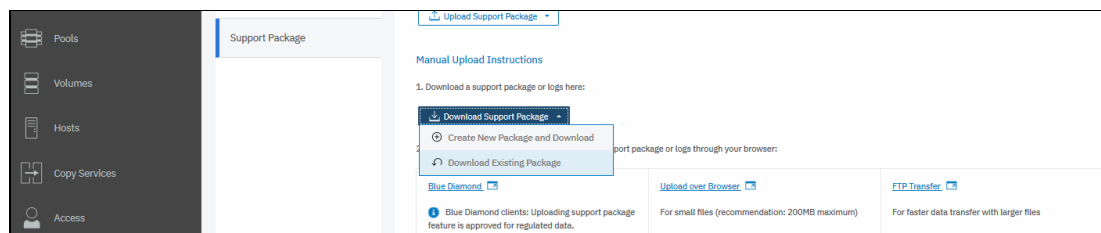


Figure 9-13 Downloading an Easy Tier heat file: Download Support Package

A download window opens that shows all the files in /dumps and its subfolders on a **configuration** node. You can filter the list by using the “easytier” keyword, select the dpa_heat file or files that are analyzed, and clicking **Download**, as shown in Figure 9-14. Save them in a convenient location (for example, to a subfolder that holds the IBM STAT executable file).

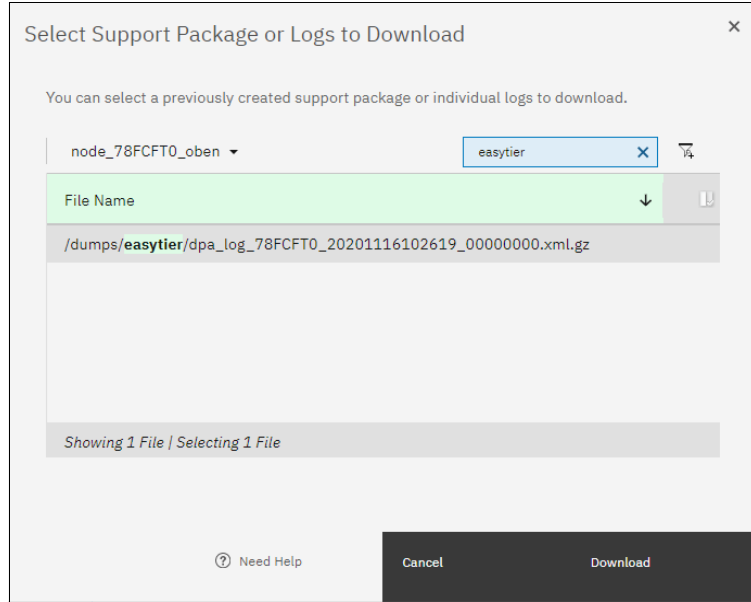


Figure 9-14 Downloading Easy Tier heat data file: dpa_heat files

You can also specify the output directory. IBM STAT creates a set of HTML files, and the user can then open the index.html file in a browser to view the results. Also, the following CSV files are created and placed in the Data_files directory:

- ▶ <panel_name>_data_movement.csv
- ▶ <panel_name>_skew_curve.csv
- ▶ <panel_name>_workload_ctg.csv

These files can be used as input data for other utilities, such as the IBM STAT Charting Utility.

For more information about how to interpret IBM STAT tool output and CSV files analysis, see *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines, SG24-7521*.

9.2 Thin-provisioned volumes

In a shared storage environment, *thin provisioning* is a method for optimizing the usage of available storage. It relies on the allocation of capacity on demand instead of the traditional method of allocating all the capacity at the time of initial provisioning. Using this principle means that storage environments can achieve higher utilization of physical storage resources by eliminating the unused allocated capacity.

Traditional storage allocation methods often provision large amounts of storage to individual hosts, but some of it remains unused (not written to), which might result in poor usage rates (often as low as 10%) of the underlying physical storage resources. Thin provisioning avoids this issue by presenting more storage capacity to the hosts than it uses from the storage pool. Physical storage resources can be expanded over time to respond to growth.

9.2.1 Concepts

The system supports thin-provisioned volumes in standard pools and in DRPs.

Each volume has a *provisioned capacity* and a *real capacity*. Provisioned capacity is the volume storage capacity that is available to a host. It is the capacity that is detected by host operating systems (OSs) and applications and can be used when creating a file system. Real capacity is the storage capacity that is reserved to a volume copy from a pool.

In a standard-provisioned volume, the provisioned capacity and real capacity are the same. However, in a thin-provisioned volume, the provisioned capacity can be much larger than the real capacity.

The provisioned capacity of a thin-provisioned volume is larger than its real capacity. As more information is written by the host to the volume, more of the real capacity is used. The system identifies read operations to unwritten parts of the provisioned capacity and returns zeros to the server without using any real capacity.

The autoexpand feature prevents a thin-provisioned volume from using up its capacity and going offline. As a thin-provisioned volume uses capacity, the autoexpand feature maintains a fixed amount of unused real capacity that is called the *contingency capacity*. For thin-provisioned volumes in standard pools, the autoexpand feature can be turned on and off. For thin-provisioned volumes in DRPs, the autoexpand feature is always enabled.

The capacity of a thin-provisioned volume is split into chunks that are called *grains*. Write I/O to grains that have not previously been written to causes real capacity to be used to store data and metadata. The grain size of thin-provisioned volumes in standard pools can be 32 KB, 64 KB, 128 KB, or 256 KB. Generally, smaller grain sizes save space but require more metadata access, which can adversely impact performance. When you use thin provisioning with IBM FlashCopy, specify the same grain size for the thin-provisioned volume and FlashCopy. The grain size of thin-provisioned volumes in DRPs cannot be changed from the default of 8 KB.

A thin-provisioned volume can be converted non-disruptively to a fully allocated volume or vice versa by using the volume mirroring function. For example, you can add a thin-provisioned copy to a fully allocated volume and then remove the fully allocated copy from the volume after it is synchronized.

The fully allocated to thin-provisioned migration procedure uses a zero-detection algorithm so that grains that contain all zeros do not cause any real capacity to be used. Usually, if the system is to detect zeros on the volume, you must use software on the host side to write zeros to all unused space on the disk or file system.

9.2.2 Implementation

For more information about creating thin-provisioned volumes, see Chapter 6, “Volumes” on page 299.

Metadata

In a standard pool, the system uses real capacity to store data that is written to the volume and metadata that describes the thin-provisioned configuration of the volume. The metadata that is required for a thin-provisioned volume is usually less than 0.1% of the provisioned capacity.

If the host uses 100% of the provisioned capacity, some extra space is required on your storage pool to store thin-provisioned metadata. In the worst case, the real size of a thin-provisioned volume can be 100.1% of its virtual capacity.

In a DRP, metadata for a thin-provisioned volume is stored separately from user data and not reflected in the volume's real capacity. Capacity reporting is handled at the pool level.

Volume parameters

When creating a thin-provisioned volume in a standard pool, some of its parameters can be modified in Custom mode, as shown in Figure 9-15.

Real capacity defines both initial volume real capacity and the amount of contingency capacity. When autoexpand is enabled, the system tries to maintain the contingency capacity always by allocating extra real capacity when hosts write to the volume.

The warning threshold can be used to send a notification when the volume is about to run out of space.

The screenshot shows a configuration window titled "Thin Provisioning". It contains the following settings:

- Real capacity:** A text input field containing the value "2" and a dropdown menu set to "% of Provisioned capacity".
- Automatically expand:** A checkbox that is checked, with the label "Enabled".
- Warning threshold:** A checkbox that is checked, with the label "Enabled", and a text input field containing the value "80" followed by the label "% of Provisioned capacity".
- Thin-Provisioned Grain Size:** A text input field containing the value "256" and a dropdown menu set to "KIB".

Figure 9-15 Volume parameters for thin provisioning

In a DRP, fine-tuning of these parameters is not required. The real capacity and warning threshold are handled at the pool level. The grain size is always 8 KB, and autoexpand is always on.

Host considerations: Do not use defragmentation applications on thin-provisioned volumes. The defragmentation process can write data to different areas of a volume, which can cause a thin-provisioned volume to grow up to its provisioned size.

9.3 UNMAP

IBM Spectrum Virtualize systems running Version 8.1.0 and later support the Small Computer System Interface (SCSI) **UNMAP** command. This command enables hosts to notify the storage controller of capacity that is no longer required, which can improve capacity savings and performance of flash storage.

9.3.1 The SCSI UNMAP command

UNMAP is a set of SCSI primitives that enable hosts to indicate to a storage system that space that is allocated to a range of blocks on a storage volume is no longer required. This command enables the storage system to take measures and optimize the system so that the space can be reused for other purposes.

When a host writes to a volume, storage is allocated from the storage pool. To free allocated space back to the pool, human intervention is needed on the storage system. The SCSI **UNMAP** feature is used to allow host OSs to unprovision storage on the storage system, which means that the resources can automatically be freed in the storage pools and used for other purposes.

One of the most common use cases is a host application, such as VMware, freeing storage within a file system. Then, the storage system can reorganize the space, such as optimizing the data on the volume or the pool so that space can be reclaimed.

A SCSI unmappable volume is a volume that can have storage unprovision and space reclamation that is triggered by the host OS. The system can pass the SCSI **UNMAP** command through to back-end flash storage and external storage controllers that support the function.

9.3.2 Back-end SCSI UNMAP

The system can generate and send SCSI **UNMAP** commands to specific back-end storage controllers and internal flash storage. Support for SCSI **UNMAP** was introduced with Version 8.1.1.

This process occurs when volumes are formatted, deleted, extents are migrated, or an **UNMAP** command is received from the host. SCSI **UNMAP** commands are sent only to the following back-end controllers:

- ▶ Beginning with Version 8.1.1: IBM FlashSystem A9000, IBM Storwize, and IBM FlashSystem family products (excluding IBM FlashSystem 840 and IBM FlashSystem 900 AE2) and IBM PureSystems® storage systems
- ▶ Beginning with 8.3.0.1: HPE Nimble storage systems

Back-end SCSI **UNMAP** commands help prevent an overprovisioned storage controller from running out of free capacity for write I/O requests, which means that when you use supported overprovisioned back-end storage, back-end SCSI **UNMAP** should be enabled.

Flash storage typically requires empty blocks to serve write I/O requests, which means **UNMAP** can improve flash performance by erasing blocks in advance.

This feature is turned on by default. It is a best practice to keep back-end **UNMAP** enabled, especially if a system is virtualizing an overprovisioned storage controller or uses FCM drives.

To verify that sending **UNMAP** commands to a back end is enabled, run the `lssystem` command, as shown in Example 9-7.

Example 9-7 Verifying the back-end UNMAP support status

```
IBM FlashSystem 7200:ITS0FS7K:superuser>lssystem | grep backend_unmap
backend_unmap on
```

9.3.3 Host SCSI UNMAP

The IBM Spectrum Virtualize system can advertise support for SCSI **UNMAP** to hosts. Hosts can then use the set of SCSI **UNMAP** commands to indicate that formerly used capacity is no longer required on a volume.

When these volumes are in DRPs, that capacity becomes reclaimable capacity and is monitored and collected, and eventually redistributed back to the pool for use by the system. Volumes in standard pools do not support automatic space reclamation after data is unmapped, and SCSI **UNMAP** commands are handled as though they were writes with zero data.

The system also sends SCSI **UNMAP** commands to back-end controllers that support them if host unmaps for corresponding blocks are received (and backend **UNMAP** is enabled).

With host SCSI **UNMAP** enabled, some host types (for example, Windows, Linux, or VMware) change their behavior when creating a file system on a volume, issuing SCSI **UNMAP** commands to the whole capacity of the volume. The format completes only after all of these **UNMAP** commands complete. Some host types run a background process (for example, **fstrim** on Linux), which periodically issues SCSI **UNMAP** commands for regions of a file system that are no longer required. Hosts might also send **UNMAP** commands when files are deleted in a file system.

Host SCSI **UNMAP** commands drive more I/O workload to back-end storage. In some circumstances (for example, volumes on a heavily loaded NL-serial-attached SCSI (SAS) array), this situation can cause an increase in response times on volumes that use the same storage. Also, host formatting time is likely to increase compared to a system that does not support the SCSI **UNMAP** command.

If you use DRPs, an overprovisioned back end that supports **UNMAP**, or FCM drives, it is a best practice to turn on SCSI **UNMAP** support. Host **UNMAP** support is enabled by default.

If only standard pools are configured and the back end is traditional (fully provisioned), consider keeping host **UNMAP** support turned off because it does not provide any benefit.

To check and modify the current settings for host SCSI **UNMAP** support, run the **lssystem** and **chsystem** CLI commands, as shown in Example 9-8.

Example 9-8 Turning on host UNMAP support

```
IBM FlashSystem 7200:ITS0FS7K:superuser>lssystem | grep host_unmap
host_unmap off
IBM FlashSystem 7200:ITS0FS7K:superuser>chsystem -hostunmap on
IBM FlashSystem 7200:ITS0FS7K:superuser>
```

Note: You can switch host **UNMAP** support on and off nondisruptively on the system side. However, hosts might need to rediscover storage, or (in the worst case) be restarted for them to stop sending **UNMAP** commands.

9.3.4 Offloading I/O throttle

Throttles are a mechanism to control the amount of resources that are used when the system is processing I/Os on supported objects. If a throttle limit is defined, the system processes the I/O for that object or delays the processing of the I/O to free resources for more critical I/O operations.

Offload commands, such as **UNMAP** and **XCOPY**, free hosts and speed the copy process by offloading the operations of certain types of hosts to a storage system. These commands are used by hosts to format new file systems, or copy volumes without the host needing to read and then write data.

Offload commands can sometimes create I/O-intensive workloads, potentially taking bandwidth from production volumes and affecting performance, especially if the underlying storage cannot handle the amount of I/O that is generated.

Throttles can be used to delay processing for offloads to free bandwidth for other more critical operations, which can improve performance but limits the rate at which host features, such as VMware VMotion, can copy data. It can also increase the time that it takes to format file systems on a host.

Note: For systems that are managing any NL storage, it might be a best practice to set the offload throttle to 100 MBps.

To implement an offload throttle, run the `mkthrottle` command with the `-type offload` parameter. In the GUI, select **Monitoring** → **Systems Hardware**, and then click **System Actions** → **Edit System Offload Throttle**, as shown in Figure 9-16.

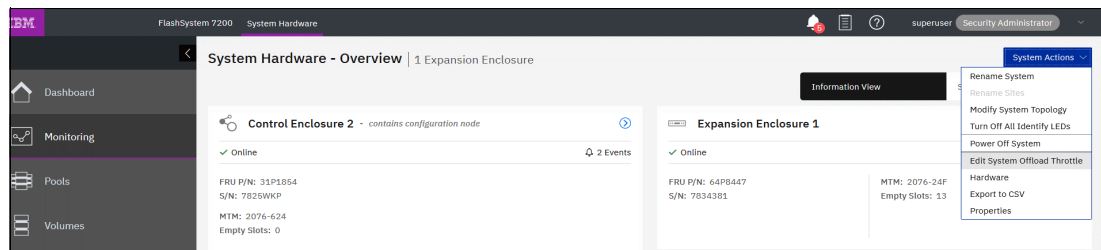


Figure 9-16 Setting an offload throttle

9.4 Data Reduction Pools

DRPs provide a set of techniques that can be used to reduce the amount of usable capacity that is required to store data, which helps increase storage efficiency and reduce storage costs. Available techniques include thin provisioning, compression, and deduplication.

DRPs automatically reclaim used capacity that is no longer needed by host systems and return it back to the pool as available capacity for future reuse.

The data reduction in DRPs is embedded in this pool type and no separate license is necessary. This situation does not apply to real-time compression (RtC), where a specific capacity-based license is needed.

Note: This book provides only an overview of DRP. For more information, see *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430.

9.4.1 Introduction to DRP

The system can use different data reduction methods simultaneously, which increases the capacity savings across the entire storage pool.

DRPs support five types of volumes:

- ▶ Fully allocated
This type provides no data reduction.
- ▶ Thin provisioned
This type provides data reduction by allocating storage on demand when writing to the volume.
- ▶ Thin and compressed
In addition to being thin provisioned, data is compressed before being written to storage.
- ▶ Thin and deduplicated
In addition to being thin provisioned, duplicates of data blocks are detected and replaced with references to the first copy.
- ▶ Thin, compressed, and deduplicated
This type achieves maximum data reduction by combining thin provisioning, compression, and deduplication.

Volumes in a DRP track when capacity is freed from hosts and possible unused capacity that can be collected and reused within the storage pool. When a host no longer needs the data that is stored on a volume, the host system uses SCSI **UNMAP** commands to release that capacity from the volume. When these volumes are in DRPs, that capacity becomes reclaimable capacity, and is monitored, collected, and eventually redistributed back to the pool for use by the system.

Note: If the usable capacity usage of a DRP exceeds more than 85%, I/O performance can be affected. The system needs 15% of usable capacity available in DRPs to ensure that capacity reclamation can be performed efficiently.

At its core, a DRP uses a Log Structured Array (LSA) to allocate capacity. An LSA enables a tree-like directory to define the physical placement of data blocks independent of size and logical location.

Each volume has a range of logical block addresses (LBAs), starting from 0 and ending with the block address that fills the capacity. The LSA enables the system to allocate data sequentially when written to volumes (in any order) and provides a directory that provides a lookup to match volume LBA with physical addresses within the array. A volume in a DRP contains directory metadata to store the mapping from logical address on the volume to physical location on the back-end storage.

This directory is too large to store in memory, so it must be read from storage as required. The lookup and maintenance of this metadata results in I/O amplification. I/O amplification occurs when a single host-generated read or write I/O results in more than one back-end storage I/O request. For example, a read I/O request might need to read some directory metadata in addition to the actual data. A write I/O request might need to read directory metadata write updated directory metadata, journal metadata, and the actual data.

Conversely, data reduction reduces the size of data that uses compression and deduplication, so less data is written to the back-end storage.

IBM Spectrum Virtualize V8.4 introduces *child pools* for DRPs. A child pool is a folder-like object within a parent DRP that contains volumes. The child pool for DRP is quota-less and its capacity is the sum of all volumes within the child pool. A child pool can be assigned to an ownership group and to segment administrative domains. The parent pool and associated child pools share MDisk, deduplication hash table, and encryption keys. Therefore, it seems advisable to use this technology to separate departments of a single client, but not different clients.

At the time of writing, VMware vSphere Virtual Volumes (VVOLs) are not supported by child pools in DRPs.

Due to the nature of the newly introduced child pools, there is a new type of volume migration that is available to move volumes within a single DRP and its affiliated child pools. With this migration, you can move volumes between all pools within one DRP entity.

9.4.2 DRP benefits

DRPs are a new type of storage pool that implement techniques such as thin provisioning, compression, and deduplication to reduce the amount of physical capacity that is required to store data. Savings in storage capacity requirements translate into reduction in the cost of storing the data.

The cost reductions that are achieved through software can facilitate the transition to all flash storage. Flash storage has lower operating costs, lower power consumption, higher density, and is cheaper to cool than disk storage. However, the cost of flash storage is still higher. Data reduction can reduce the total cost of ownership (TCO) of an all-flash system to be competitive with HDDs.

One benefit of DRP is in the form of capacity savings that are achieved by deduplication and compression. Real-time deduplication identifies duplicate data blocks during write I/O operations and stores a reference to the first copy of the data instead of writing the data to the storage pool a second time. It does this task by maintaining a fingerprint database containing hashes of data blocks already written to the pool. If new data that is written by hosts matches an entry in this database, then a reference is generated in the directory metadata instead of writing the new data.

Compression reduces the size of the host data that is written to the storage pool. DRP uses the Lempel-Ziv based RtC and decompression algorithm. It offers a new implementation of data compression that is fully integrated into the IBM Spectrum Virtualize I/O stack. It makes optimal use of node resources such as memory and CPU cores, and uses hardware acceleration on supported platforms efficiently. DRP compression operates on small block sizes, which results in consistent and predictable performance.

Deduplication and compression can be combined, in which case data is first deduplicated and then compressed. Therefore, deduplication references are created on the compressed data that is stored on the physical domain.

DRP supports end-to-end SCSI **UNMAP** functions. Hosts use the set of SCSI **UNMAP** commands to indicate that the formerly used capacity is no longer required on a target volume. Reclaimable capacity is unused capacity that is created when data is overwritten, volumes are deleted, or when data is marked as unneeded by a host by using the SCSI **UNMAP** command. That capacity can be collected and reused on the system.

DRPs, the directory, and the actual reduction techniques are designed around optimizing for flash and future solid-state storage technologies. All metadata operations are 4 KB, which is ideal for flash storage to maintain low and consistent latency. All data read operations are 8 KB (before reduction) and designed to minimize latency because flash storage is suitable for small block workload with high IOPS. All write operations are coalesced into 256 KB sequential writes to simplify the garbage collection on flash devices and gain full stride writes from RAID arrays.

DRP works well with Easy Tier. The directory metadata of DRPs does not fit in memory, so it is stored on disk by using dedicated metadata volumes that are separate from the actual data. The metadata volumes are small but frequently accessed by small block I/O requests. Performance gains are expected because they are optimal candidates for promotion to the fastest tier of storage through Easy Tier. In contrast, data volumes with large but frequently rewritten data is grouped to consolidate “heat”. Easy Tier can accurately identify active data.

RAID Reconstruct Read (3R) is a technology to increase the reliability and availability of data that is stored in DRPs. 3R is introduced in IBM Spectrum Virtualize V8.4.

All reads are evaluated, and if there is a mismatch, the data is reconstructed by using the parity information. To eliminate rereading of corrupted data, the affiliate cache block is marked invalid. This process works for internal and external back-end devices.

9.4.3 Planning for DRP

Before configuring and using DRPs in production environments, you must plan the capacity and performance. DRP has different performance characteristics than standard pools, so existing sizing models cannot be used directly without modifications.

For more information about how to estimate the capacity savings that are achieved by compression and deduplication, see 9.5, “Saving estimations for compression and deduplication” on page 545.

The following software and hardware requirements must be met for DRP compression and deduplication:

- ▶ The system must run Version 8.1.3.2 or higher.
- ▶ IBM FlashSystem 5010 is not supported.
- ▶ IBM FlashSystem 5030 needs the Cache Upgrade option (#ALGA).
- ▶ All other supported platforms need at least 32 GB of cache.

In most cases, it is a best practice to enable compression for all thin-provisioned and deduplicated volumes. Overhead in DRPs is caused by metadata handling, which is the same for compressed volumes and thin-provisioned volumes without compression.

In the IBM FlashSystem 5030 system, the limitation in CPU power and the lack of a hardware accelerator might lead to a performance impact.

If the system contains self-compressing drives, DRPs provide a major benefit only if deduplication is used and the estimated deduplication savings are significant. If there is no plan to use deduplication or the expected deduplication ratio is low, consider using fully allocated volumes instead and use drive compression for capacity savings. For more information about how to estimate deduplication savings, see 9.5.2, “Evaluating compression and deduplication” on page 547.

In systems with self-compressing drives, certain system configurations make determining accurate physical capacity on the system difficult. If the system contains self-compressing drives and DRPs with thin-provisioned volumes without compression, the system cannot determine the accurate amount of physical capacity that is used on the system. In this case, overcommitting and losing access to write operations is possible. To prevent this situation from happening, use compressed volumes (with or without deduplication) or fully allocated volumes. Separate compressed volumes and fully allocated volumes by using separate pools. Similar considerations apply to configurations with compressing back-end storage controllers, as described in 9.6, “Overprovisioning and data reduction on external storage” on page 548.

There is a maximum number of four DRPs in a system. When this limit is reached, only more standard pools can be created.

A DRP uses a customer data volume per I/O group to store volume data. There is a limit on the maximum size of a customer data volume of 128,000 extents per I/O group, which places a limit on the maximum physical capacity in a pool after data reduction that depends on the extent size, number of DRPs, and number of I/O groups, as shown in Table 9-2. DRPs have a minimum extent size of 1024 MB.

Table 9-2 Maximum physical capacity after data reduction

Extent size	1 DRP - 1 I/O group	1 DRP - 4 I/O groups	4 DRP - 4 I/O groups
1024 MB	128 TiB	512 TiB	2 PiB
2048 MB	256 TiB	1 PiB	4 PiB
4096 MB	512 TiB	2 PiB	8 PiB
8192 MB	1 PiB	4 PiB	16 PiB

Overwriting data, unmapping data, and deleting volumes cause reclaimable capacity in the pool to increase. Garbage collection is performed in the background to convert reclaimable capacity to available capacity. This operation requires free capacity in the pool to operate efficiently without impacting I/O performance. A best practice is to ensure that the provisioned capacity with the DRP does not exceed 85% of the total usable capacity of the DRP.

To ensure that garbage collection is working properly, there is minimum capacity limit in a single DRP depending on extent size and number of I/O groups, as shown in Table 9-3. Even when there are no volumes in the pool, some of the space is used to store metadata. The required metadata capacity depends on the total capacity of the storage pool and on the extent size, which should be considered when planning capacity.

Table 9-3 Minimum capacity in a single Data Reduction Pool

Extent size	1 I/O group	4 I/O groups
1024 MB	255 GiB	1 TiB
2048 MB	0.5 TiB	2 TiB
4096 MB	1 TiB	4 TiB
8192 MB	2 TiB	8 TiB

Note: The default extent size in a DRP is 4 GB. If the estimated total capacity in the pool exceeds the documented limits, choose a larger extent size. If the estimated total capacity is relatively small, consider using a smaller extent size for a smaller metadata impact and lower minimum capacity limit.

For more information about the considerations of using data reduction on the system and the back-end storage, see 9.6, “Overprovisioning and data reduction on external storage” on page 548.

9.4.4 Implementing DRP with compression and deduplication

To use all data reduction technologies on the system, you must create a DRP, create volumes within the DRP, and map these volumes to hosts that support SCSI **UNMAP** commands. The implementation process for DRP is like standard pools, but has its own specifics.

Creating pools and volumes

To create a DRP, select **Pools** → **Pools**, and select **Data Reduction** in the Create Pool dialog. For more information about how to create a storage pool and populate it with MDisks, see Chapter 5, “Storage pools” on page 237.

To create a volume within a DRP, select **Volumes** → **Volumes**, and click **Create Volumes**.

Figure 9-17 on page 539 shows the Create Volumes dialog. In the **Capacity Savings** menu, the following selections are available: **None**, **thin provisioned**, and **Compressed**. If **Compressed** or **thin provisioned** is selected, the **Deduplicated** option also becomes available and can be selected.

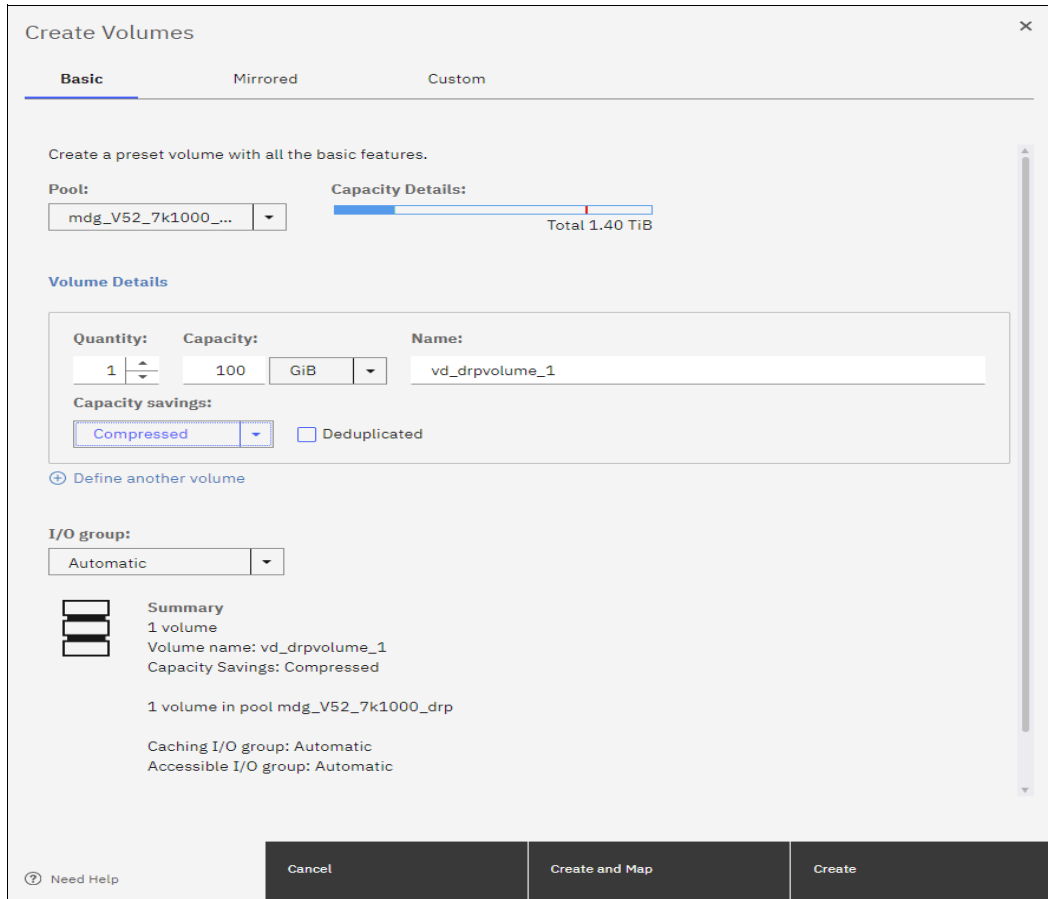


Figure 9-17 Creating a compressed volume

Capacity monitoring

Capacity monitoring in DRPs is mainly done on the system and storage pool levels. Use the Dashboard in the GUI to view a summary of the capacity usage and capacity savings of the entire system.

The Pools page in the management GUI is used for reporting on the storage pool level and displays Usable Capacity and Capacity Details. Usable Capacity indicates the amount of capacity that is available for storing data on a pool after formatting and RAID techniques are applied. Capacity Details is the capacity that is available for volumes before any capacity savings methods are applied. To monitor this capacity, select **Pools** → **Pools**, as shown in Figure 9-18.


















Name	State	Usable Capacity	Capacity Details	Data Reduction
> mdg_fs820-1_flash	✓ Online	 920.50 GiB / 8.00 TiB (11%)	 1,020.50 GiB / 8.00 TiB (12%)	No
mdg_quorum	✓ Online	 0 bytes / 3.50 GiB (0%)	 0 bytes / 3.50 GiB (0%)	No
> mdg_v51_10k600_hybrid	✓ Online	 5.95 TiB / 6.75 TiB (88%)	 6.11 TiB / 6.75 TiB (91%)	No
▼ mdg_v52_7k1000	✓ Online	 4.47 TiB / 7.02 TiB (64%)	 4.64 TiB / 7.02 TiB (66%)	No
childpool_mirror_mdg_v52_7k1000	✓ Online	 31.75 GiB / 100.00 GiB (32%)		No
childpool_v52	✓ Online	 0 bytes / 100.00 GiB (0%)		No
MigrationPool_256	✓ Online	 0 bytes		No
rl_pool_v51_10k600_hybrid	✓ Online	 711.00 GiB / 1.90 TiB (36%)	 711.00 GiB / 1.90 TiB (36%)	No
rl_pool_v52_7k1000	✓ Online	 809.25 GiB / 1.41 TiB (56%)	 809.25 GiB / 1.41 TiB (56%)	No
mdg_v52_7k1000_drp	✓ Online	 17.00 GiB / 1.40 TiB (1%)	 267.00 GiB / 1.40 TiB (19%)	Yes

Figure 9-18 Data Reduction Pool capacity overview

To see more detailed capacity reporting including the warning threshold and capacity savings, open the pool properties dialog by right-clicking a pool and selecting **Properties**. This dialog shows the savings that are achieved by thin provisioning, compression, and deduplication, and the total data reduction savings in the pool, as shown in Figure 9-19 on page 541. In addition, the Reclaimable capacity is shown, which is unused capacity that is created when data is overwritten, volumes are deleted, or when data is marked as unneeded by a host by using the SCSI **UNMAP** command. This capacity is converted to available capacity by the garbage collection background process.

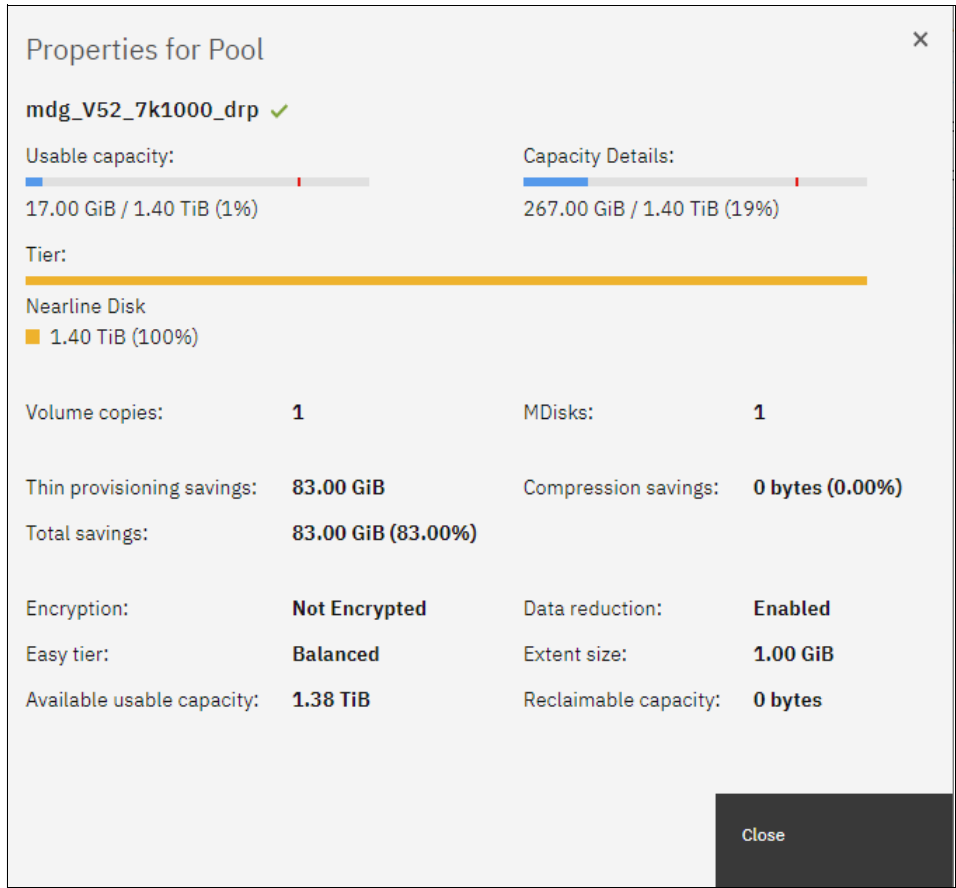


Figure 9-19 Capacity reporting in a Data Reduction Pool

The capacity reporting shows 267 GiB capacity usage, and there is only a single virtual disk (VDisk) copy with 100 GiB provisioned capacity. This capacity is the result of a DRP reservation for deduplication and compression. Some extents are marked used, and only some bytes are written. With increasing usage of the pool, the impact of this reservation is decreasing.

Thin-provisioned, compressed, and deduplicated volumes do not provide detailed per-volume capacity reporting, as shown in the Volumes by Pool window in Figure 9-20. Only the Capacity (which is the provisioned capacity that is available to hosts) is shown. Real capacity, Used capacity, and Compression savings are not applicable for volumes with capacity savings. Only fully allocated volumes display those parameters.



Figure 9-20 Volumes in a DRP pool

Per-volume compression savings are not visible directly, but they can be accurately estimated by using the IBM Comprestimator, which is described in 9.5.1, “Evaluating compression savings by using IBM Comprestimator” on page 545. The IBM Comprestimator can be used on compressed volumes to analyze the volume level compression savings.

The CLI can be used for limited capacity reporting on the volume level. The `used_capacity_before_reduction` entry indicates the total amount of data that is written to a thin-provisioned or compressed volume copy in a data reduction storage pool before data reduction occurs. This field is empty for fully allocated volume copies and volume copies not in a DRP.

To find this value, run the `lsvdisk` command with a volume name or ID as a parameter, as shown in Example 9-9. It shows a thin-provisioned volume without compression and deduplication with a virtual size of 1 TiB that is provisioned to the host. A 53 GB file was written from the host.

Example 9-9 Data Reduction Pool volume capacity reporting on the CLI

```
IBM FlashSystem 7200:ITS0FS7K:superuser>lsvdisk thin_provisioned
id 34
name vdisk1

capacity 1.00TB

used_capacity
real_capacity
free_capacity

tier tier_scm
tier_capacity 0.00MB
tier tier0_flash
tier_capacity 0.00MB
tier tier1_flash
tier_capacity 0.00MB
tier tier_enterprise
tier_capacity 0.00MB
tier tier_nearline
tier_capacity 0.00MB

compressed_copy no
uncompressed_used_capacity
deduplicated_copy no
used_capacity_before_reduction 53.04GB
```

The used, real, and free capacity, and the capacity that is stored on each storage tier, is not shown for volumes (except fully allocated volumes) in DRPs.

Capacity reporting on the pool level is available by running the `lsmdiskgrp` command with the pool ID or name as a parameter, as shown in Example 9-10.

Example 9-10 Data Reduction Pool capacity reporting on the CLI

```
IBM FlashSystem 7200:ITS0FS7K:superuser>lsmdiskgrp 1 | grep -E
"capacity|compression|tier tier"
capacity 5.00TB
free_capacity 2.87TB
virtual_capacity 4.00TB
```

```
used_capacity 1.14TB
real_capacity 1.14TB
tier tier_scm
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier tier0_flash
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier tier1_flash
tier_capacity 5.00TB
tier_free_capacity 3.85TB
tier tier_enterprise
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier tier_nearline
tier_capacity 0.00MB
tier_free_capacity 0.00MB
compression_active no
compression_virtual_capacity 0.00MB
compression_compressed_capacity 0.00MB
compression_uncompressed_capacity 0.00MB
child_mdisk_grp_capacity 0.00MB
used_capacity_before_reduction 143.68GB
used_capacity_after_reduction 94.64GB
overhead_capacity 52.00GB
deduplication_capacity_saving 36.20GB
reclaimable_capacity 0.00MB
physical_capacity 5.00TB
physical_free_capacity 3.85TB
```

Compression-related properties are not valid for DRPs.

For more information about every reported value, see [IBM FlashSystem 9200 documentation](#) and expand **Command-line interface** → **Storage pool commands** → **lsmdiskgrp**.

Migrating to and from a DRP

Data migration from or to a DRP is done by using volume mirroring. A second copy in the target pool is added to the source volume, and the original copy is optionally removed after the synchronization process completes. If the volume already has two copies, one of the copies must be removed or a more complex migration scheme that uses FlashCopy, RC, host mirroring, or similar must be used.

Also, real-time compressed volumes with data-reduced DRP volumes cannot co-exist in a single I/O group. Therefore, migrating such volumes has extra considerations. One possible solution might be to inflate real-time compressed volumes in standard pools and migrate these volumes in a second step to data-reduced volumes in a DRP pool.

Note: All volumes that cannot coexist with data-reduced DRP volumes must be migrated in a single step.

Depending on the system configuration and the type of migration, a one-step migration or a two-step migration is necessary. The reason is that compressed volumes in standard pools cannot coexist with deduplicated volumes in DRPs. Therefore, a two-step migration is required in the following scenarios.

The migration processes work as follows:

1. To create a second copy, right-click the source volume and choose **Add Volume Copy**, as shown in Figure 9-21. Choose the target pool of the migration for the second copy and select the capacity savings.

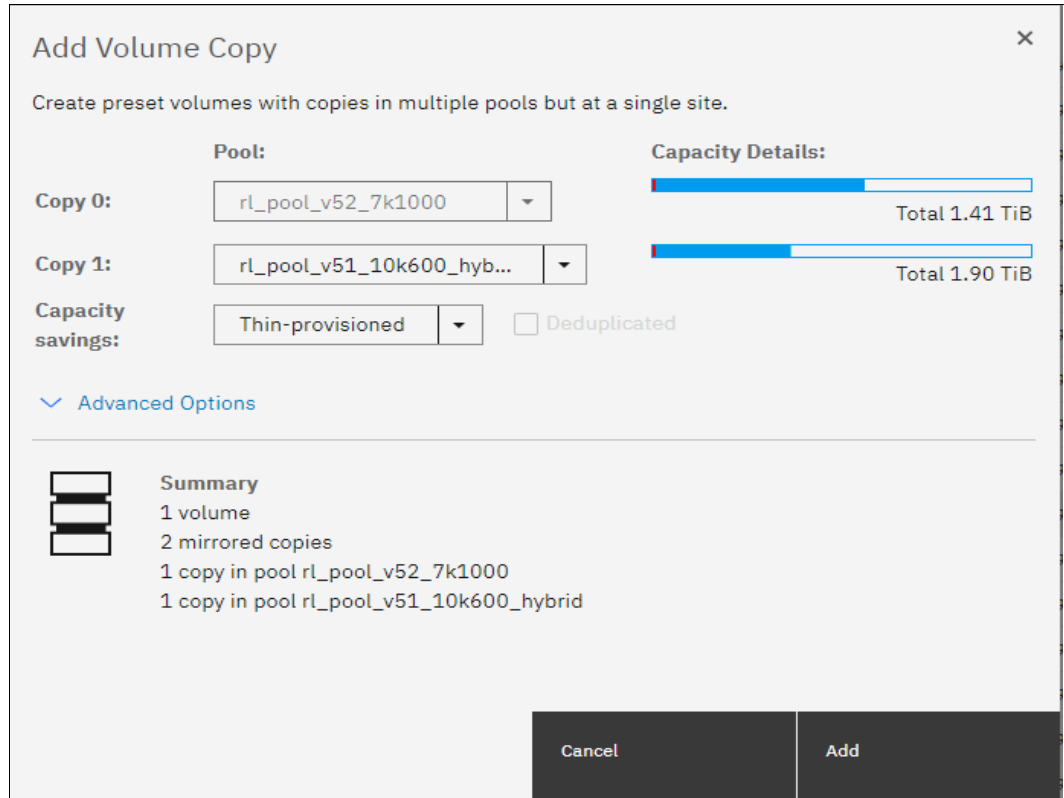


Figure 9-21 Add Volume Copy dialog

2. After you click **Add**, synchronization starts. The time that synchronization takes to complete depends on the size of the volume, system performance, and the configured migration rate. You can increase the synchronization rate by right-clicking the volume and selecting **Modify Mirror Sync Rate**.

When both copies are synchronized, *Yes* is displayed for both copies in the Synchronized column in the Volumes window. You can track the synchronization process by using the Running tasks window, as shown in Figure 9-22. After it reaches 100% and the copies are in-sync, you can complete migration by deleting the source copy.

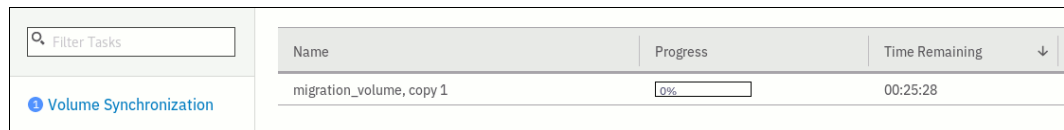


Figure 9-22 Synchronization progress

Garbage collection and volume deletion

DRP includes built-in capabilities to enable garbage collection of unused storage capacity. *Garbage collection* is a DRP process that reduces the amount of data that is stored on external storage systems and internal drives by reclaiming previously used storage resources that are no longer needed by host systems.

When a DRP is created, the system monitors the pool for reclaimable capacity from host **UNMAP** operations. When space is freed from a host OS, it is a process called *unmapping*. Hosts indicate that the allocated capacity is no longer required on a target volume. The freed space is collected and reused by the system automatically without having to reallocate the capacity manually.

Removing thin-provisioned or compressed volume copies in a DRP is an asynchronous operation. Volume copies that are removed from the system enter the *deleting* state, during which the used capacity of the copies is converted to reclaimable capacity in the pool by using a background deletion process. The removal process of deduplicated volume copies searches and moves deduplication references that other volumes might have to the deleting volume copies. This task is done to ensure that deduplicated data that was on deleted copies continues to be available for other volumes in the system.

After this process completes, the volume copies are deleted and disappear from the system configuration. In a second step, garbage collection can give the reclaimable capacity that is generated in the first step back to the pool as available capacity, which means that the used capacity of a removed volume is not available for reuse immediately after the removal.

The time that it takes to delete a thin-provisioned or compressed volume copy depends on the size of the volume, the system configuration, and the workload. For deduplicated copies, the duration also depends on the amount and size of other deduplicated copies in the pool, which means that it might take a long time to delete a small deduplicated copy if there are many other deduplicated volumes in the same pool. The deletion process is a background process that might impact system overall performance.

The deleting state of a volume or volume copy can be seen by running the `lsvdisk` command. The GUI hides volumes in this state, but it shows deleting volume copies if the volume contains another copy.

Note: Removing thin-provisioned or compressed volume copies in a DRP might take a long time to complete. Used capacity is not immediately given back to the pool as available capacity.

When one copy of a mirrored volume is in the deleting state, it is not possible to add a copy to the volume before the deletion finishes. If a new copy must be added without waiting for the deletion to complete, first split the copy that must be deleted into a new volume, and then delete the new volume and add a new second copy to the original volume. To split a copy into a new volume, right-click the copy and select **Split into New Volume** in the GUI or run the `splitvdiskcopy` command on the CLI.

9.5 Saving estimations for compression and deduplication

This section provides information about the tools that are used for sizing the environment for compression and deduplication.

9.5.1 Evaluating compression savings by using IBM Comprestimator

IBM Comprestimator is a utility that estimates the capacity savings that can be achieved when compression is used for storage volumes. The utility is integrated into the system by using the GUI and the CLI. It can also be installed and used on host systems.

Starting with IBM Spectrum Virtualize V8.4, the integrated Comprestimator is always enabled and running continuously, thus providing up-to-date compression estimation over the entire cluster, both in GUI and IBM Storage Insights.

IBM Comprestimator provides a quick and accurate estimation of compression and thin-provisioning benefits. The utility performs read-only operations, so it does not affect the data that is stored on the volume.

If the compression savings prove to be beneficial in your environment, volume mirroring can be used to convert volumes to compressed volumes.

To see the results and the date of the latest estimation cycle, as shown in Figure 9-23, Go to the **Volumes** window, right-click any volume, and select **Space Savings** → **Estimate Compression Savings**. If no analysis was done, the system suggests running it. A new estimation of all volumes can be started from this dialog. To run or rerun analysis on a single volume, select **Analyze** in the **Space Savings** submenu.

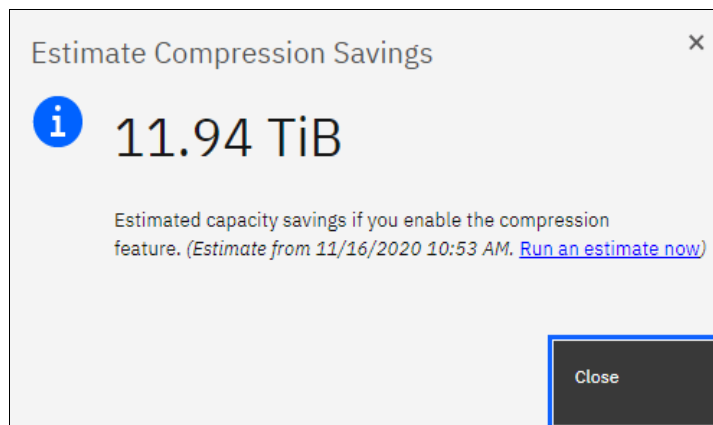


Figure 9-23 Estimate Compression Savings

To analyze all the volumes on the system from the CLI, run the **analyzevdiskbysystem** command.

The command analyzes all the current volumes that are created on the system. Volumes that are created during or after the analysis are not included and can be analyzed individually. The time that it takes to analyze all the volumes on system depends on the number of volumes that are being analyzed, and results can be expected at about a minute per volume. For example, if a system has 50 volumes, compression savings analysis takes approximately 50 minutes.

You can run an analysis on a single volume by specifying its name or ID as a parameter for the **analyzevdisk** CLI command.

To check the progress of the analysis, run the **lsvdiskanalysisprogress** command. This command displays the total number of volumes on the system, total number of volumes that are remaining to be analyzed, and estimated time of completion.

To display information for the thin provisioning and compression estimation analysis report for all volumes, run the **lsvdiskanalysis** command.

If you are using a version of IBM Spectrum Virtualize that is older than Version 7.6 or if you want to estimate the compression savings of another IBM or non-IBM storage system, the separate IBM Comprestimator Utility can be installed on a host that is connected to the device that needs to be analyzed. For more information and the latest version of this utility, see [IBM Comprestimator Utility Version 1.5.3.1](#).

Consider the following best practices for using IBM Comprestimator:

- ▶ Run the IBM Comprestimator Utility before implementing an IBM Spectrum Virtualize solution and DRPs.
- ▶ Download the latest version of the IBM Comprestimator Utility if you are not using one that is included in your IBM Spectrum Virtualize solution.
- ▶ Use IBM Comprestimator to analyze volumes that contain as much active data as possible rather than volumes that are nearly empty or newly created to ensure more accuracy when sizing your environment for compression and DRPs.

Note: IBM Comprestimator can run for a long period (a few hours) when it is scanning a relatively empty device. The utility randomly selects and reads 256 KB samples from the device. If the sample is empty (that is, full of null values), it is skipped. A minimum number of samples with data is required to provide an accurate estimation. When a device is mostly empty, many random samples are empty. As a result, the utility runs for a longer time as it tries to gather enough non-empty samples that are required for an accurate estimate. The scan is stopped if the number of empty samples is over 95%.

9.5.2 Evaluating compression and deduplication

To help with the profiling and analysis of user workloads that are to be migrated to the new system, IBM provides a highly accurate Data Reduction Estimation Tool (DRET) that supports both deduplication and compression. The tool operates by scanning target workloads on any legacy array (from IBM or a third party) and then merging all scan results to provide an integrated system-level data reduction estimate. It provides a report of what it would expect the deduplication and compression savings to be from data that is written to an existing disk.

The DRET utility uses advanced mathematical and statistical algorithms to perform an analysis with a low memory footprint. The utility runs on a host that has access to the devices to be analyzed. It performs only read operations, so it has no effect on the data that is stored on the device.

The following sections provide information about installing DRET on a host and using it to analyze devices on it. Depending on the environment configuration, in many cases DRET is used on more than one host to analyze more data types.

When DRET is used to analyze a block device that is used by a file system, all underlying data in the device is analyzed regardless of whether this data belongs to files that were already deleted from the file system. For example, you can fill a 100 GB file system and make it 100% used, and then delete all the files in the file system to make it 0% used. When scanning the block device that is used for storing the file system in this example, DRET accesses the data that belongs to the files that are deleted.

Important: The preferred method of using DRET is to analyze volumes that contain as much active data as possible rather than volumes that are mostly empty of data, which increases the accuracy level and reduces the risk of analyzing old data that is deleted but might still have traces on the device.

For more information and the latest version of this utility, see [Data Reduction Estimator Tool Version 1.04](#).

9.6 Overprovisioning and data reduction on external storage

Starting with IBM Spectrum Virtualize V8.1.x, overprovisioning on selected back-end controllers is supported, which means that if back-end storage performs data deduplication or data compression on LUs that are provisioned from it, they still can be used as external MDisks on the system. However, more configuration and monitoring considerations must be accounted for.

Overprovisioned MDisks from controllers that are supported by this feature can be used as managed mode MDisks in the system and can be added to storage pools (including DRPs).

Implementation steps for overprovisioned MDisks are the same as for fully allocated storage controllers. The system detects whether the MDisk is overprovisioned, its total physical capacity, and used and remaining physical capacity. It detects whether SCSI **UNMAP** commands are supported by the back end. By sending SCSI **UNMAP** commands to overprovisioned MDisks, the system marks data that is no longer in use. Then, the garbage collection processes on the back end can free unused capacity and convert it to free space.

At the time of writing, the following back-end controllers are supported by overprovisioned MDisks:

- ▶ IBM FlashSystem A9000 V12.1.0 and later
- ▶ IBM FlashSystem 900 V1.4 and later
- ▶ IBM FlashSystem V9000 AE2 and AE3 expansions
- ▶ IBM Storwize or IBM FlashSystem family systems Version 8.1.0 and later
- ▶ PureSystems storage
- ▶ HPE Nimble

Extra caution is required when planning and monitoring capacity for such configurations. Table 9-4 shows an overview of configuration guidelines when using overprovisioned external storage controllers.

Table 9-4 Using data reduction at two levels

System	Back end	Comments
DRP	Fully allocated	<i>Recommended.</i> Use DRP on the system to plan for compression and deduplication. DRP at the top level provides the best application capacity reporting.
Fully allocated	Overprovisioned, single tier of storage	<i>Recommended with appropriate precautions.</i> Track physical capacity use carefully to avoid out-of-space conditions. The system can report physical use but does not manage to avoid out-of-space conditions. There is no visibility of each application's use at the system layer. If the back end runs out of space, there is a limited ability to recover. Consider creating a sacrificial emergency space volume.
DRP with compression	Overprovisioned	<i>Recommended with appropriate precautions.</i> Assume 1:1 compression in back-end storage and do not overcommit capacity in the back end. Small extra savings are achieved from compressing DRP metadata.

System	Back end	Comments
Fully allocated	Overprovisioned, multiple tiers of storage	<i>Use with great care.</i> Easy Tier is unaware of physical capacity in tiers of a hybrid pool, so it tends to fill the top tier with the hottest data. Changes in the compressibility of data in the top tier can overcommit the storage, which leads to out-of-space conditions.
DRP with thin-provisioned or fully allocated volumes	Overprovisioned	<i>Avoid.</i> Difficult to understand the physical capacity usage of the uncompressed volumes. High risk of overcommitting the back end. If a mix of DRP and fully allocated volumes is required, use separate pools.
DRP	DRP	<i>Avoid.</i> Creates two levels of I/O amplification and capacity impact. DRP at the bottom layer provides no benefit.

When using DRPs with a compressing back-end controller, use compression in DRP and avoid overcommitting the back end by assuming a 1:1 compression ratio in back-end storage. Small extra savings are realized from compressing metadata.

If the back-end controller uses FCM drives that are always compressing with hardware acceleration, the same methodology should be used. The eventual capacity savings should be used by creating more MDisks to be implemented in the DRP.

Note: Fully allocated volumes that are above overprovisioned MDisk configurations must be avoided or used with extreme caution because it can lead to overcommitting back-end storage.

The concept of provisioning groups is used for capacity reporting and monitoring of overprovisioned external storage controllers. A provisioning group is an object that represents a set of MDisks that share physical resources. Each overprovisioned MDisk is part of a provisioning group that defines the physical storage resources that are available to a set of MDisks.

Storage controllers report the usable capacity of an overprovisioned MDisk based on its provisioning group. If multiple MDisks are part of the same provisioning group, then all these MDisks share the physical storage resources and report the same usable capacity. However, this usable capacity is not available to each MDisk individually because it is shared between all these MDisks.

Provisioning groups are used differently depending on the back-end storage, as shown in the following examples:

- ▶ IBM FlashSystem A9000 and IBM FlashSystem 900: The entire subsystem forms one provisioning group.
- ▶ Storwize and IBM FlashSystem family systems: The storage pool forms a provisioning group, which enables more than one independent provisioning group in a system.
- ▶ RAID with compressing drives: An array is a provisioning group that presents the physical storage that is in use much like an external array.

Capacity usage should primarily be monitored on the overprovisioned back-end storage controller itself.

From the system, capacity usage can be monitored on overprovisioned MDisks by using one of the following methods:

- ▶ The GUI dashboard, as shown in Figure 9-24.

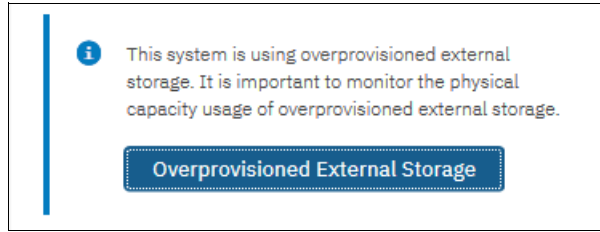


Figure 9-24 Dashboard button for overprovisioned storage monitoring

Click **Overprovisioned External Storage** to show an overview of overprovisioned MDisks and provisioning groups that are used by the system, as shown in Figure 9-25.

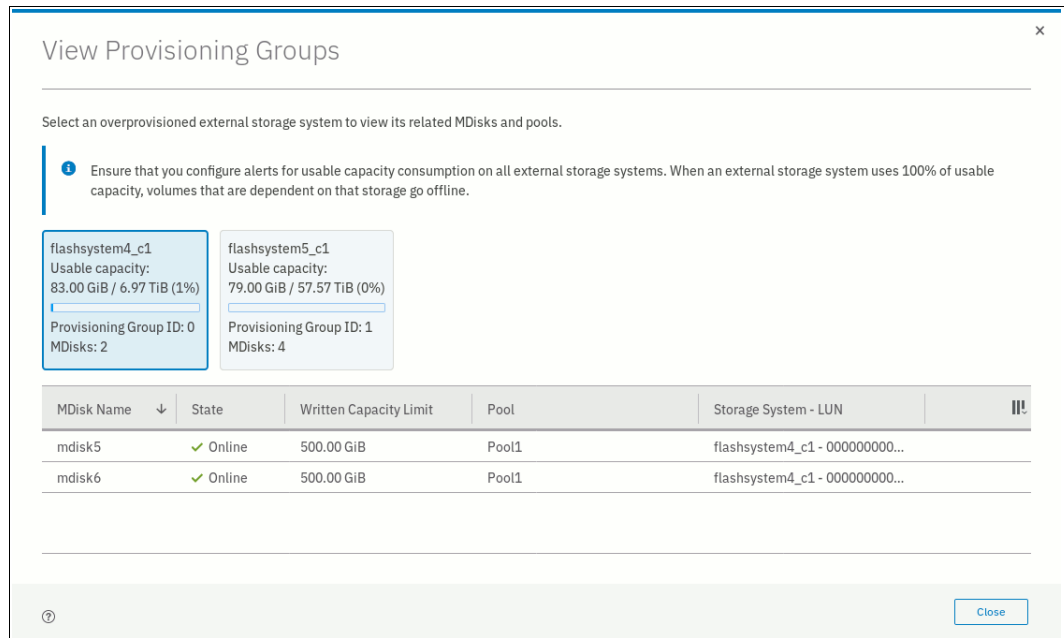


Figure 9-25 View Provisioning Groups

- ▶ The MDisk properties window, which opens by selecting **Pools** → **MDisks by Pools**, right-clicking an MDisk, and then selecting the **Properties** option, as shown in Figure 9-26.

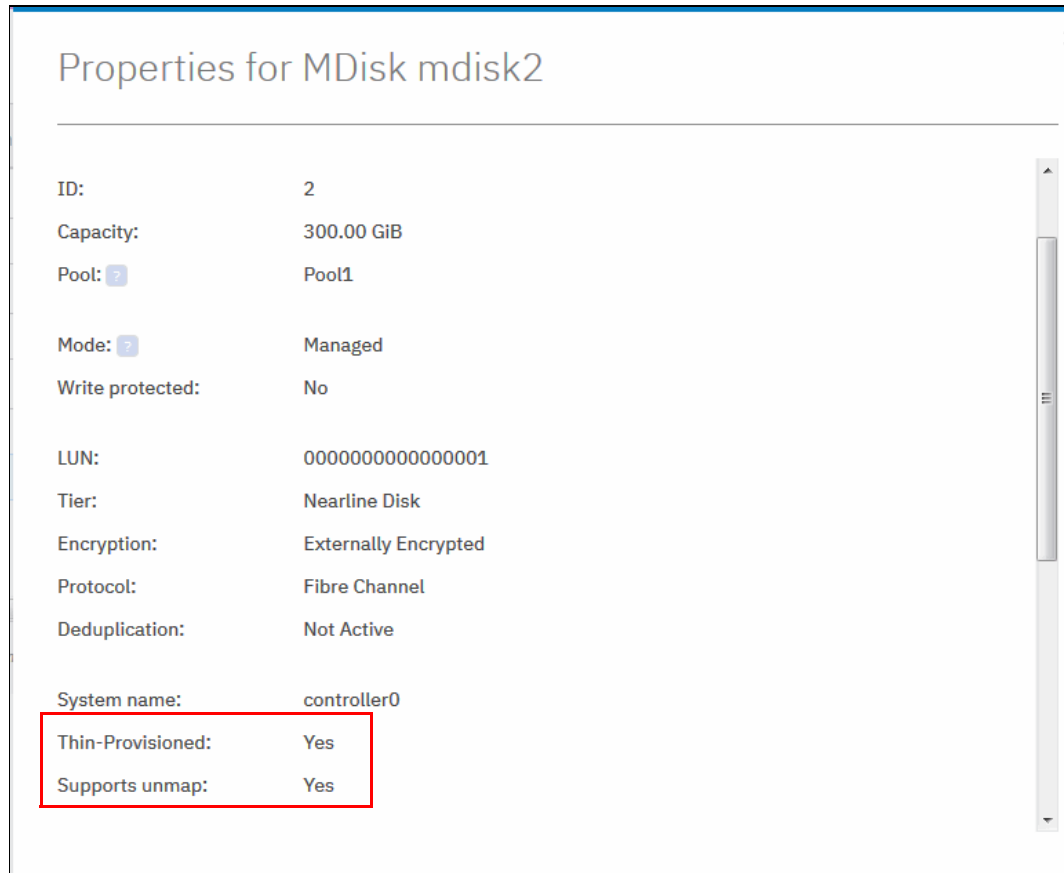


Figure 9-26 Thin-provisioned MDisk properties

- ▶ Running `lsmdisk` with an MDisk name or ID as a parameter displays the full properties of a thin-provisioned volume, as shown in Example 9-11.

Example 9-11 The `lsmdisk` parameters for thin-provisioned MDisks

```
IBM FlashSystem 7200:ITS0FS7K:superuser>lsmdisk mdisk2
id 2
name mdisk2
status online
mode managed
<...>
dedupe no
<...>
over_provisioned yes
supports_unmap yes
provisioning_group_id
physical_capacity 299.00GB
physical_free_capacity 288.00GB
write_protected no
allocated_capacity 11.00GB
effective_used_capacity 300.00GB
```

The overprovisioning status and SCSI UNMAP support for the selected MDisk are displayed.

The **physical_capacity** and **physical_free_capacity** parameters belong to the MDisk's provisioning group. They indicate the total physical storage capacity and formatted available physical space in the provisioning group that contains this MDisk.

Note: It is not a best practice to create multiple storage pools from MDisks in a single provisioning group.



Advanced Copy Services

This chapter describes the Advanced Copy Services that are a group of functions that provide different methods of data copy. It also describes the storage software capabilities to support the interaction with hybrid clouds. These functions are enabled by IBM Spectrum Virtualize software.

This chapter includes the following topics:

- ▶ 10.1, “IBM FlashCopy” on page 554
- ▶ 10.2, “Managing FlashCopy by using the GUI” on page 584
- ▶ 10.3, “Transparent Cloud Tiering” on page 621
- ▶ 10.4, “Implementing Transparent Cloud Tiering” on page 624
- ▶ 10.5, “Volume mirroring and migration options” on page 634
- ▶ 10.6, “Remote Copy” on page 636
- ▶ 10.7, “Remote Copy commands” on page 665
- ▶ 10.8, “Native IP replication” on page 672
- ▶ 10.9, “Managing Remote Copy by using the GUI” on page 693
- ▶ 10.10, “Remote Copy memory allocation” on page 717
- ▶ 10.11, “Troubleshooting Remote Copy” on page 718

10.1 IBM FlashCopy

Through the IBM FlashCopy function of the IBM Spectrum Virtualize, you can perform a *point-in-time (PiT) copy* of one or more volumes. This section describes the inner workings of FlashCopy and provides more information about its configuration and use.

You can use FlashCopy to help you solve critical and challenging business needs that require duplication of data of your source volume. Volumes can remain online and active while you create consistent copies of the data sets. Because the copy is performed at the block level, it operates below the host operating system (OS) and its cache. Therefore, the copy is not apparent to the host unless it is mapped.

While the FlashCopy operation is performed, the source volume is frozen briefly to initialize the FlashCopy bitmap after which I/O can resume. Although several FlashCopy options require the data to be copied from the source to the target in the background (which can take time to complete), the resulting data on the target volume is presented so that the copy appears to complete immediately. This feature means that the copy can immediately be mapped to a host and is directly accessible for read *and* write operations.

10.1.1 Business requirements for FlashCopy

When you are deciding whether FlashCopy addresses your needs, you must adopt a combined business and technical view of the problems that you want to solve. First, determine the needs from a business perspective. Then, determine whether FlashCopy can address the technical needs of those business requirements.

The business applications for FlashCopy are wide-ranging. Common use cases for FlashCopy include, but are not limited to, the following examples of rapidly creating:

- ▶ Consistent backups of dynamically changing data
- ▶ Consistent copies of production data to facilitate data movement or migration between hosts
- ▶ Copies of production data sets for application development and testing, auditing purposes and data mining, and for quality assurance

Regardless of your business needs, FlashCopy within the IBM Spectrum Virtualize is flexible and offers a broad feature set, which makes it applicable to several scenarios.

Back up improvements with FlashCopy

FlashCopy does not reduce the time that it takes to perform a backup to traditional backup infrastructure. However, it can be used to minimize and under certain conditions, eliminate application downtime that is associated with performing backups. FlashCopy can also transfer the resource usage of performing intensive backups from production systems.

After the FlashCopy is performed, the resulting image of the data can be backed up to tape, as though it were the source system. After the copy to tape is completed, the image data is redundant and the target volumes can be discarded. For time-limited applications, such as these examples, “no copy” or incremental FlashCopy is used most often. The use of these methods puts less load on your servers infrastructure.

When FlashCopy is used for backup purposes, the target data often is managed as read-only *at the OS level*. This approach provides extra security by ensuring that your target data was not modified and remains true to the source.

Restore with FlashCopy

FlashCopy can perform a restore from any FlashCopy mapping. Therefore, you can restore (or copy) from the target to the source of your regular FlashCopy relationships. When restoring data from FlashCopy, this method can be qualified as reversing the direction of the FlashCopy mappings.

This capability has the following benefits:

- ▶ Pairing mistakes are not a concern. You trigger a restore.
- ▶ The process appears instantaneous.
- ▶ You can maintain a pristine image of your data while you are restoring what was the primary data.

This approach can be used for various applications, such as recovering your production database application after an errant batch process that caused extensive damage.

Best practices: Although restoring from a FlashCopy is quicker than a traditional tape media restore, you must not use restoring from a FlashCopy as a substitute for good backup and archiving practices. Instead, keep one to several iterations of your FlashCopies so that you can near-instantly recover your data from the most recent history, and keep your long-term backup and archive as suitable for your business.

In addition to the restore option that copies the original blocks from the target volume to modified blocks on the source volume, the target can be used to perform a restore of individual files. To do that, you make the target available on a host. It is suggested to not make the target available to the source host because seeing duplicates of disks causes problems for most host OSs. Copy the files to the source by using normal host data copy methods for your environment.

For more information about how to use reverse FlashCopy, see 10.1.12, “Reverse FlashCopy” on page 575.

Moving and migrating data with FlashCopy

FlashCopy can be used to facilitate the movement or migration of data between hosts while minimizing downtime for applications. By using FlashCopy, application data can be copied from source volumes to new target volumes while applications remain online. After the volumes are fully copied and synchronized, the application can be brought down and then immediately brought back up on the new server that is accessing the new FlashCopy target volumes.

This method differs from the other migration methods, which are described later in this chapter. Common uses for this capability are host and back-end storage hardware refreshes.

Application testing with FlashCopy

It is often important to test a new version of an application or OS that is using actual production data. This testing ensures the highest quality possible for your environment. FlashCopy makes this type of testing easy to accomplish without putting the production data at risk or requiring downtime to create a constant copy.

You can create a FlashCopy of your source and use that for your testing. This copy is a duplicate of your production data down to the block level so that even physical disk identifiers are copied. Therefore, it is impossible for your applications to tell the difference.

You can also use the FlashCopy feature to create restart points for long running batch jobs. This option means that if a batch job fails several days into its run, it might be possible to restart the job from a saved copy of its data rather than rerunning the entire multiday job.

10.1.2 FlashCopy principles and terminology

The FlashCopy function creates a PiT or time-zero (T0) copy of data that is taken from a source volume and stored on a target volume by using *copy-on-write* (CoW) and copy on-demand, or a *redirect-on-write* (RoW) mechanism.

When a FlashCopy operation starts, a checkpoint creates a *bitmap table* that indicates that no part of the source volume was copied. Each bit in the bitmap table represents one region of the source volume and its corresponding region on the target volume. Each region is called a *grain*.

The relationship between two volumes defines the way data are copied and is called a *FlashCopy mapping*.

FlashCopy mappings between multiple volumes can be grouped in a Consistency group to ensure their PiT (or T0) is identical for all of them. A simple one-to-one FlashCopy mapping does not need to belong to a consistency group.

Figure 10-1 shows the basic terms that are used with FlashCopy. All elements are explained later in this chapter.

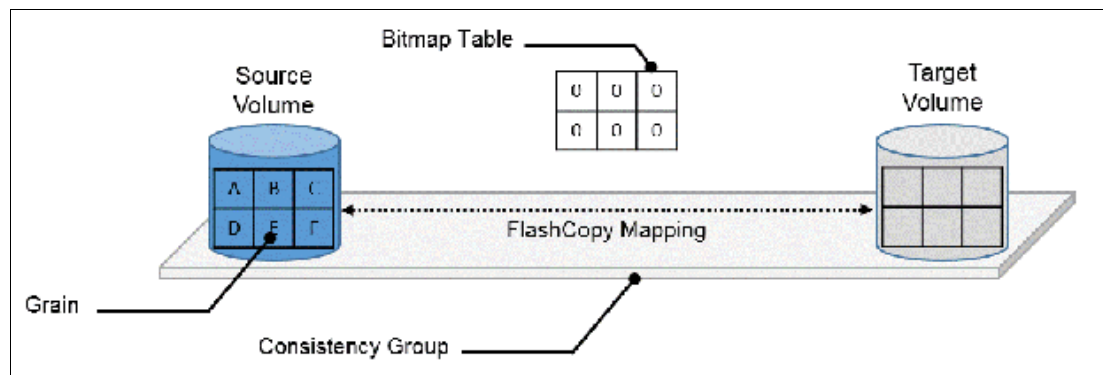


Figure 10-1 FlashCopy terminology

10.1.3 FlashCopy mapping

The relationship between the source volume and the target volume is defined by a FlashCopy mapping. The FlashCopy mapping can have three different types, four attributes, and seven different states.

The FlashCopy mapping can be one of the following types:

- ▶ **Snapshot:** Sometimes referred to as *nocopy*, a snapshot is a PiT copy of a volume without a background copy of the data from the source volume to the target. Only the changed blocks on the source volume are copied. The target copy cannot be used without an active link to the source.
- ▶ **Clone:** Sometimes referred to as *full copy*, a clone is a PiT copy of a volume with background copy of the data from the source volume to the target. All blocks from the source volume are copied to the target volume. The target copy becomes a usable independent volume.

- ▶ **Backup:** Sometimes referred to as *incremental*, a backup FlashCopy mapping consists of a PiT full copy of a source volume, plus periodic increments or “deltas” of data that changed between two points in time.

The FlashCopy mapping has four property attributes (clean rate, copy rate, autodelete, incremental) and seven different states that are described later in this chapter. Users can perform the following actions on a FlashCopy mapping:

- ▶ **Create:** Define a source and target, and set the properties of the mapping.
- ▶ **Prepare:** The system must be prepared before a FlashCopy copy starts. It flushes the cache and makes it “transparent” for a short time, so no data is lost.
- ▶ **Start:** The FlashCopy mapping is started and the copy begins immediately. The target volume is immediately accessible.
- ▶ **Stop:** The FlashCopy mapping is stopped (by the system or by the user). Depending on the state of the mapping, the target volume is usable or not usable.
- ▶ **Modify:** Some properties of the FlashCopy mapping can be modified after creation.
- ▶ **Delete:** Delete the FlashCopy mapping. This action does not delete volumes (source or target) from the mapping.

The source and target volumes must be the same size. The minimum granularity that IBM Spectrum Virtualize supports for FlashCopy is an entire volume. It is not possible to use FlashCopy to copy only part of a volume.

Important: As with any PiT copy technology, you are bound by OS and application requirements for interdependent data and the restriction to an entire volume.

The source and target volumes must belong to the same IBM Spectrum Virtualize based system. They do not have to be in the same I/O group or storage pool, although it is recommended that they have the same preferred node for the best performance.

Volumes that are members of a FlashCopy mapping cannot have their size increased or decreased while they are members of the FlashCopy mapping.

All FlashCopy operations occur on FlashCopy mappings. FlashCopy does not alter the volumes. However, multiple operations can occur at the same time on multiple FlashCopy mappings because of the use of consistency groups.

10.1.4 Consistency groups

To overcome the issue of dependent writes across volumes and to create a consistent image of the client data, a FlashCopy operation must be performed on multiple volumes as an atomic operation. To accomplish this method, the IBM Spectrum Virtualize supports the concept of *consistency groups*.

Consistency groups address the requirement to preserve PiT data consistency across multiple volumes for applications that include related data that spans multiple volumes. For these volumes, consistency groups maintain the integrity of the FlashCopy by ensuring that “dependent writes” are run in the application’s intended sequence. Also, consistency groups provide an easy way to manage several mappings.

FlashCopy mappings can be part of a consistency group, even if only one mapping exists in the consistency group. If a FlashCopy mapping is not part of any consistency group, it is referred to as *stand-alone*.

Dependent writes

It is crucial to use consistency groups when a data set spans multiple volumes. Consider the following typical sequence of writes for a database update transaction:

1. A write is run to update the database log, which indicates that a database update is about to be performed.
2. A second write is run to perform the update to the database.
3. A third write is run to update the database log, which indicates that the database update completed successfully.

The database ensures the correct ordering of these writes by waiting for each step to complete before the next step is started. However, if the database log (updates 1 and 3) and the database (update 2) are on separate volumes, it is possible for the FlashCopy of the database volume to occur before the FlashCopy of the database log. This sequence can result in the target volumes seeing writes 1 and 3 but not 2 because the FlashCopy of the database volume occurred before the write was completed.

In this case, if the database was restarted by using the backup that was made from the FlashCopy target volumes, the database log indicates that the transaction completed successfully. In fact, it did not complete successfully because the FlashCopy of the volume with the database file was started (the bitmap was created) before the write completed to the volume. Therefore, the transaction is lost and the integrity of the database is in question.

Most of the actions that the user can perform on a FlashCopy mapping are the same for consistency groups.

10.1.5 Crash consistent copy and hosts considerations

FlashCopy consistency groups do not provide application consistency. It ensures only that volume points-in-time are consistent between them.

Because FlashCopy is at the block level, it is necessary to understand the interaction between your application and the host OS. From a logical standpoint, it is easiest to think of these objects as “layers” that sit on top of one another. The application is the topmost layer, and beneath it is the OS layer.

Both of these layers have various levels and methods of caching data to provide better speed. Therefore, because the IBM Spectrum Virtualize and FlashCopy sit below these layers, they are unaware of the cache at the application or OS layers.

To ensure the integrity of the copy that is made, it is necessary to flush the host OS and application cache for any outstanding reads or writes before the FlashCopy operation is performed. Failing to flush the host OS and application cache produces what is referred to as a *crash consistent* copy.

The resulting copy requires the same type of recovery procedure, such as log replay and file system checks, that is required following a host crash. FlashCopies that are crash consistent often can be used after file system and application recovery procedures.

Various OSs and applications provide facilities to stop I/O operations and ensure that all data is flushed from host cache. If these facilities are available, they can be used to prepare for a FlashCopy operation. When this type of facility is unavailable, the host cache must be flushed manually by quiescing the application and unmounting the file system or drives.

The target volumes are overwritten with a complete image of the source volumes. Before the FlashCopy mappings are started, any data that is held on the host OS (or application) caches for the target volumes must be discarded. The easiest way to ensure that no data is held in these caches is to unmount the target volumes before the FlashCopy operation starts.

Best practice: From a practical standpoint, when you have an application that is backed by a database and you want to make a FlashCopy of that application's data, it is sufficient in most cases to use the write-suspend method that is available in most modern databases. This approach is possible because the database maintains strict control over I/O.

This method is as opposed to flushing data from the application and backing database, which is always the suggested method because it is safer. However, this method can be used when facilities do not exist or your environment includes time sensitivity.

IBM FlashCopy application-integrated solutions

IBM FlashCopy is not application-aware, and a third-party tool is needed to link the application to the FlashCopy operations.

IBM Spectrum Protect Snapshot protects data with integrated, application-aware snapshot backup and restore capabilities that use FlashCopy technologies in the IBM Spectrum Virtualize.

You can protect data that is stored by IBM Db2® SAP, Oracle, Microsoft Exchange, and Microsoft SQL Server applications. You can create and manage volume-level snapshots for file systems and custom applications.

In addition, you can use IBM Spectrum Protect Snapshot to manage frequent, near-instant, nondisruptive, application-aware backups and restores that use integrated application and VMware snapshot technologies. IBM Spectrum Protect Snapshot can be widely used in IBM and non-IBM storage systems.

Other IBM products are also available for application-aware backup and restore capabilities, such as IBM Spectrum Protect Plus and IBM Copy Data Management. For more information about these offerings, speak to your IBM representative.

Note: To see how IBM Spectrum Protect Snapshot, IBM Spectrum Protect Plus, and IBM Copy Data Management can help your business, see [IBM Documentation](#).

10.1.6 Grains and bitmap: I/O indirection

When a FlashCopy operation starts, a checkpoint is made of the source volume. No data is copied at the time that a start operation occurs. Instead, the checkpoint creates a bitmap that indicates that no part of the source volume was copied. Each bit in the bitmap represents one region of the source volume. Each region is called a *grain*.

You can think of the bitmap as a simple table of ones or zeros. The table tracks the difference between a source volume grains and a target volume grains. At the creation of the FlashCopy mapping, the table is filled with zeros, which indicates that no grain is copied yet.

When a grain is copied from source to target, the region of the bitmap that refers to that grain is updated (for example, from “0” to “1”), as shown in Figure 10-2.

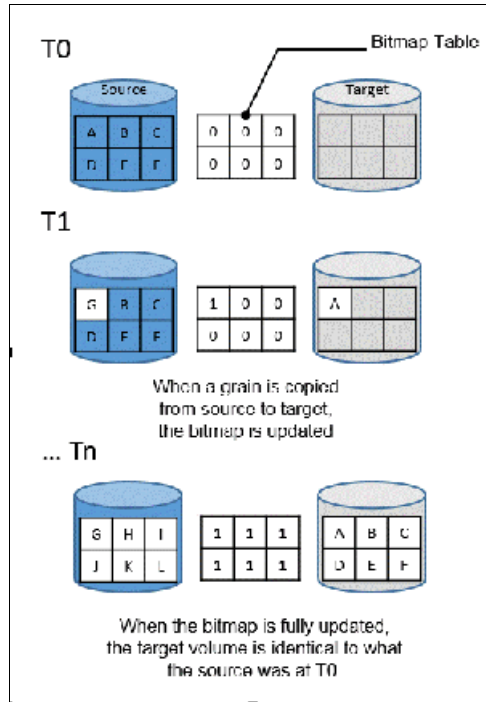


Figure 10-2 A simplified representation of grains and bitmap

The grain size can be 64 KB or 256 KB (the default is 256 KB). The grain size cannot be selected by the user when a FlashCopy mapping is created from the GUI. The FlashCopy bitmap contains 1 bit for each grain. The bit records whether the associated grain is split by copying the grain from the source to the target.

After a FlashCopy mapping is created, the grain size for that FlashCopy mapping cannot be changed. When a FlashCopy mapping is created, the grain size of that mapping is used if the grain size parameter is not specified and one of the volumes in the mapping is part of a FlashCopy mapping.

If neither volume in the new mapping is part of another FlashCopy mapping and at least one of the volumes in the mapping is a compressed volume, the default grain size is 64 KB for performance considerations. Other than in this situation, the default grain size is 256 KB.

Copy-on-write, redirect-on-write, and Copy on Demand

With IBM Spectrum Virtualize V8.4 and later, FlashCopy uses a CoW mechanism to copy data from a source volume to a target volume in standard pools (non-Data Reduction Pools (DRPs)), and a RoW mechanism in DRPs. In earlier versions, it uses only CoW regardless of the Pool type.

With CoW, as shown in Figure 10-3, when data is written on a source volume, the grain where the to-be-changed blocks are stored is first copied to the target volume and then modified on the source volume. The bitmap is updated to track the copy.

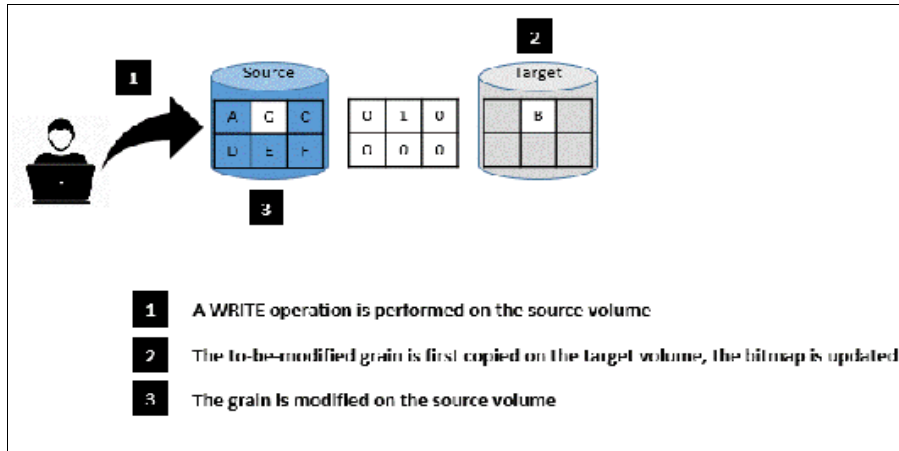


Figure 10-3 Copy-on-write steps

With RoW, when the source volume is modified, the updated grain is written directly to a new block in the DRP customer data volume. The source volume metadata and FlashCopy bitmap are then updated to reflect this update. RoW was introduced with IBM Spectrum Virtualize V8.4 for DRPs only. Compared to CoW, RoW reduces the back-end spectrum by removing the copy operation, which improves the overall performance of FlashCopy operations.

Note: At the time of writing, RoW is used only for volumes with supported deduplication, without a mirroring relationship, and when both the source and target volumes are within the same pool and I/O group. The selection between CoW versus RoW is automatically done by the base code under these conditions.

With IBM FlashCopy, the target volume is immediately accessible for read *and* write operations. Therefore, a target volume can be modified even if it is part of a FlashCopy mapping. In standard pools, as shown in Figure 10-4, when a write operation is performed on the *target* volume, the grain that contains the blocks to be changed is first copied from the source (CoD). Then, the grain is modified with the new value. The bitmap is modified so that the grain from the source is *not* copied again, even if it is changed or a background copy is enabled.

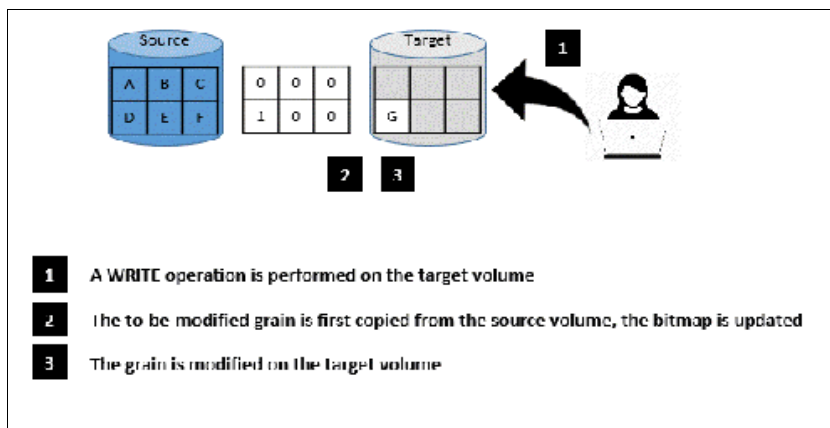


Figure 10-4 Copy on Demand steps

Starting with IBM Spectrum Virtualize V8.4 and later, this behavior is slightly different in DRPs. The software reads the grain to be updated from the source volume, modifies the grain with the new value in the cache, writes the modified grain to the DRP customer data volume, and updates the FlashCopy bitmap.

Note: If all the blocks of the grain to be modified are changed, there is no need to first read or copy the source grain. There is no CoD, so the source grain is directly modified at the target volume.

FlashCopy indirection layer

The FlashCopy indirection layer governs the I/O to the source and target volumes when a FlashCopy mapping is started, which is done by using the FlashCopy bitmap. The purpose of the FlashCopy indirection layer is to enable the source and target volumes for read and write I/O immediately after the FlashCopy is started.

The indirection Layer intercepts any I/O coming from a host (read or write operation) and addressed to a FlashCopy volume (source or target). It determines whether the addressed volume is a source or a target, its direction (read or write), and the state of the bitmap table for the FlashCopy mapping that the addressed volume is in. It then decides what operation to perform. The different I/O indirections are described next.

Read from the source Volume

When a user performs a read operation on the source volume, there is no redirection. The operation is similar to what is done with a volume that is not part of a FlashCopy mapping.

Write on the source Volume

Performing a write operation on the source volume modifies a block or a set of blocks, which modifies a grain on the source. It generates one of the following actions, depending on the state of the grain to be modified.

Consider the following points:

- ▶ If the bitmap indicates that the grain was copied, the source grain is changed and the target volume and the bitmap table remain unchanged, as shown in Figure 10-5.

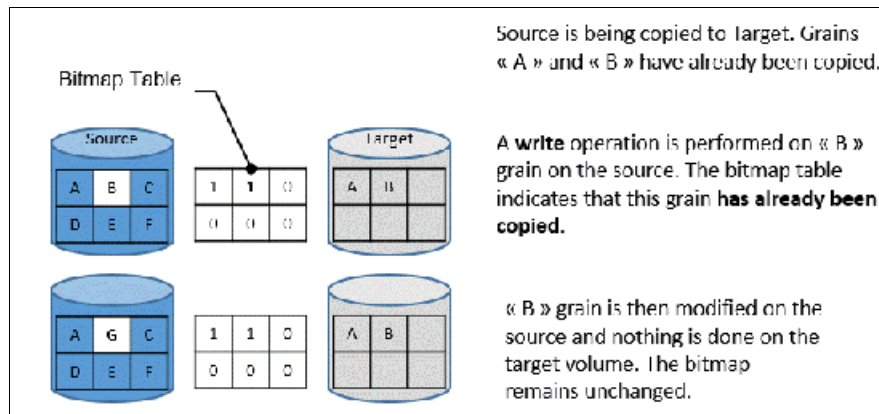


Figure 10-5 Modifying an already copied grain on the source

- ▶ If the bitmap indicates that the grain is not yet copied, the grain is first copied to the target (CoW), the bitmap table is updated, and the grain is modified on the source, as shown in Figure 10-6. This process is true for standard pools in IBM Spectrum Virtualize V8.4 or later or any pool type if the version is lower than Version 8.4.

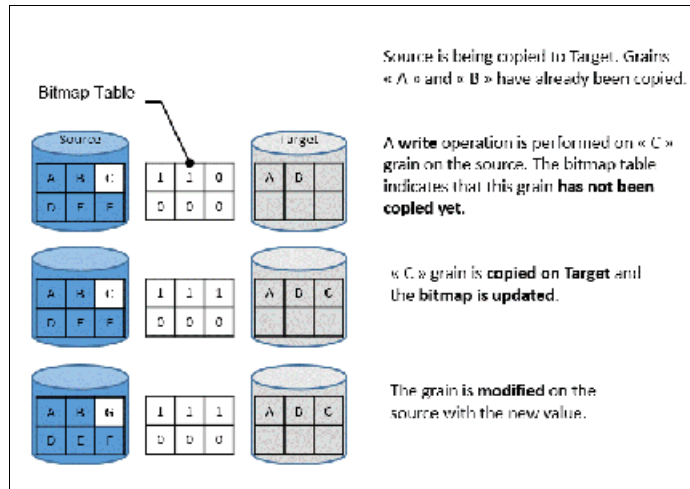


Figure 10-6 Modifying a non-copied grain on the source

- ▶ If this pool is a DRP in a system running IBM Spectrum Virtualize V8.4 or later, the system does a RoW operation, as described in “Copy-on-write, redirect-on-write, and Copy on Demand” on page 560.

Write on a target volume

Because FlashCopy target volumes are immediately accessible in Read and Write mode, it is possible to perform write operations on the target volume when the FlashCopy mapping is started. Performing a write operation on the target generates one of the following actions, depending on the bitmap:

- ▶ If the bitmap indicates the grain to be modified on the target was not yet copied, it is first copied from the source (copy on demand). The bitmap is updated, and the grain is modified on the target with the new value, as shown in Figure 10-7. The source volume remains unchanged.

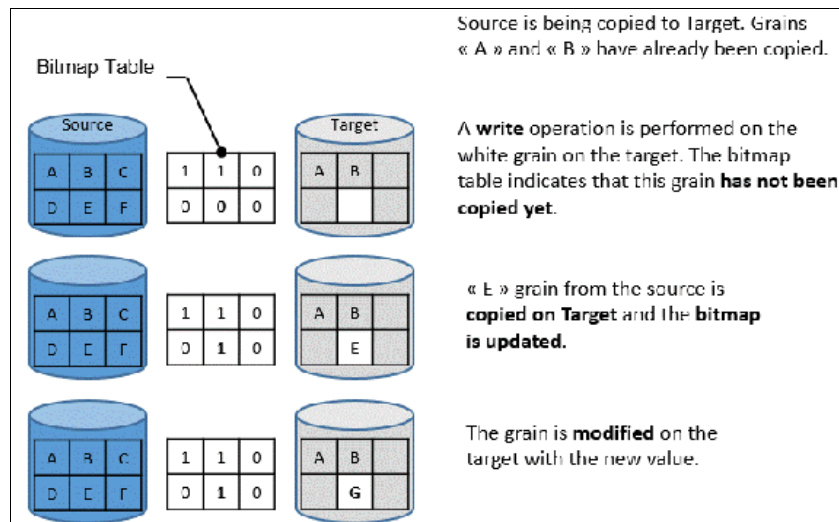


Figure 10-7 Modifying a non-copied grain on the target

Note: If the entire grain is to be modified and not only part of it (some blocks only), the copy on demand is bypassed. The bitmap is updated, and the grain on the target is modified but not copied first.

- ▶ If the bitmap indicates the grain to be modified on the target was copied, it is directly changed. The bitmap is *not* updated, and the grain is modified on the target with the new value, as shown in Figure 10-8.

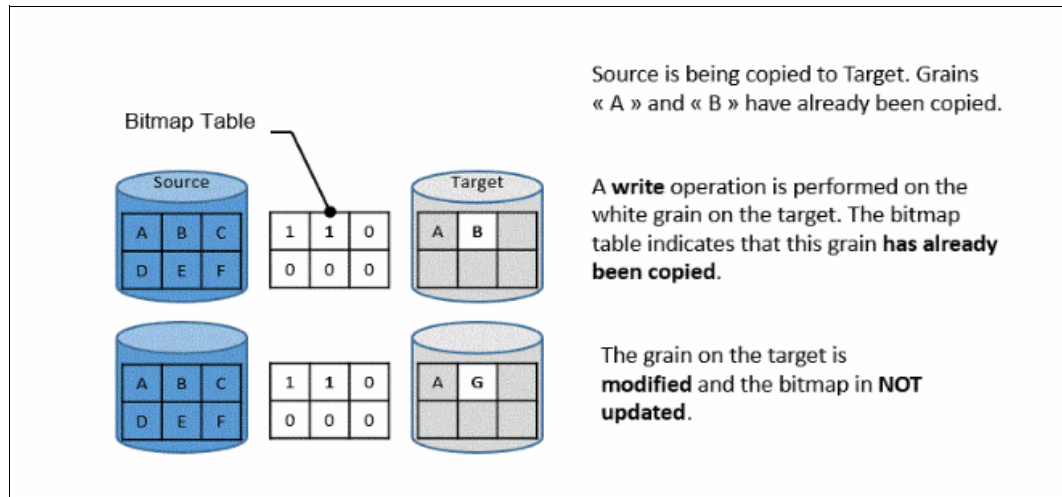


Figure 10-8 Modifying an already copied grain on the target

Note: The bitmap is not updated in that case. Otherwise, it might be copied from the source late if a background copy is ongoing or if write operations are made on the source. That process over-writes the changed grain on the target.

Read from a target volume

Performing a read operation on the target volume returns the value in the grain on the source or on the target, depending on the bitmap. Consider the following points:

- ▶ If the bitmap indicates that the grain was copied from the source or that the grain was modified on the target, the grain on the target is read, as shown in Figure 10-9 on page 565.
- ▶ If the bitmap indicates that the grain was not yet copied from the source or was not modified on the target, the grain on the source is read, as shown in Figure 10-9 on page 565.

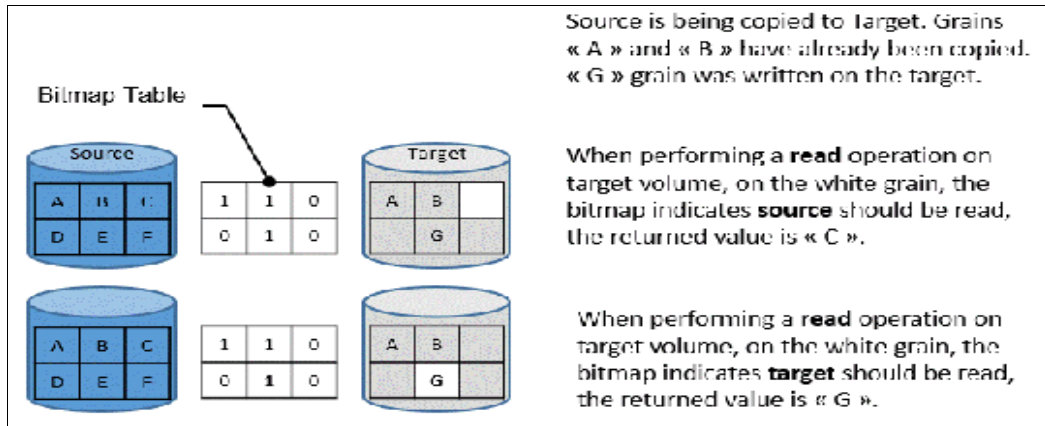


Figure 10-9 Reading a grain on a target

If source has multiple targets, the Indirection layer algorithm behaves differently on Target I/Os. For more information about multi-target operations, see 10.1.11, “Multiple target FlashCopy” on page 570.

10.1.7 Interacting with the cache

IBM Spectrum Virtualize based systems have their cache divided into upper and lower cache. The upper cache serves mostly as a write cache and hides the write latency from the hosts and application. The lower cache is a read/write cache that optimizes I/O to and from disks.

Figure 10-10 shows the IBM Spectrum Virtualize software stack, which includes the cache architecture.

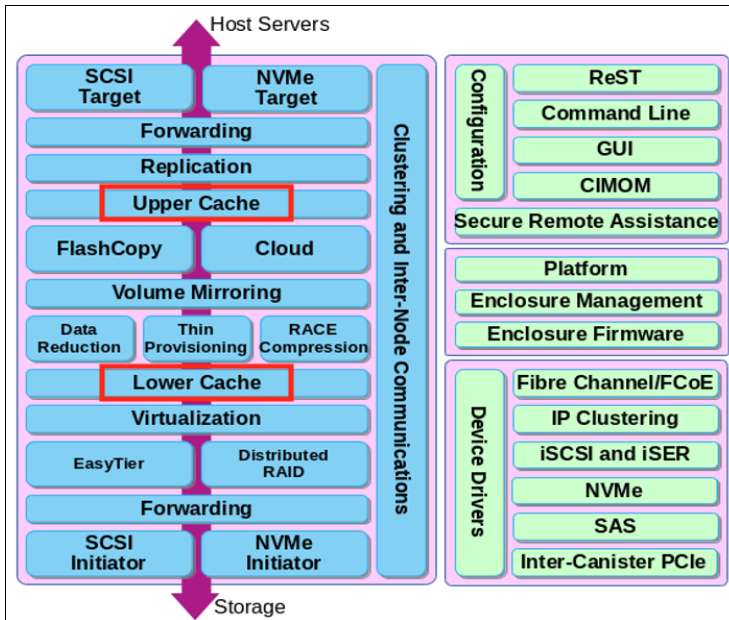


Figure 10-10 IBM Spectrum Virtualize software architecture

The CoW process introduces significant latency into write operations. To isolate the active application from this extra latency, the FlashCopy indirection layer is placed logically between the upper and lower cache. Therefore, the extra latency that is introduced by the CoW process is encountered only by the internal cache operations and not by the application. With IBM Spectrum Virtualize V8.4 and later, the RoW mechanism that is used for DRPs aims to reduce the impact that is introduced by CoW.

Also, the two-level cache provides more performance improvements to the FlashCopy mechanism. Because the FlashCopy layer is above the lower cache in the IBM Spectrum Virtualize software stack, it can benefit from read prefetching and coalescing writes to back-end storage. FlashCopy benefits from the two-level cache because the upper cache write data does not have to go directly to back-end storage, but to the lower cache layer instead.

10.1.8 Background Copy Rate

The Background Copy Rate is a property of a FlashCopy mapping. A grain copy from the source to the target can occur when triggered by a write operation on the source or target volume, or when background copy is enabled. With background copy enabled, the target volume eventually becomes a clone of the source volume at the time the mapping was started (T0). When the copy is completed, the mapping can be removed between the two volumes and you can end up with two independent volumes.

The background copy rate property determines the speed at which grains are copied as a background operation, immediately after the FlashCopy mapping is started. That speed is defined by the user when the FlashCopy mapping is created, and can be changed dynamically for each individual mapping, whatever its state. Mapping copy rate values can be 0 - 150, with the corresponding speeds that are listed in Table 10-1.

Table 10-1 Copy rate values

User-specified copy rate attribute value	Data copied/sec	256 KB grains/sec	64 KB grains/sec
1 - 10	128 kibibytes (KiB)	0.5	2
11 - 20	256 KiB	1	4
21 - 30	512 KiB	2	8
31 - 40	1 mebibyte (MiB)	4	16
41 - 50	2 MiB	8	32
51 - 60	4 MiB	16	64
61 - 70	8 MiB	32	128
71 - 80	16 MiB	64	256
81 - 90	32 MiB	128	512
91 - 100	64 MiB	256	1024
101 - 110	128 MiB	512	2048
111 - 120	256 MiB	1024	4096
121 - 130	512 MiB	2048	8192

User-specified copy rate attribute value	Data copied/sec	256 KB grains/sec	64 KB grains/sec
131 - 140	1 GiB	4096	16384
141 - 150	2 GiB	8192	32768

When the background copy function is not performed (copy rate = 0), the target volume remains a valid copy of the source data only while the FlashCopy mapping remains in place.

The *grains per second* numbers represent the maximum number of grains that IBM Spectrum Virtualize copies per second. This amount assumes that the bandwidth to the managed disks (MDisks) can accommodate this rate.

If IBM Spectrum Virtualize cannot achieve these copy rates because of insufficient bandwidth from the nodes to the MDisks, the background copy I/O contends for resources on an equal basis with the I/O that is arriving from the hosts. Background copy I/O and I/O that is arriving from the hosts tend to see an increase in latency and a consequential reduction in throughput.

Background copy and foreground I/O continue to progress, and do not stop, hang, or cause the node to fail.

The background copy is performed by one of the nodes that belong to the I/O group in which the source volume is stored. This responsibility is moved to the other node in the I/O group if the node that performs the background and stopping copy fails.

10.1.9 Incremental FlashCopy

When a FlashCopy mapping is stopped (because the entire source volume was copied onto the target volume or a user manually stopped it), the bitmap table is reset. Therefore, when the same FlashCopy is started again, the copy process is restarted from the beginning.

Running the **-incremental** option when creating the FlashCopy mapping allows the system to keep the bitmap as it is when the mapping is stopped. Therefore, when the mapping is started again (at another PiT), the bitmap is reused and only changes between the two copies are applied to the target.

A system that provides Incremental FlashCopy capability allows the system administrator to refresh a target volume without having to wait for a full copy of the source volume to be complete. At the point of refreshing the target volume, if the data was changed on the source or target volumes for a particular grain, the grain from the source volume is copied to the target.

The advantages of Incremental FlashCopy are useful only if a previous full copy of the source volume was obtained. Incremental FlashCopy helps with only further recovery time objectives (RTOs, which are the time that is needed to recover data from a previous state), it does not help with the initial RTO.

For example, as shown in Figure 10-11 on page 568, a FlashCopy mapping was defined between a source volume and a target volume by using the **-incremental** option.

Consider the following points:

- ▶ The mapping is started on the Copy1 date. A *full copy* of the source volume is made, and the bitmap is updated every time that a grain is copied. At the end of Copy1, all grains are copied and the target volume is an exact replica of the source volume at the beginning of Copy1. Although the mapping is stopped, the bitmap is maintained because of the `-incremental` option.
- ▶ Changes are made on the source volume and the bitmap is updated, although the FlashCopy mapping is not active. For example, grains E and C on the source are changed in G and H, their corresponding bits are changed in the bitmap. The target volume is untouched.
- ▶ The mapping is started again on Copy2 date. The bitmap indicates that only grains E and C were changed, so only G and H are copied on the target volume. The other grains do not need to be copied because they were copied the first time. The copy time is much quicker than for the first copy as only a fraction of the source volume is copied.

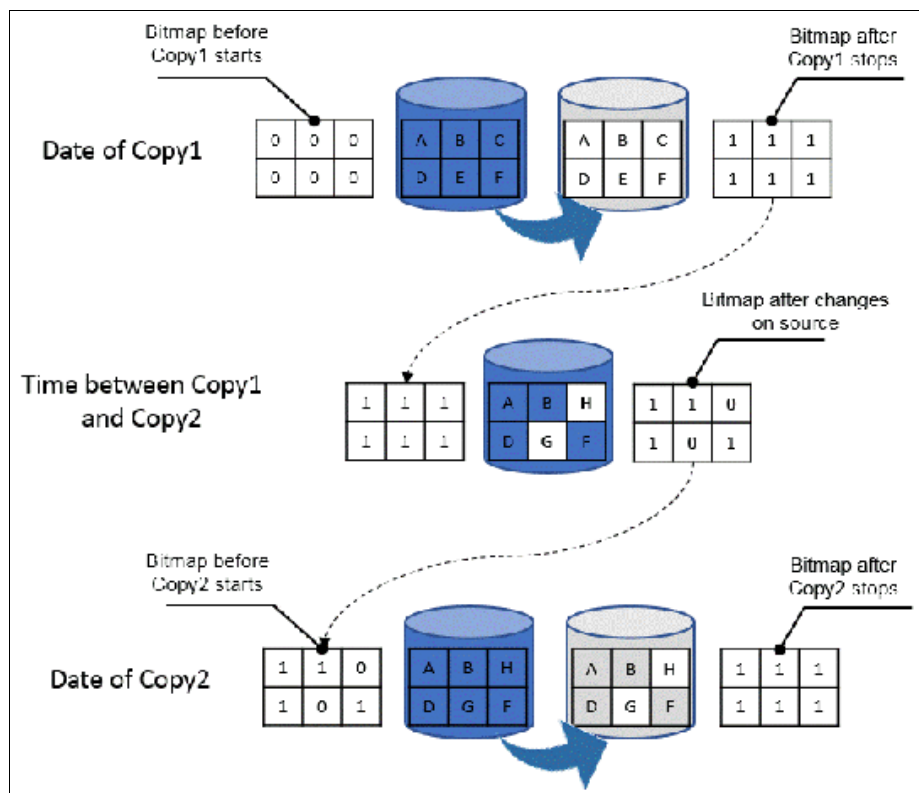


Figure 10-11 Incremental FlashCopy example

10.1.10 Starting FlashCopy mappings and consistency groups

You can prepare, start, or stop FlashCopy on a stand-alone mapping or a consistency group.

When the command-line interface (CLI) is used to perform FlashCopy on volumes, run a `prestartfcmap` or `prestartfcconsistgrp` command *before* you start a FlashCopy (regardless of the type and options specified). These commands put the cache into write-through mode and provides a flushing of the I/O that is bound for your volume. After FlashCopy is started, an effective copy of a source volume to a target volume is created.

The content of the source volume is presented immediately on the target volume and the original content of the target volume is lost.

Then, FlashCopy commands can be run to the FlashCopy consistency group and simultaneously for all the FlashCopy mappings that are defined in the consistency group. For example, when a FlashCopy **start** command is run to the consistency group, all of the FlashCopy mappings in the consistency group are started at the same time. This simultaneous start results in a PiT copy that is consistent across all of the FlashCopy mappings that are contained in the consistency group.

Rather than running **prestartfcmap** or **prestartfcconsistgrp**, you can also use the **-prep** parameter in the **startfcmap** or **startfcconsistgrp** command to prepare and start FlashCopy in one step.

Important: After an individual FlashCopy mapping is added to a consistency group, it can be managed as part of the group only. Operations, such as prepare, start, and stop, are no longer allowed on the individual mapping.

FlashCopy mapping states

At any point, a mapping is in one of the following states:

► Idle or copied

The source and target volumes act as independent volumes, even if a mapping exists between the two. Read and write caching is enabled for the source and the target volumes. If the mapping is incremental and the background copy is complete, the mapping records only the differences between the source and target volumes. If the connection to both nodes in the I/O group that the mapping is assigned to is lost, the source and target volumes are offline.

► Copying

The copy is in progress. Read and write caching is enabled on the source and the target volumes.

► Prepared

The mapping is ready to start. The target volume is online, but is not accessible. The target volume cannot perform read or write caching. Read and write caching is failed by the Small Computer System Interface (SCSI) front end as a hardware error. If the mapping is incremental and a previous mapping completed, the mapping records only the differences between the source and target volumes. If the connection to both nodes in the I/O group that the mapping is assigned to is lost, the source and target volumes go offline.

► Preparing

The target volume is online, but not accessible. The target volume cannot perform read or write caching. Read and write caching is failed by the SCSI front end as a hardware error. Any changed write data for the source volume is flushed from the cache. Any read or write data for the target volume is discarded from the cache. If the mapping is incremental and a previous mapping completed, the mapping records only the differences between the source and target volumes. If the connection to both nodes in the I/O group that the mapping is assigned to is lost, the source and target volumes go offline.

► Stopped

The mapping is stopped because you issued a stop command or an I/O error occurred. The target volume is offline and its data is lost. To access the target volume, you must restart or delete the mapping. The source volume is accessible and the read and write cache is enabled. If the mapping is incremental, the mapping is recording write operations to the source volume. If the connection to both nodes in the I/O group that the mapping is assigned to is lost, the source and target volumes go offline.

► Stopping

The mapping is copying data to another mapping. If the background copy process is complete, the target volume is online while the stopping copy process completes. If the background copy process is incomplete, data is discarded from the target volume cache. The target volume is offline while the stopping copy process runs. The source volume is accessible for I/O operations.

► Suspended

The mapping started, but it did not complete. Access to the metadata is lost, which causes the source and target volume to go offline. When access to the metadata is restored, the mapping returns to the copying or stopping state and the source and target volumes return online. The background copy process resumes. If the data was not flushed and was written to the source or target volume before the suspension, it is in the cache until the mapping leaves the suspended state.

Summary of FlashCopy mapping states

Table 10-2 lists the various FlashCopy mapping states and the corresponding states of the source and target volumes.

Table 10-2 FlashCopy mapping state summary

State	Source		Target	
	Online/Offline	Cache state	Online/Offline	Cache state
Idling/Copied	Online	Write-back	Online	Write-back
Copying	Online	Write-back	Online	Write-back
Stopped	Online	Write-back	Offline	N/A
Stopping	Online	Write-back	► Online if copy complete ► Offline if copy incomplete	N/A
Suspended	Offline	Write-back	Offline	N/A
Preparing	Online	Write-through	Online but not accessible	N/A
Prepared	Online	Write-through	Online but not accessible	N/A

10.1.11 Multiple target FlashCopy

A volume can be the source of multiple target volumes. A target volume can also be the source of another target volume. However, a target volume can have only one source volume. A source volume can have multiple target volumes in one or multiple consistency groups. A consistency group can contain multiple FlashCopy mappings (source-target relations). A source volume can belong to multiple consistency groups.

Figure 10-12 shows these different possibilities.

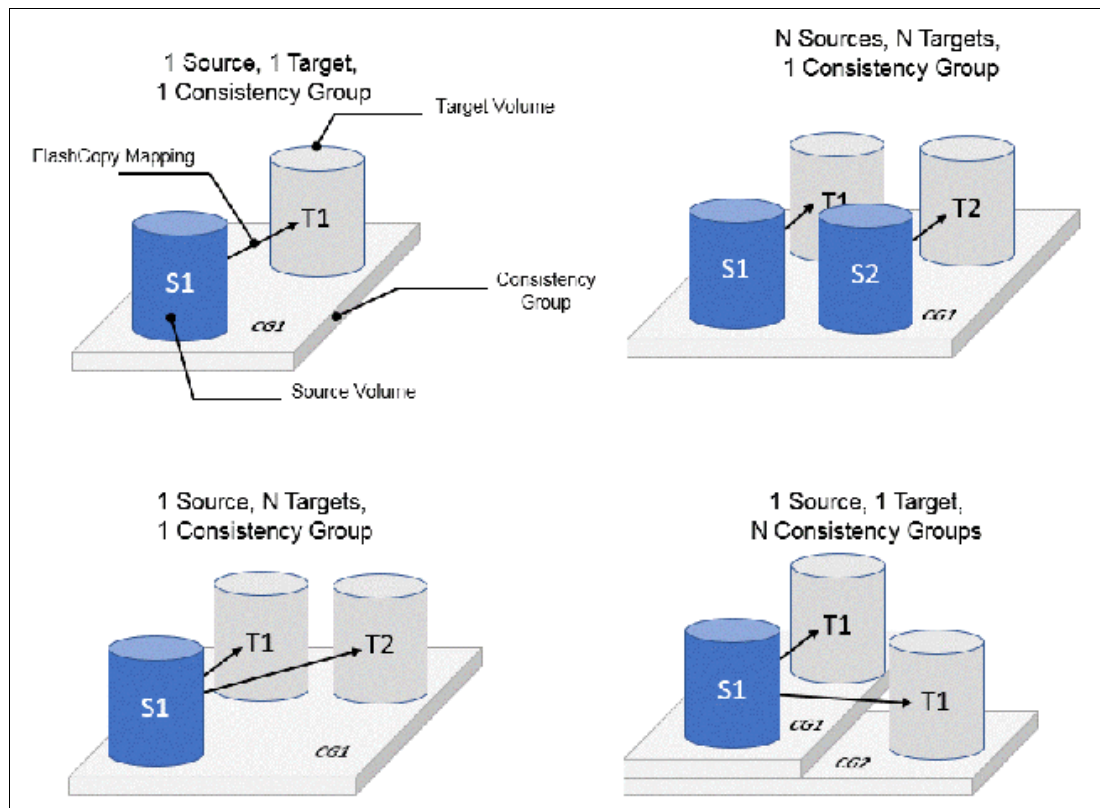


Figure 10-12 Consistency groups and mappings combinations

Every source-target relationship is a FlashCopy mapping and is maintained with its own bitmap table. No consistency group bitmap table exists.

When a source volume is in a FlashCopy mapping with multiple targets, in multiple consistency groups, it allows the copy of a single source at multiple points in time and therefore, keeps multiple versions of a single volume.

Consistency group with multiple target FlashCopy

A consistency group aggregates FlashCopy mappings, not volumes. Therefore, where a source volume has multiple FlashCopy mappings, they can be in the same or separate consistency groups.

If a particular volume is the source volume for multiple FlashCopy mappings, you might want to create separate consistency groups to separate each mapping of the same source volume. Regardless of whether the source volume with multiple target volumes is in the same consistency group or in separate consistency groups, the resulting FlashCopy produces multiple identical copies of the source data.

Dependencies

When a source volume has multiple target volumes, a mapping is created for each source-target relationship. When data is changed on the source volume, it is first copied to the target volume because of the CoW mechanism that is used by FlashCopy. If running IBM Spectrum Virtualize V8.4 or later, for DRP pools the software uses a RoW mechanism instead, as described in “Copy-on-write, redirect-on-write, and Copy on Demand” on page 560.

You can create up to 256 targets for a single source volume. Therefore, a single write operation on the source volume might result in 256 write operations (one per target volume) when using CoW. This configuration generates a large workload that the system might not be able to handle, which might lead to a heavy performance impact on front-end operations.

To avoid any significant effect on performance because of multiple targets, FlashCopy creates dependencies between the targets. Dependencies can be considered as “hidden” FlashCopy mappings that are not visible to and cannot be managed by the user. A dependency is created between the most recent target and the previous one (in order of start time). Figure 10-13 shows an example of a source volume with three targets.

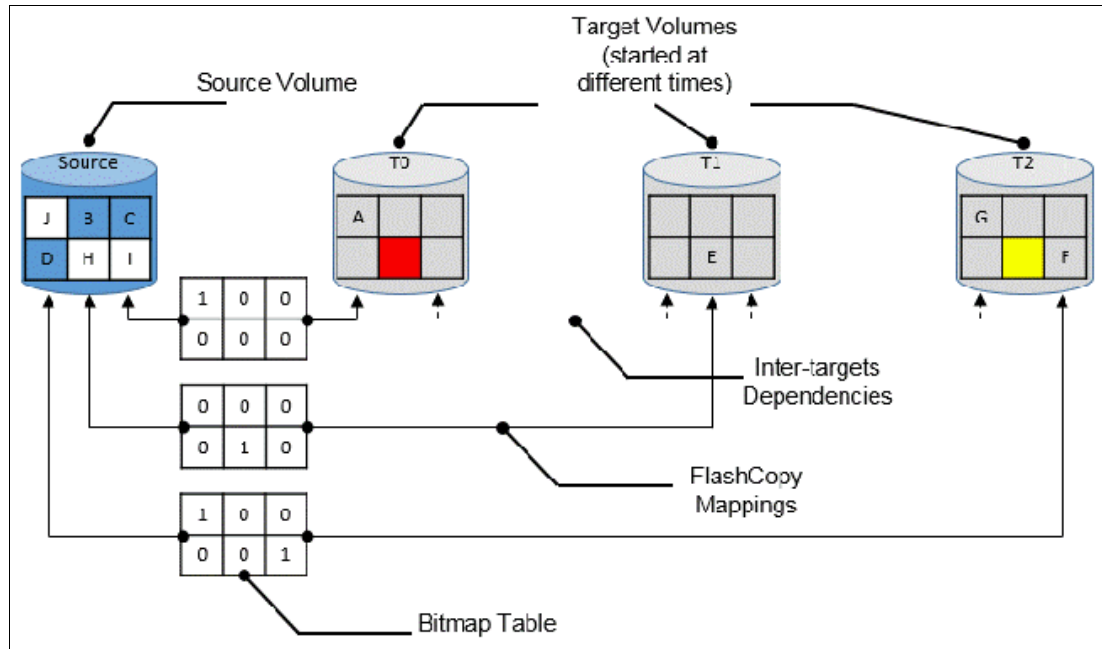


Figure 10-13 FlashCopy dependencies example

When the three targets are started, Target T0 was started first and considered the “oldest.” Target T1 was started next and is considered “next oldest,” and finally, Target T2 was started last and considered the “most recent” or “newest.” The “next oldest” target for T2 is T1. The “next oldest” target for T1 is T0. T1 is newer than T2, and T0 is newer than T1.

Source read with multiple target FlashCopy

No specific behavior is shown for read operations on source volumes when multiple targets exist for that volume. The data is always read from the source.

Source write with multiple target FlashCopy (CoW)

A write to the source volume does not cause its data to be copied to all of the targets. Instead, it is copied to the most recent target volume only. For example, consider the sequence of events that are listed in Table 10-3 for a source volume and three targets that are started at different times. In this example, no background copy exists. The “most recent” target is indicated with an asterisk.

Table 10-3 Sequence example of write I/Os on a source with multiple targets

Time	Source volume	Target T0	Target T1	Target T2
Time 0: mapping with T0 is started	A B C D E F	___ ^a ___	Not started	Not started
Time 1: change of “A” is made on source (->“G”)	G B C D E F	A __ ^a ___	Not started	Not started
Time 2: mapping with T1 is started	G B C D E F	A __ ___	___ ^a ___	Not started
Time 3: change of “E” is made on source (->“H”)	G B C D H F	A __ ___	___ ^a _ E _	Not started
Time 4: mapping with T2 is started	G B C D H F	A __ ___	___ _ E _	___ ^a ___
Time 5: change of “F” is made on source (->“I”)	G B C D H I	A __ ___	___ _ E _	___ ^a __ F
Time 6: change of “G” is made on source (->“J”)	J B C D H I	A __ ___	___ _ E _	G __ ^a __ F
Time 7: stop of Source-T2 mapping	J B C D H I	A __ ___	G __ ^a _ E F	Stopped
Time 8: stop of Source-T1 mapping	J B C D H I	A __ ^a _ E F	Stopped	Stopped

a. “Most recent” target

An intermediate target disk (not the oldest or the newest) treats the set of newer target volumes and the true source volume as a type of composite source. It treats all older volumes as a kind of target (and behaves like a source to them).

Target read with multiple target FlashCopy

Target reading with multiple targets depends on whether the grain was copied. Consider the following points:

- ▶ If the grain that is read is copied from the source to the target, the read returns data from the target that is read.
- ▶ If the grain is not yet copied, each of the newer mappings is examined in turn. The read is performed from the first copy (the oldest) that is found. If none is found, the read is performed from the source.

For example, in Figure 10-13 on page 572, if the yellow grain on T2 is read, it returns “H” because no newer target than T2 exists. Therefore, the source is read.

As another example, in Figure 10-13 on page 572, if the red grain on T0 is read, it returns “E” because two newer targets exist for T0, and T1 is the oldest of those targets.

Target write with multiple target FlashCopy (Copy on Demand)

A write to an intermediate or the newest target volume must consider the state of the grain within its own mapping and the state of the grain of the next oldest mapping. Consider the following points:

- ▶ If the grain in the target that is written is copied and if the grain of the next oldest mapping is not yet copied, the grain must be copied before the write can proceed to preserve the contents of the next oldest mapping.
 For example, in Figure 10-13 on page 572, if the grain “G” is changed on T2, it must be copied to T1 (next oldest not yet copied) first and then changed on T2.
- ▶ If the grain in the target that is being written is not yet copied, the grain is copied from the oldest copied grain in the mappings that are newer than the target, or from the source if none is copied. For example, in Figure 10-13 on page 572, if the red grain on T0 is written, it is first copied from T1 (data “E”). After this copy is done, the write can be applied to the target.

Table 10-4 lists the indirection layer algorithm in a multi-target FlashCopy.

Table 10-4 Summary table of the FlashCopy indirection layer algorithm

Accessed volume	Was the grain copied?	Host I/O operation	
		Read	Write
Source	No	Read from the source volume.	Copy grain to most recently started target for this source, then write to the source.
	Yes	Read from the source volume.	Write to the source volume.
Target	No	If any newer targets exist for this source in which this grain was copied, read from the oldest of these targets. Otherwise, read from the source.	Hold the write. Check the dependency target volumes to see whether the grain was copied. If the grain is not copied to the next oldest target for this source, copy the grain to the next oldest target. Then, write to the target.
	Yes	Read from the target volume.	Write to the target volume.

Stopping process in a multiple target FlashCopy: Cleaning Mode

When a mapping that contains a target that includes dependent mappings is stopped, the mapping enters the stopping state. It then begins copying all grains that are uniquely held on the target volume of the mapping that is being stopped to the next oldest mapping that is in the copying state. The mapping remains in the stopping state until all grains are copied, and then enters the stopped state. This mode is referred to as the *Cleaning Mode*.

For example, if the mapping Source-T2 was stopped, the mapping enters the stopping state while the cleaning process copies the data of T2 to T1 (next oldest). After all of the data is copied, Source-T2 mapping enters the stopped state, and T1 is no longer dependent upon T2. However, T0 remains dependent upon T1.

For example, as shown in Table 10-3 on page 573, if you stop the Source-T2 mapping on “Time 7,” then the grains that are not yet copied on T1 are copied from T2 to T1. Reading T1 is then like reading the source at the time T1 was started (“Time 2”).

As another example, with Table 10-3 on page 573, if you stop the Source-T1 mapping on “Time 8,” the grains that are not yet copied on T0 are copied from T1 to T0. Reading T0 is then similar to reading the source at the time T0 was started (“Time 0”).

If you stop the Source-T1 mapping while Source-T0 mapping and Source-T2 are still in copying mode, the grains that are not yet copied on T0 are copied from T1 to T0 (next oldest). T0 now depends upon T2.

Your target volume is still accessible while the cleaning process is running. When the system is operating in this mode, it is possible that host I/O operations can prevent the cleaning process from reaching 100% if the I/O operations continue to copy new data to the target volumes.

Cleaning rate

The data rate at which data is copied from the target of the mapping being stopped to the next oldest target is determined by the *cleaning rate*. This property of FlashCopy mapping can be changed dynamically. It is measured as is the copyrate property, but both properties are independent. Table 10-5 lists the relationship of the cleaning rate values to the attempted number of grains to be split per second.

Table 10-5 Cleaning rate values

User-specified copy rate attribute value	Data copied/sec	256 KB grains/sec	64 KB grains/sec
1 - 10	128 KiB	0.5	2
11 - 20	256 KiB	1	4
21 - 30	512 KiB	2	8
31 - 40	1 MiB	4	16
41 - 50	2 MiB	8	32
51 - 60	4 MiB	16	64
61 - 70	8 MiB	32	128
71 - 80	16 MiB	64	256
81 - 90	32 MiB	128	512
91 - 100	64 MiB	256	1024
101 - 110	128 MiB	512	2048
111 - 120	256 MiB	1024	4096
121 - 130	512 MiB	2048	8192
131 - 140	1 GiB	4096	16384
141 - 150	2 GiB	8192	32768

10.1.12 Reverse FlashCopy

Reverse FlashCopy enables FlashCopy targets to become restore points for the source without breaking the FlashCopy mapping, and without having to wait for the original copy operation to complete. A FlashCopy source supports multiple targets (up to 256) and multiple rollback points.

A key advantage of the IBM Spectrum Virtualize Multiple Target Reverse FlashCopy function is that the reverse FlashCopy does not delete the original target. This feature enables processes that use the target, such as a tape backup or tests, to continue uninterrupted.

IBM Spectrum Virtualize also can create an optional copy of the source volume to be made before the reverse copy operation starts. This ability to restore back to the original source data can be useful for diagnostic purposes.

The production disk is instantly available with the backup data. Figure 10-14 shows an example of Reverse FlashCopy with a simple FlashCopy mapping (single target).

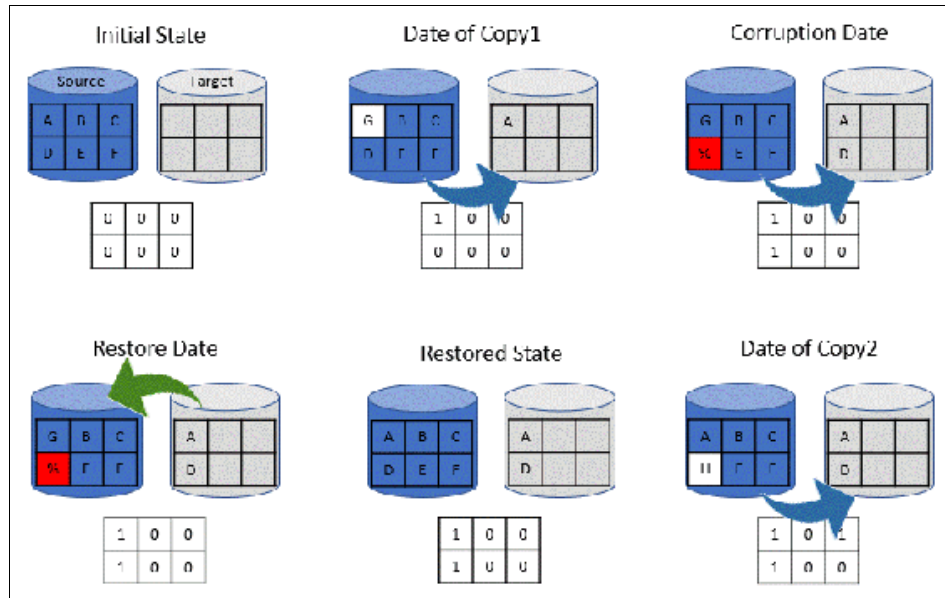


Figure 10-14 A reverse FlashCopy example for data restoration

This example assumes that a simple FlashCopy mapping was created between the “source” volume and “target” volume, and no background copy is set.

When the FlashCopy mapping starts (Date of Copy1), if source volume is changed (write operations on grain “A”), the modified grains are first copied to target, the bitmap table is updated, and the source grain is modified (from “A” to “G”).

At a specific time (“Corruption Date”), data is modified on another grain (grain “D” below), so it is first written on the target volume and the bitmap table is updated. Unfortunately, the new data is corrupted on source volume.

The storage administrator can then use the Reverse FlashCopy feature by completing the following steps:

1. Create a mapping from target to source (if not already created). Because FlashCopy recognizes that the target volume of this new mapping is a source in another mapping, it does not create another bitmap table. It uses the existing bitmap table instead, with its updated bits.
2. Start the new mapping. Because of the existing bitmap table, only the *modified* grains are copied.

After the restoration is complete, at the “Restored State” time, source volume data is similar to what it was before the Corruption Date. The copy can resume with the restored data (Date of Copy2) and, for example, data on the source volume can be modified (“D” grain is changed in “H” grain in the example below). In this last case, because “D” grain was copied, it is not copied again on target volume.

Consistency groups are reversed by creating a set of reverse FlashCopy mappings and adding them to a new reverse consistency group. Consistency groups cannot contain more than one FlashCopy mapping with the same target volume.

10.1.13 FlashCopy and image mode volumes

FlashCopy can be used with image mode volumes. Because the source and target volumes must be the same size, you must create a target volume with the same size as the image mode volume when you are creating a FlashCopy mapping. To accomplish this task by using the CLI, run the `svcinfo lsvdisk -bytes volumename` command. The size in bytes is then used to create the volume that is used in the FlashCopy mapping.

This method provides an exact number of bytes because image mode volumes might not line up one-to-one on other measurement unit boundaries. Example 10-1 shows the size of the ITS0-RS-TST volume. The ITS0-TST01 volume is then created, which specifies the same size.

Example 10-1 Listing the size of a volume in bytes and creating a volume of equal size

```
IBM_2145:ITS0-SV1:superuser>lsvdisk -bytes ITS0-RS-TST
id 42
name ITS0-RS-TST
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name Pool0
capacity 21474836480
type striped
formatted no
formatting yes
mdisk_id
mdisk_name
FC_id
.....

IBM_2145:ITS0-SV1:superuser>mkvdisk -mdiskgrp Pool0 -iogrp 0 -size 21474836480
-unit b -name ITS0-TST01
Virtual Disk, id [43], successfully created
IBM_2145:ITS0-SV1:superuser>

IBM_2145:ITS0-SV1:superuser>lsvdisk -delim " "
42 ITS0-RS-TST 0 io_grp0 online 0 Pool0 20.00GB striped
600507680C9B8000480000000000002C 0 1 not_empty 0 no 0 0 Pool0 yes no 42
ITS0-RS-TST
43 ITS0-TST01 0 io_grp0 online 0 Pool0 20.00GB image
600507680C9B8000480000000000002D 0 1 not_empty 0 no 0 0 Pool0 yes no 43 ITS0-TST01
IBM_2145:ITS0-SV1:superuser>
```

Tip: Alternatively, you can run the `expandvdisksize` and `shrinkvdisksize` volume commands to modify the size of the volume.

These actions must be performed before a mapping is created.

10.1.14 FlashCopy mapping events

This section describes the events that modify the states of a FlashCopy. It also describes the mapping events that are listed in Table 10-6.

Overview of a FlashCopy sequence of events: The FlashCopy sequence includes the following tasks:

1. Associate the source data set with a target location (one or more source and target volumes).
2. Create a FlashCopy mapping for each source volume to the corresponding target volume. The target volume must be equal in size to the source volume.
3. Discontinue access to the target (application dependent).
4. Prepare (pre-trigger) the FlashCopy:
 - a. Flush the cache for the source.
 - b. Discard the cache for the target.
5. Start (trigger) the FlashCopy:
 - a. Pause I/O (briefly) on the source.
 - b. Resume I/O on the source.
 - c. Start I/O on the target.

Table 10-6 Mapping events

Mapping event	Description
Create	A FlashCopy mapping is created between the specified source volume and the specified target volume. The operation fails if any one of the following conditions is true: <ul style="list-style-type: none"> ▶ The source volume is a member of 256 FlashCopy mappings. ▶ The node has insufficient bitmap memory. ▶ The source and target volumes are different sizes.
Prepare	The <code>prestartfcmap</code> or <code>prestartfcconsistgrp</code> command is directed to a consistency group for FlashCopy mappings that are members of a normal consistency group or to the mapping name for FlashCopy mappings that are stand-alone mappings. The <code>prestartfcmap</code> or <code>prestartfcconsistgrp</code> command places the FlashCopy mapping into the Preparing state. The <code>prestartfcmap</code> or <code>prestartfcconsistgrp</code> command can corrupt any data that was on the target volume because cached writes are discarded. Even if the FlashCopy mapping is never started, the data from the target might be changed logically during the act of preparing to start the FlashCopy mapping.
Flush done	The FlashCopy mapping automatically moves from the preparing state to the prepared state after all cached data for the source is flushed and all cached data for the target is no longer valid.

Mapping event	Description
Start	<p>When all of the FlashCopy mappings in a consistency group are in the prepared state, the FlashCopy mappings can be started. To preserve the cross-volume consistency group, the start of all of the FlashCopy mappings in the consistency group must be synchronized correctly concerning I/Os that are directed at the volumes by running the startfcmap or startfcconsistgrp command.</p> <p>The following actions occur during the running of the startfcmap command or the startfcconsistgrp command:</p> <ul style="list-style-type: none"> ▶ New reads and writes to all source volumes in the consistency group are paused in the cache layer until all ongoing reads and writes beneath the cache layer are completed. ▶ After all FlashCopy mappings in the consistency group are paused, the internal cluster state is set to enable FlashCopy operations. ▶ After the cluster state is set for all FlashCopy mappings in the consistency group, read and write operations continue on the source volumes. ▶ The target volumes are brought online. <p>As part of the startfcmap or startfcconsistgrp command, read and write caching is enabled for the source and target volumes.</p>
Modify	<p>The following FlashCopy mapping properties can be modified:</p> <ul style="list-style-type: none"> ▶ FlashCopy mapping name. ▶ Clean rate. ▶ Consistency group. ▶ Copy rate (for background copy or stopping copy priority). ▶ Automatic deletion of the mapping when the background copy is complete.
Stop	<p>The following separate mechanisms can be used to stop a FlashCopy mapping:</p> <ul style="list-style-type: none"> ▶ Issue a command. ▶ An I/O error occurred.
Delete	<p>This command requests that the specified FlashCopy mapping is deleted. If the FlashCopy mapping is in the copying state, the force flag must be used.</p>
Flush failed	<p>If the flush of data from the cache cannot be completed, the FlashCopy mapping enters the stopped state.</p>
Copy complete	<p>After all of the source data is copied to the target and there are no dependent mappings, the state is set to copied. If the option to automatically delete the mapping after the background copy completes is specified, the FlashCopy mapping is deleted automatically. If this option is not specified, the FlashCopy mapping is not deleted automatically and can be reactivated by preparing and starting again.</p>
Bitmap online/offline	<p>The node failed.</p>

10.1.15 Thin-provisioned FlashCopy

FlashCopy source and target volumes can be thin-provisioned.

Source or target thin-provisioned

The most common configuration is a fully allocated source and a thin-provisioned target. By using this configuration, the target uses a smaller amount of real storage than the source.

With this configuration, use a copyrate equal to 0 only. In this state, the virtual capacity of the target volume is identical to the capacity of the source volume, but the real capacity (the one used on the storage system) is lower, as shown on Figure 10-15. The real size of the target volume increases with writes that are performed on the source volume, on not already copied grains. Eventually, if the entire source volume is written (unlikely), the real capacity of the target volume is identical to the source's volume.

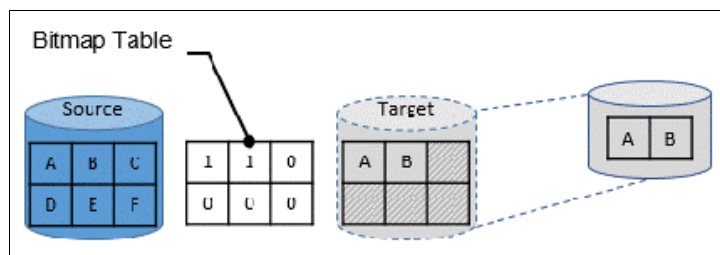


Figure 10-15 Thin-provisioned target volume

Source and target thin-provisioned

When the source and target volumes are thin-provisioned, only the data that is allocated to the source is copied to the target. In this configuration, the background copy option has no effect.

Performance: The best performance is obtained when the grain size of the thin-provisioned volume is the same as the grain size of the FlashCopy mapping.

Thin-provisioned incremental FlashCopy

The implementation of thin-provisioned volumes does not preclude the use of incremental FlashCopy on the same volumes. It does not make sense to have a fully allocated source volume and then use incremental FlashCopy (which is always a full copy the first time) to copy this fully allocated source volume to a thin-provisioned target volume. However, this action is not prohibited.

Consider the following optional configurations:

- ▶ A thin-provisioned source volume can be copied incrementally by using FlashCopy to a thin-provisioned target volume. Whenever the FlashCopy is performed, only data that was modified is recopied to the target. If space is allocated on the target because of I/O to the target volume, this space is not reclaimed with subsequent FlashCopy operations.
- ▶ A fully allocated source volume can be copied incrementally by using FlashCopy to another fully allocated volume at the same time as it is being copied to multiple thin-provisioned targets (taken at separate points in time). By using this combination, a single full backup can be kept for recovery purposes, and the backup workload is separated from the production workload. At the same time, older thin-provisioned backups can be retained.

10.1.16 Serialization of I/O by FlashCopy

In general, the FlashCopy function in the IBM Spectrum Virtualize introduces no explicit serialization into the I/O path. Therefore, many concurrent I/Os are allowed to the source and target volumes.

However, a lock exists for each grain and this lock can be in shared or exclusive mode. For multiple targets, a common lock is shared, and the mappings are derived from a particular source volume. The lock is used in the following modes under the following conditions:

- ▶ The lock is held in shared mode during a read from the target volume, which touches a grain that was not copied from the source.
- ▶ The lock is held in exclusive mode while a grain is being copied from the source to the target.

If the lock is held in shared mode and another process wants to use the lock in shared mode, this request is granted unless a process is waiting to use the lock in exclusive mode.

If the lock is held in shared mode and it is requested to be exclusive, the requesting process must wait until all holders of the shared lock free it.

Similarly, if the lock is held in exclusive mode, a process wanting to use the lock in shared or exclusive mode must wait for it to be freed.

10.1.17 Event handling

When a FlashCopy mapping is not copying or stopping, the FlashCopy function does not affect the handling or reporting of events for error conditions that are encountered in the I/O path. Event handling and reporting are affected only by FlashCopy when a FlashCopy mapping is copying or stopping, that is, actively moving data.

Node failure

Normally, two copies of the FlashCopy bitmap are maintained. One copy of the FlashCopy bitmap is on each of the two nodes that make up the I/O group of the source volume. When a node fails, one copy of the bitmap for all FlashCopy mappings whose source volume is a member of the failing node's I/O group becomes inaccessible.

FlashCopy continues with a single copy of the FlashCopy bitmap that is stored as non-volatile in the remaining node in the source I/O group. The system metadata is updated to indicate that the missing node no longer holds a current bitmap. When the failing node recovers or a replacement node is added to the I/O group, the bitmap redundancy is restored.

Path failure (Path Offline state)

In a fully functioning system, all of the nodes have a software representation of every volume in the system within their application hierarchy.

Because the storage area network (SAN) that links IBM Spectrum Virtualize nodes to each other and to the MDisks is made up of many independent links, it is possible for a subset of the nodes to be temporarily isolated from several of the MDisks. When this situation occurs, the MDisks are said to be *Path Offline* on certain nodes.

Other nodes: Other nodes might see the MDisks as Online because their connection to the MDisks still exists.

Path Offline for the source Volume

If a FlashCopy mapping is in the `copying` state and the source volume goes path offline, this path offline state is propagated to all target volumes up to, but not including, the target volume for the newest mapping that is 100% copied but remains in the `copying` state. If no mappings are 100% copied, all of the target volumes are taken offline. `Path offline` is a state that exists on a per-node basis. Other nodes might not be affected. If the source volume comes online, the target and source volumes are brought back online.

Path Offline for the target Volume

If a target volume goes path offline but the source volume is still online and if any dependent mappings exist, those target volumes also go path offline. The source volume remains online.

10.1.18 Asynchronous notifications

FlashCopy raises informational event log entries for certain mapping and consistency group state transitions. These state transitions occur as a result of configuration events that complete asynchronously. The informational events can be used to generate Simple Network Management Protocol (SNMP) traps to notify the user.

Other configuration events complete synchronously, and no informational events are logged as a result of the following events:

▶ `PREPARE_COMPLETED`

This state transition is logged when the FlashCopy mapping or consistency group enters the prepared state as a result of a user request to prepare. The user can now start (or stop) the mapping or consistency group.

▶ `COPY_COMPLETED`

This state transition is logged when the FlashCopy mapping or consistency group enters the `idle_or_copied` state when it was in the `copying` or `stopping` state. This state transition indicates that the target disk now contains a complete copy and no longer depends on the source.

▶ `STOP_COMPLETED`

This state transition is logged when the FlashCopy mapping or consistency group enters the stopped state as a result of a user request to stop. It is logged after the automatic copy process completes. This state transition includes mappings where no copying needed to be performed. This state transition differs from the event that is logged when a mapping or group enters the stopped state as a result of an I/O error.

10.1.19 Interoperation with Metro Mirror and Global Mirror

A volume can be part of any copy relationship; that is, FlashCopy, Metro Mirror (MM), or Remote Mirror. Therefore, FlashCopy can work with MM/Global Mirror (GM) to provide better protection of the data.

For example, you can perform an MM copy to duplicate data from Site_A to Site_B, and then perform a daily FlashCopy to back up the data to another location.

Note: A volume cannot be part of FlashCopy, MM, or Remote Mirror, if it is set to Transparent Cloud Tiering (TCT) function.

Table 10-7 on page 583 lists the supported combinations of FlashCopy and Remote Copy (RC). In the table, *RC* refers to MM and GM.

Table 10-7 FlashCopy and remote copy interaction

Component	RC primary site	RC secondary site
FlashCopy source	Supported.	Supported latency: When the FlashCopy relationship is in the preparing and prepared states, the cache at the RC secondary site operates in write-through mode. This process adds latency to the latent RC relationship.
FlashCopy target	This combination is supported and has the following restrictions: <ul style="list-style-type: none"> ▶ Running a stop -force might cause the RC relationship to be fully resynchronized. ▶ Code level must be 6.2.x or later. ▶ I/O group must be the same. 	This combination is supported with the major restriction that the FlashCopy mapping cannot be copying, stopping, or suspended. Otherwise, the restrictions are the same as at the RC primary site.

10.1.20 FlashCopy attributes and limitations

The FlashCopy function in IBM Spectrum Virtualize features the following attributes:

- ▶ The target is the T0 copy of the source, which is known as *FlashCopy mapping target*.
- ▶ FlashCopy produces an exact copy of the source volume, including any metadata that was written by the host OS, logical volume manager (LVM), and applications.
- ▶ The source volume and target volume are available (almost) immediately following the FlashCopy operation.
- ▶ The source and target volumes:
 - Must be the same “virtual” size.
 - Must be on the same IBM Spectrum Virtualize system.
 - Do not need to be in the same I/O group or storage pool, although it is recommended for them to have the same preferred node for best performance.
- ▶ The storage pool extent sizes can differ between the source and target.
- ▶ The target volumes can be the source volumes for other FlashCopy mappings (*cascaded FlashCopy*). However, a target volume can have only one source copy.
- ▶ Consistency groups are supported to enable FlashCopy across multiple volumes at the same time.
- ▶ The target volume can be updated independently of the source volume.
- ▶ Bitmaps that are governing I/O redirection (I/O indirection layer) are maintained in both nodes of the IBM Spectrum Virtualize I/O group to prevent a single point of failure (SPOF).
- ▶ FlashCopy mapping and consistency groups can be automatically withdrawn after the completion of the background copy.
- ▶ Thin-provisioned FlashCopy (or Snapshot in the GUI) use disk space only when updates are made to the source or target data, and not for the entire capacity of a volume copy.
- ▶ FlashCopy licensing is based on the virtual capacity of the source volumes.

- ▶ Incremental FlashCopy copies all of the data when you first start FlashCopy, and then only the changes when you stop and start FlashCopy mapping again. Incremental FlashCopy can substantially reduce the time that is required to re-create an independent image.
- ▶ Reverse FlashCopy enables FlashCopy targets to become restore points for the source without breaking the FlashCopy relationship, and without having to wait for the original copy operation to complete.
- ▶ The size of the source and target volumes cannot be altered (increased or decreased) while a FlashCopy mapping is defined.

The IBM FlashCopy limitations for IBM Spectrum Virtualize V8.4 are listed in Table 10-8.

Table 10-8 FlashCopy limitations in Version 8.4

Property	Maximum number
FlashCopy mappings per system	10000
FlashCopy targets per source	256
FlashCopy mappings per consistency group	512
FlashCopy consistency groups per system	500
Total FlashCopy volume capacity per I/O group	4096 TiB

10.2 Managing FlashCopy by using the GUI

It is often easier to work with the FlashCopy function from the GUI if you have a reasonable number of host mappings. However, in enterprise data centers with many host mappings, use the CLI to run your FlashCopy commands.

10.2.1 FlashCopy presets

The IBM Spectrum Virtualize GUI interface provides three FlashCopy presets (Snapshot, Clone, and Backup) to simplify the more common FlashCopy operations.

Although these presets meet most FlashCopy requirements, they do not support all possible FlashCopy options. If more specialized options are required that are not supported by the presets, the options must be performed by using CLI commands.

This section describes the preset options and their use cases.

Snapshot

This preset creates a PiT copy that tracks only the changes that are made, either at the source or target volumes. The snapshot is not intended to be an independent copy. Instead, the copy is used to maintain a view of the production data at the time that the snapshot is created. Therefore, the snapshot holds only the data from regions of the production volume that changed since the snapshot was created. Because the snapshot preset uses thin provisioning, only the capacity that is required for the changes is used.

Snapshot uses the following preset parameters:

- ▶ Background copy: None
- ▶ Incremental: No
- ▶ Delete after completion: No

- ▶ Cleaning rate: No
- ▶ Primary copy source pool: Target pool

Use case

The user wants to produce a copy of a volume without affecting the availability of the volume. The user does not anticipate many changes to be made to the source or target volume; a significant proportion of the volumes remains unchanged.

By ensuring that only changes require a copy of data to be made, the total amount of disk space that is required for the copy is reduced. Therefore, many Snapshot copies can be used in the environment.

Snapshots are useful for providing protection against corruption or similar issues with the validity of the data, but they do not provide protection from physical controller failures. Snapshots can also provide a vehicle for performing repeatable testing (including “what-if” modeling that is based on production data) without requiring a full copy of the data to be provisioned.

For example, in Figure 10-16, the source volume user can still work on the original data volume (as with a production volume) and the target volumes can be accessed instantly. Users of target volumes can modify the content and perform “what-if” tests; for example, versioning. Storage administrators do not need to perform full copies of a volume for temporary tests. However, the target volumes must remain linked to the source. When the link is broken (FlashCopy mapping stopped or deleted), the target volumes become unusable.

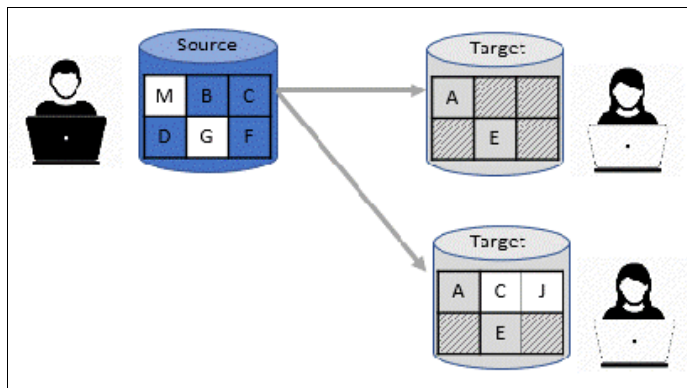


Figure 10-16 FlashCopy snapshot preset example

Clone

The clone preset creates a replica of the volume, which can be changed without affecting the original volume. After the copy completes, the mapping that was created by the preset is automatically deleted.

Clone uses the following preset parameters:

- ▶ Background copy rate: 50
- ▶ Incremental: No
- ▶ Delete after completion: Yes
- ▶ Cleaning rate: 50
- ▶ Primary copy source pool: Target pool

Use case

Users want a copy of the volume that they can modify without affecting the original volume. After the clone is established, it is not expected that it is refreshed or that the original production data must be referenced again. If the source is thin-provisioned, the target is thin-provisioned for the auto-create target.

Backup

The backup preset creates an incremental PiT replica of the production data. After the copy completes, the backup view can be refreshed from the production data, with minimal copying of data from the production volume to the backup volume.

Backup uses the following preset parameters:

- ▶ Background Copy rate: 50
- ▶ Incremental: Yes
- ▶ Delete after completion: No
- ▶ Cleaning rate: 50
- ▶ Primary copy source pool: Target pool

Use case

The user wants to create a copy of the volume that can be used as a backup if the source becomes unavailable, such as because of loss of the underlying physical controller. The user plans to periodically update the secondary copy, and does not want to suffer from the resource demands of creating a copy each time.

Incremental FlashCopy times are faster than full copy, which helps to reduce the window where the new backup is not yet fully effective. If the source is thin-provisioned, the target is also thin-provisioned in this option for the auto-create target.

Another use case, which is not supported by the name, is to create and maintain (periodically refresh) an independent image that can be subjected to intensive I/O (for example, data mining) without affecting the source volume's performance.

Note: IBM Spectrum Virtualize in general and FlashCopy in particular are not backup solutions on their own. For example, a FlashCopy backup preset does not schedule a regular copy of your volumes. Instead, it overwrites the mapping target and does not make a copy of it before starting a new “backup” operation. It is the user's responsibility to handle the target volumes (for example, saving them to tapes) and the scheduling of the FlashCopy operations.

10.2.2 FlashCopy window

This section describes the tasks that you can perform at a FlashCopy level by using the IBM Spectrum Virtualize GUI.

When the IBM Spectrum Virtualize GUI is used, FlashCopy components can be seen in different windows. Three windows are related to FlashCopy and are available by using the **Copy Services** menu, as shown in Figure 10-17 on page 587.

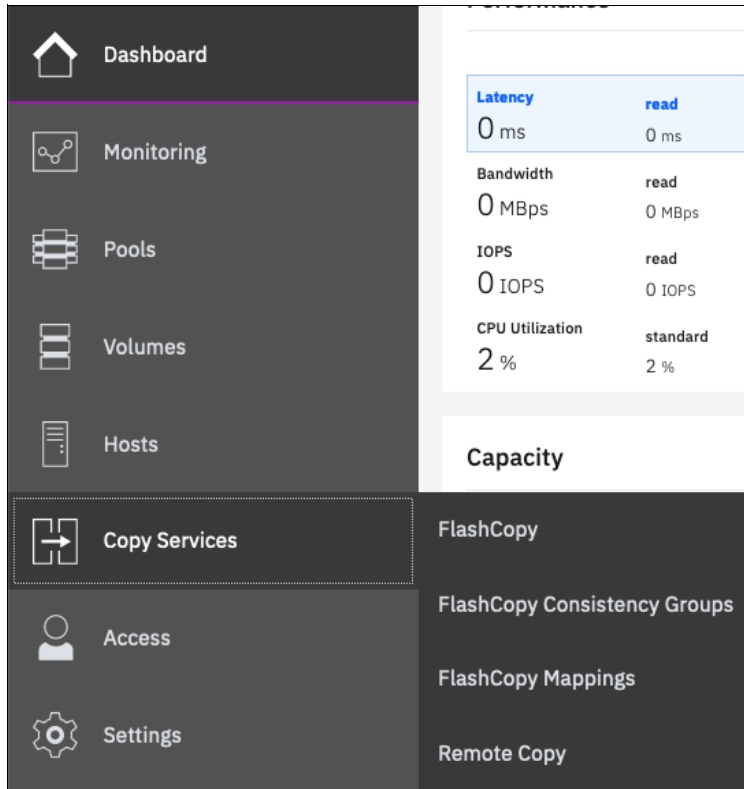


Figure 10-17 Copy Services menu

The FlashCopy window is accessible by clicking **Copy Services** → **FlashCopy**. It displays all of the volumes that are defined in the system. Volumes that are part of a FlashCopy mapping appear, as shown in Figure 10-18. By clicking a source volume, you can display the list of its target volumes.

Volume Name	Status	Progress	Capacity	Group	Flash Time	
ITSO-FC-VOL-01			10.00 GiB			
ITSO-FC-VOL-01_03	✓ Copied	100%			Oct 22, 2019, 3:21:06 PM	
ITSO-FC-VOL-01_05	⌛ Copying	0%			Oct 22, 2019, 3:21:21 PM	
ITSO-FC-VOL-01_04	⌛ Copying	0%			Oct 22, 2019, 3:21:16 PM	
ITSO-FC-VOL-01_01	✓ Copied	100%		fccstgrp1	Oct 18, 2019, 2:20:11 PM	

Figure 10-18 Source and target volumes displayed in the FlashCopy window

All volumes are listed in this window, and target volumes appear twice (as a regular volume and as a target volume in a FlashCopy mapping).

Consider the following points:

- ▶ The Consistency Group window is accessible by clicking **Copy Services** → **Consistency Groups**. Use the Consistency Groups window (as shown in Figure 10-19) to list the FlashCopy mappings that are part of consistency groups and part of no consistency groups.

Mapping Name	↑	Status	Source Volume	Target Volume	Progress	Flash Tim
∨ Not in a Group						
fcmmap6		✓ Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_03	100%	Oct 22,
fcmmap7		⌛ Copying	ITSO-FC-VOL-01	ITSO-FC-VOL-01_04	0%	Oct 22,
fcmmap8		⌛ Copying	ITSO-FC-VOL-01	ITSO-FC-VOL-01_05	0%	Oct 22,
∨ fccstgrp1 Idle or Copied						
fcmmap0		✓ Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_01	100%	Oct 18,
fcmmap1		✓ Copied	ITSO-FC-VOL-04	ITSO-FC-VOL-04_01	100%	Oct 18,
fcmmap2		✓ Copied	ITSO-FC-VOL-03	ITSO-FC-VOL-03_01	100%	Oct 18,
fcmmap3		✓ Copied	ITSO-FC-VOL-05	ITSO-FC-VOL-05_01	100%	Oct 18,
fcmmap4		✓ Copied	ITSO-FC-VOL-02	ITSO-FC-VOL-02_01	100%	Oct 18,

Figure 10-19 Consistency Groups window

- ▶ The FlashCopy Mappings window is accessible by clicking **Copy Services** → **FlashCopy Mappings**. Use the FlashCopy Mappings window (as shown in Figure 10-20) to display the list of mappings between source volumes and target volumes.

Mapping Name	↑	Status	Source Volume	Target Volume	Progress	Group	Flash Tim
fcmmap0		✓ Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_01	100%	fccstgrp1	Oct 18, 2019, 2:
fcmmap1		✓ Copied	ITSO-FC-VOL-04	ITSO-FC-VOL-04_01	100%	fccstgrp1	Oct 18, 2019, 2:
fcmmap2		✓ Copied	ITSO-FC-VOL-03	ITSO-FC-VOL-03_01	100%	fccstgrp1	Oct 18, 2019, 2:
fcmmap3		✓ Copied	ITSO-FC-VOL-05	ITSO-FC-VOL-05_01	100%	fccstgrp1	Oct 18, 2019, 2:
fcmmap4		✓ Copied	ITSO-FC-VOL-02	ITSO-FC-VOL-02_01	100%	fccstgrp1	Oct 18, 2019, 2:
fcmmap6		✓ Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_03	100%		Oct 22, 2019, 3:
fcmmap7		⌛ Copying	ITSO-FC-VOL-01	ITSO-FC-VOL-01_04	0%		Oct 22, 2019, 3:
fcmmap8		⌛ Copying	ITSO-FC-VOL-01	ITSO-FC-VOL-01_05	0%		Oct 22, 2019, 3:

Showing 8 FC mappings | Selecting 0 FC mappings

Figure 10-20 FlashCopy mapping window

10.2.3 Creating a FlashCopy mapping

This section describes creating FlashCopy mappings for volumes and their targets.

Open the FlashCopy window from the **Copy Services** menu, as shown in Figure 10-21. Select the volume for which you want to create the FlashCopy mapping. Right-click the volume or click the **Actions** menu.

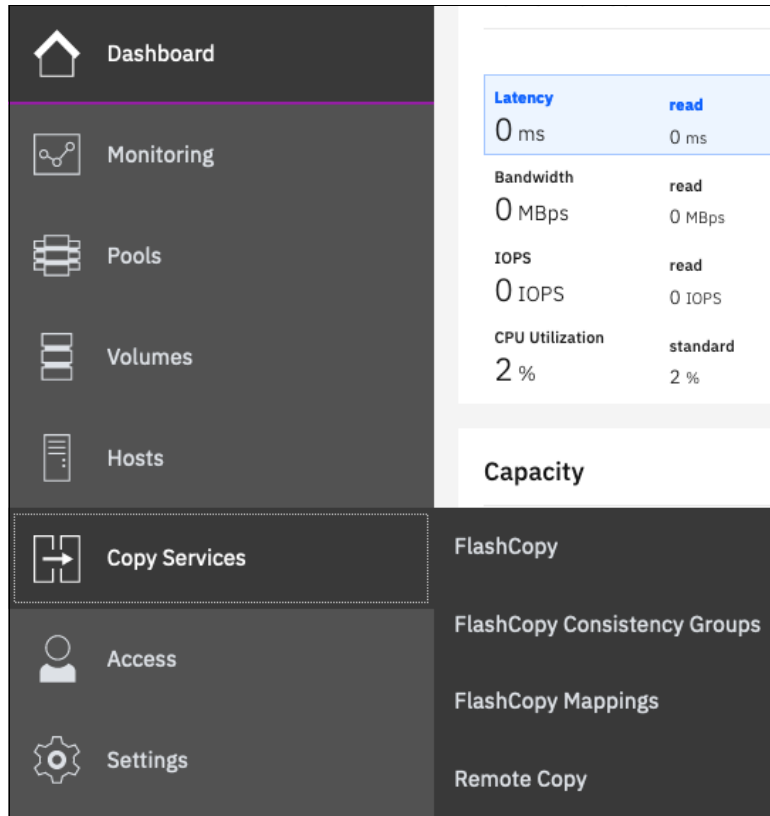


Figure 10-21 FlashCopy window

Multiple FlashCopy mappings: To create multiple FlashCopy mappings at the same time, select multiple volumes by pressing and holding **Ctrl** and clicking the entries that you want.

Depending on whether you created the target volumes for your FlashCopy mappings or you want the system to create the target volumes for you, the following options are available:

- ▶ If you created the target volumes, see “Creating a FlashCopy mapping with existing target Volumes” on page 590.
- ▶ If you want the system to create the target volumes for you, see “Creating a FlashCopy mapping and target volumes” on page 595.

Creating a FlashCopy mapping with existing target Volumes

Complete the following steps to use existing target volumes for the FlashCopy mappings.

Attention: When starting a FlashCopy mapping from a source volume to a target volume, data that is on the target is over-written. The system does not prevent you from selecting a target volume that is mapped to a host and contains data.

1. Right-click the volume that you want to create a FlashCopy mapping for, and select **Advanced FlashCopy** → **Use Existing Target Volumes**, as shown in Figure 10-22.

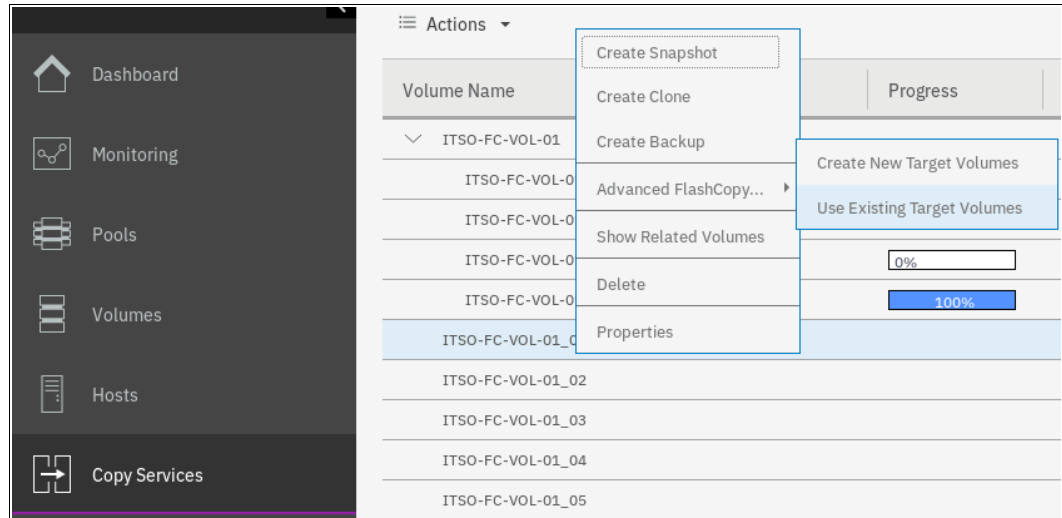


Figure 10-22 Creating a FlashCopy mapping with an existing target

The Create FlashCopy Mapping window opens, as shown in Figure 10-23 on page 591.

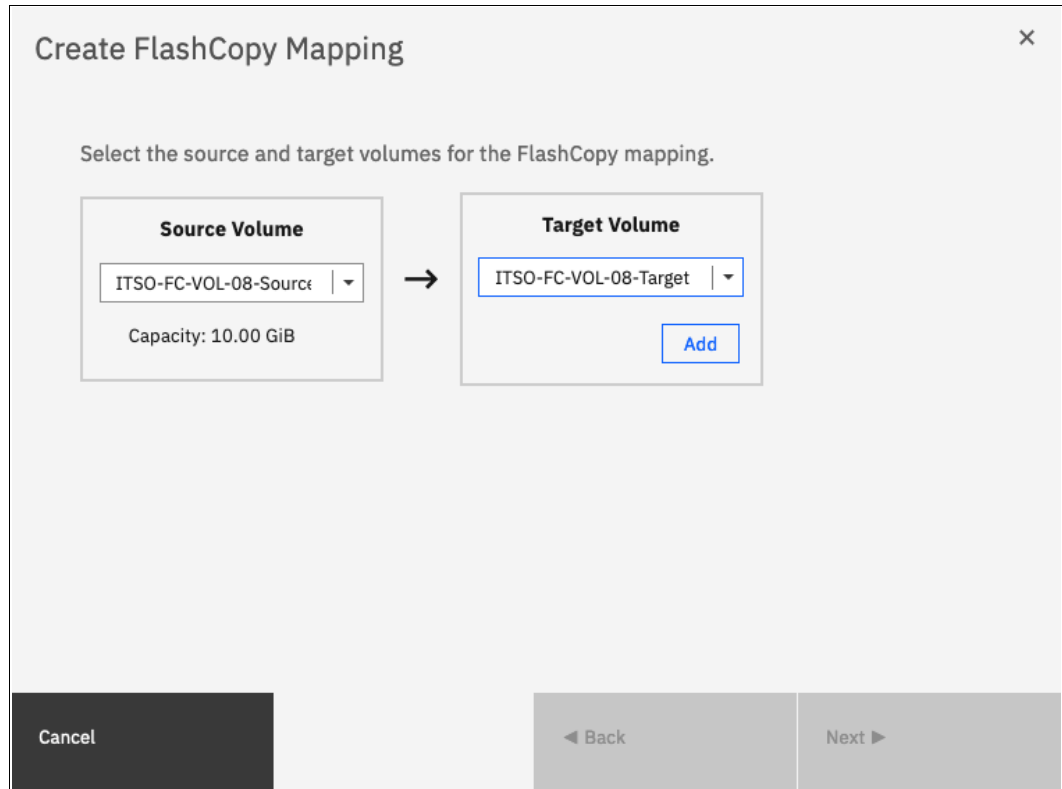


Figure 10-23 Selecting source and target for a FlashCopy mapping

In this window, you create the mapping between the selected source volume and the target volume you want to create a mapping with. Then, click **Add**.

Important: The source volume and the target volume must be of equal size. Therefore, only targets of the same size are shown in the list for a source volume.

Volumes that are a target in a FlashCopy mapping cannot be a target in a new mapping. Therefore, only volumes that are not targets can be selected.

To remove a mapping that was created, click **X** (see Figure 10-24). Otherwise, click **Next** after you create all of the mappings that you need.

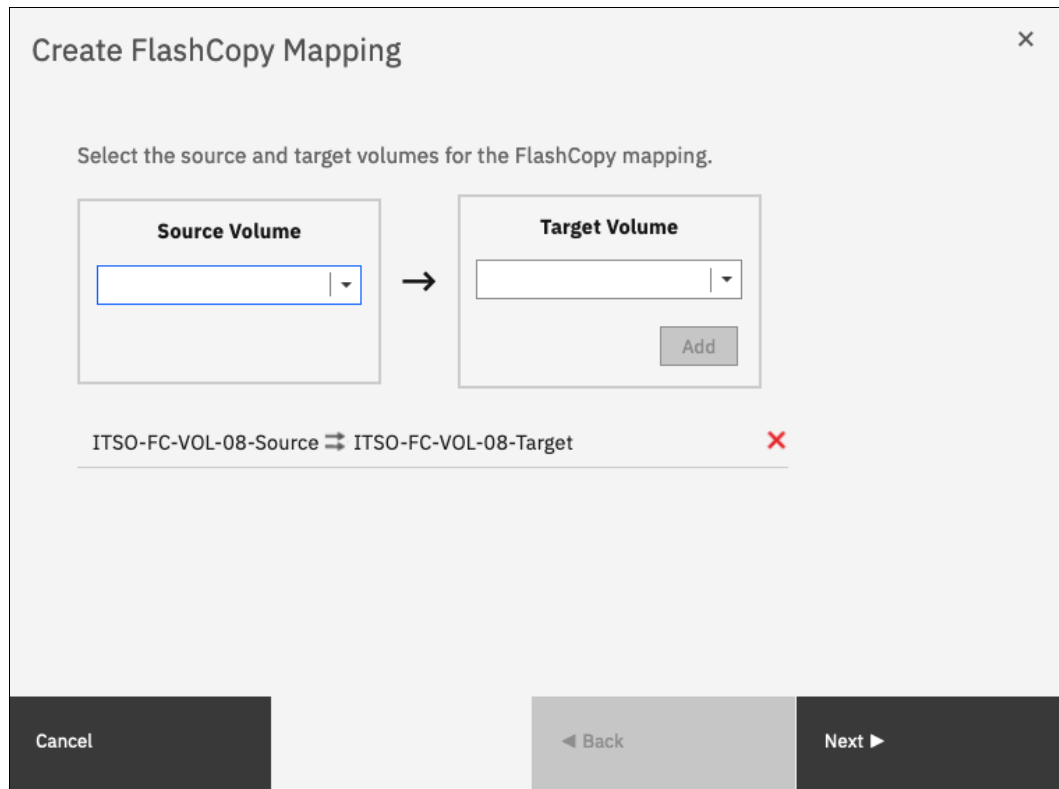


Figure 10-24 Viewing the source and target at creation time

2. In the next window, select one FlashCopy preset. The GUI provides the following presets to simplify common FlashCopy operations, as shown in Figure 10-25 on page 593. For more information about the presets, see 10.2.1, “FlashCopy presets” on page 584:
 - Snapshot: Creates a PiT snapshot copy of the source volume.
 - Clone: Creates a PiT replica of the source volume.
 - Backup: Creates an incremental FlashCopy mapping that can be used to recover data or objects if the system experiences data loss. These backups can be copied multiple times from source and target volumes.

Note: If you want to create a simple Snapshot of a volume, you likely want the target volume to be defined as thin-provisioned to save space on your system. If you use an existing target, ensure it is thin-provisioned first. The use of the Snapshot preset does not make the system check whether the target volume is thin-provisioned.

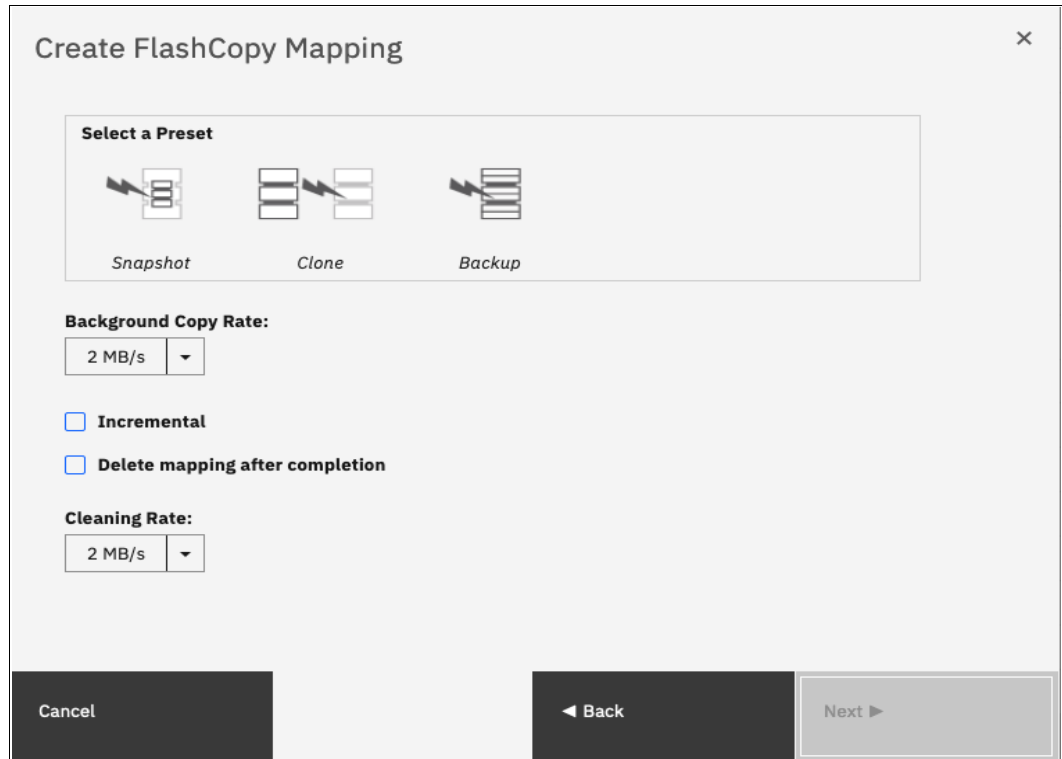


Figure 10-25 FlashCopy mapping preset selection

When selecting a preset, some options, such as Background Copy Rate, Incremental, and Delete mapping after completion, are automatically changed or selected. You can still change the automatic settings, but this action is not recommended for the following reasons:

- If you select the **Backup** preset but then clear **Incremental** or select **Delete mapping after completion**, you lose the benefits of the incremental FlashCopy and must copy the entire source volume each time you start the mapping.
- If you select the **Snapshot** preset but then change the **Background Copy Rate**, you with a full copy of your source volume.

For more information about the Background Copy Rate and the Cleaning Rate, see Table 10-1 on page 566 or Table 10-5 on page 575.

When your FlashCopy mapping setup is ready, click **Next**.

3. You can choose whether to add the mappings to a consistency group, as shown in Figure 10-26.

If you want to include this FlashCopy mapping in a consistency group, select **Yes, add the mappings to a consistency group** and select the consistency group from the drop-down menu.

Figure 10-26 shows a dialog box titled "Create FlashCopy Mapping" with a close button (X) in the top right corner. The main text asks, "Do you want to add the FlashCopy mappings to a consistency group?". There are two radio button options: "No, do not add the mappings to a consistency group." (unselected) and "Yes, add the mappings to a consistency group." (selected). Next to the "Yes" option is a text input field containing "fccstgrp0" and a dropdown arrow. At the bottom, there are three buttons: "Cancel", "< Back", and "Finish".

Figure 10-26 Selecting a consistency group for the FlashCopy mapping

4. It is possible to add a FlashCopy mapping to a consistency group or to remove a FlashCopy mapping from a consistency group after they are created. If you do not know at this stage what to do, you can change it later. Click **Finish**.

The FlashCopy mapping is now ready for use. It is visible in the three different windows: FlashCopy, FlashCopy mappings, and Consistency Groups.

Note: Creating a FlashCopy mapping does *not* automatically start any copy. You must manually start the mapping.

Creating a FlashCopy mapping and target volumes

Complete the following steps to create target volumes for FlashCopy mapping:

1. Right-click the volume that you want to create a FlashCopy mapping for and select **Advanced FlashCopy** → **Create New Target Volumes**, as shown in Figure 10-27.

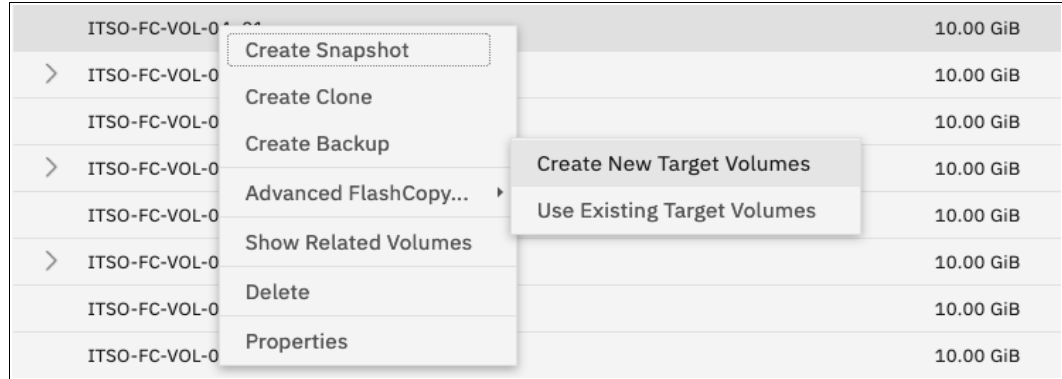


Figure 10-27 Creating a FlashCopy mapping and targets

2. In the next window, select one FlashCopy preset. The GUI provides the following presets to simplify common FlashCopy operations, as shown in Figure 10-28.

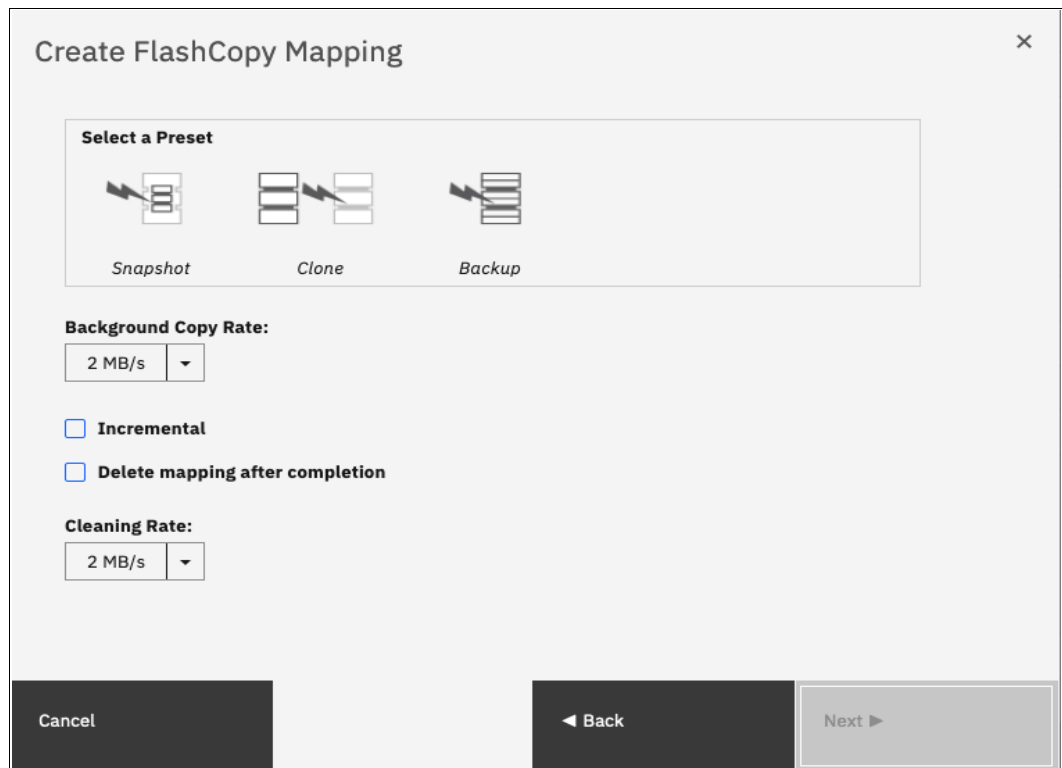


Figure 10-28 FlashCopy mapping preset selection

For more information about the presets, see 10.2.1, “FlashCopy presets” on page 584.

- Snapshot: Creates a PiT snapshot copy of the source volume.
- Clone: Creates a PiT replica of the source volume.

- **Backup:** Creates an incremental FlashCopy mapping that can be used to recover data or objects if the system experiences data loss. These backups can be copied multiple times from source and target volumes.

Note: If you want to create a simple Snapshot of a volume, you likely want the target volume to be defined as thin-provisioned to save space on your system. If you use an existing target, ensure it is thin-provisioned first. The use of the Snapshot preset does not make the system check whether the target volume is thin-provisioned.

When selecting a preset, some options, such as Background Copy Rate, Incremental, and Delete mapping, after completion are automatically changed or selected. You can still change the automatic settings, but this action is not recommended for the following reasons:

- If you select the **Backup** preset but then clear **Incremental** or select **Delete mapping after completion**, you lose the benefits of the incremental FlashCopy. You must copy the entire source volume each time you start the mapping.
- If you select the **Snapshot** preset but then change the **Background Copy Rate**, you have a full copy of your source volume.

For more information about the Background Copy Rate and the Cleaning Rate, see Table 10-1 on page 566 or Table 10-5 on page 575.

When your FlashCopy mapping setup is ready, click **Next**.

3. You can choose whether to add the mappings to a consistency group, as shown in Figure 10-29.

If you want to include this FlashCopy mapping in a consistency group, select **Yes, add the mappings to a consistency group**, and select the consistency group from the drop-down menu.

Create FlashCopy Mapping ×

Do you want to add the FlashCopy mappings to a consistency group?

No, do not add the mappings to a consistency group.

Yes, add the mappings to a consistency group. ▼

Cancel ◀ Back Finish

Figure 10-29 Selecting a consistency group for the FlashCopy mapping

4. It is possible to add a FlashCopy mapping to a consistency group or to remove a FlashCopy mapping from a consistency group after they are created. If you do not know at this stage what to do, you can change it later. Click **Next**.

5. The system prompts the user to select the pool that is used to automatically create targets, as shown in Figure 10-30. Click **Next**.

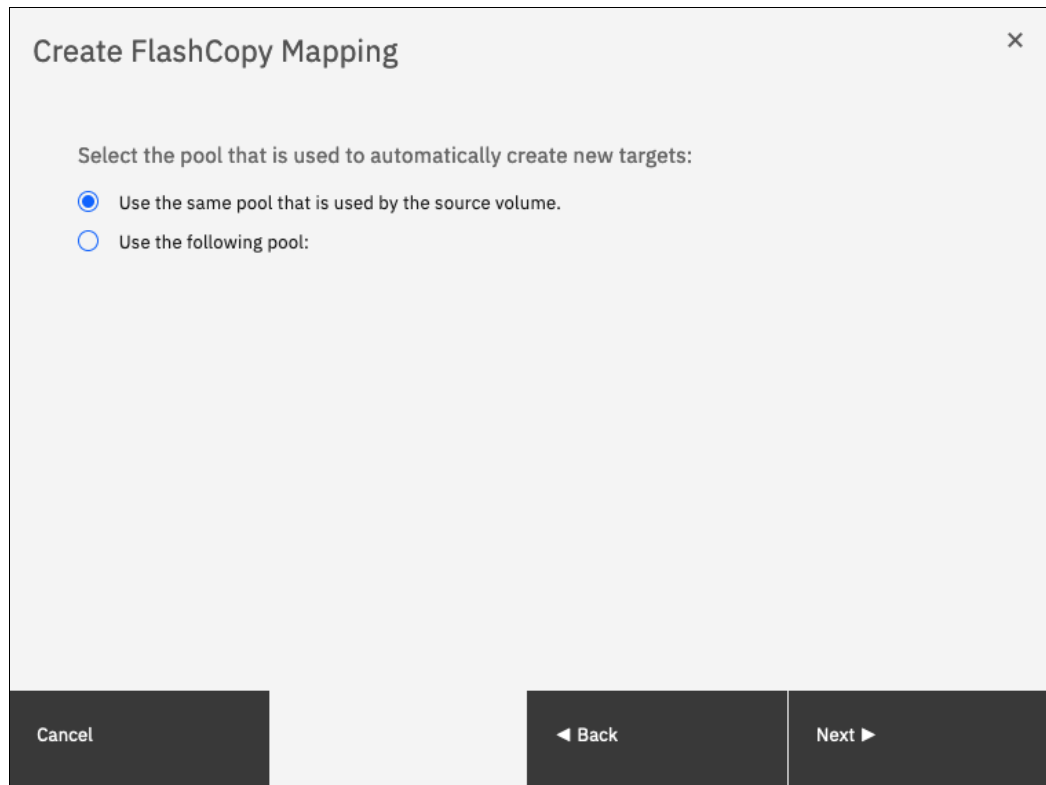


Figure 10-30 Selecting the pool

6. The system prompts the user how to define the new volumes that are created, as shown in Figure 10-31 on page 599. It can be None, Thin-provisioned, or Inherit from source volume. If Inherit from source volume is selected, the system checks the type of the source volume and then creates a target of the same type. Click **Finish**.

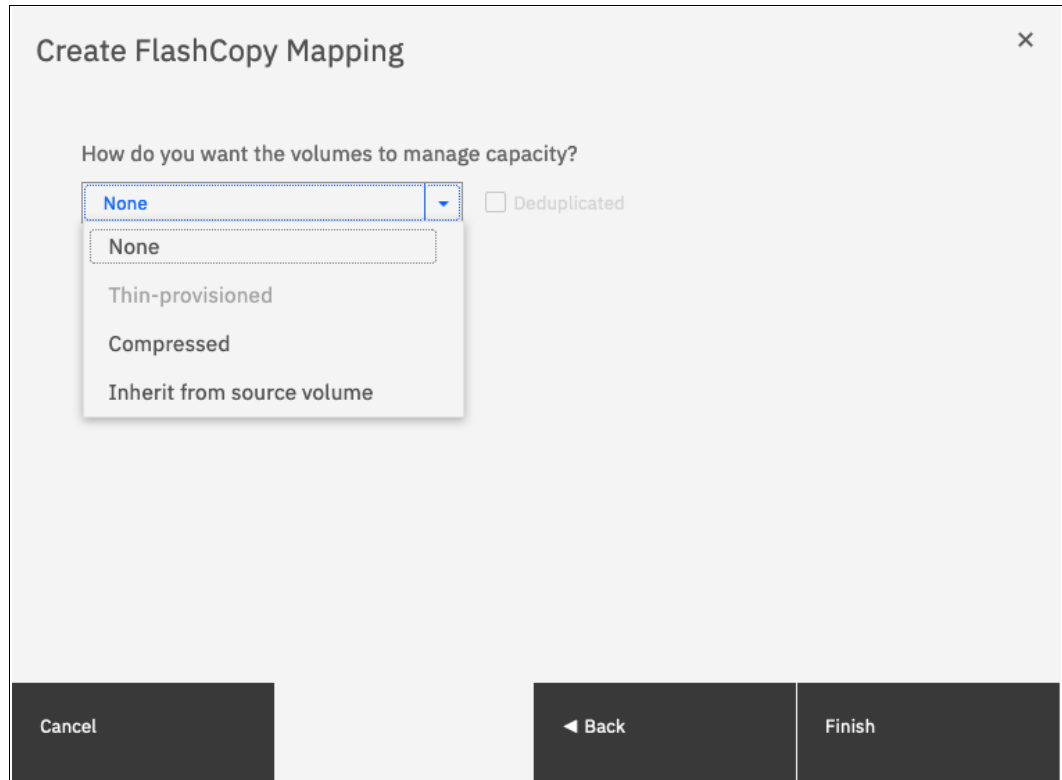


Figure 10-31 Selecting the type of volumes for the created targets

Note: If you selected multiple source volumes to create FlashCopy mappings, selecting **Inherit properties from source Volume** applies to each newly created target volume. For example, if you selected a compressed volume and a generic volume as sources for the new FlashCopy mappings, the system creates a compressed target and a generic target.

The FlashCopy mapping is now ready for use. It is visible in the three different windows: FlashCopy, FlashCopy mappings, and consistency groups.

10.2.4 Single-click snapshot

The *snapshot* creates a PiT backup of production data. The snapshot is not intended to be an independent copy. Instead, it is used to maintain a view of the production data at the time that the snapshot is created. Therefore, the snapshot holds only the data from regions of the production volume that changed since the snapshot was created. Because the snapshot preset uses thin provisioning, only the capacity that is required for the changes is used.

Snapshot uses the following preset parameters:

- ▶ Background copy: No
- ▶ Incremental: No
- ▶ Delete after completion: No
- ▶ Cleaning rate: No
- ▶ Primary copy source pool: Target pool

To create and start a snapshot, complete the following steps:

1. Open the FlashCopy window from the **Copy Services** → **FlashCopy** menu.
2. Select the volume that you want to create a snapshot of, and right-click it or click **Actions** → **Create Snapshot**, as shown in Figure 10-32.

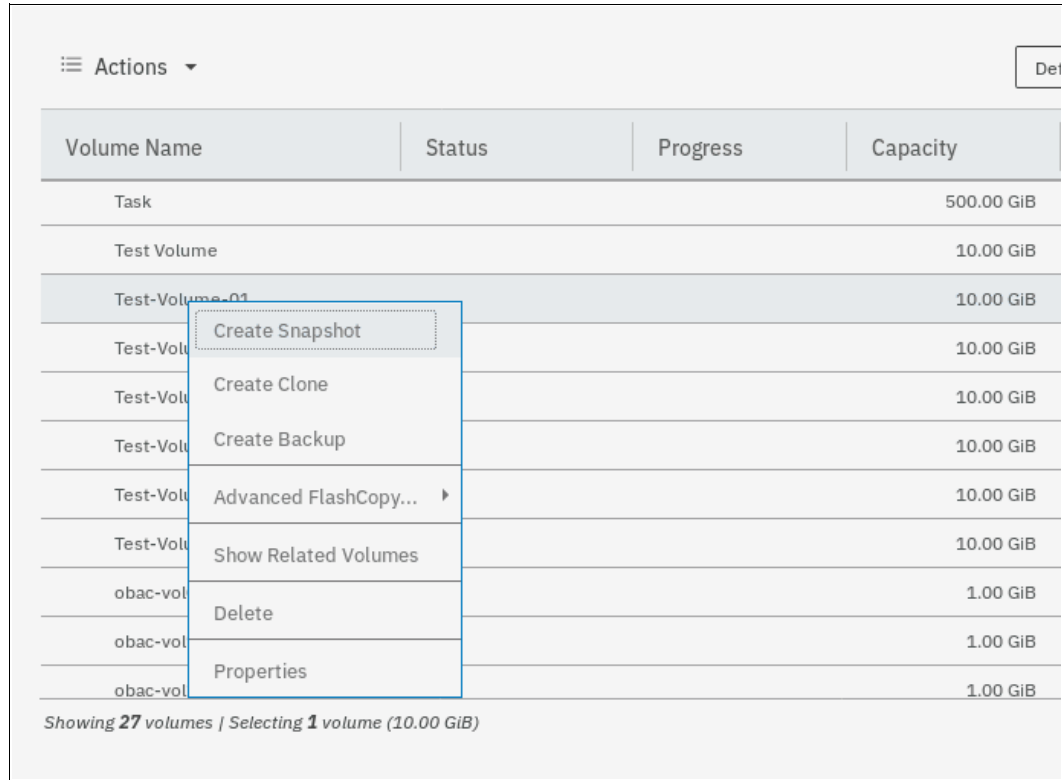


Figure 10-32 Single-click snapshot creation and start

3. You can select multiple volumes at a time, which creates as many snapshots automatically. The system then automatically groups the FlashCopy mappings in a new consistency group, as shown in Figure 10-33 on page 601.

Volume Name	Status	Progress	Capacity
Task			500.00 GiB
Test Volume			10.00 GiB
Test-Volume-01			10.00 GiB
Test-Volume-02			10.00 GiB
Test-Volume-03			10.00 GiB
Test-Volume-04			10.00 GiB
Test-Volume-05			10.00 GiB
Test-Volume-06			10.00 GiB
obac-vol0			1.00 GiB
obac-vol1			1.00 GiB
obac-vol2			1.00 GiB

Showing 27 volumes | Selecting 4 volumes (40.00 GiB)

Figure 10-33 Selecting single-click snapshot creation and start

For each selected source volume, the following actions occur:

- A FlashCopy mapping is automatically created. By default, it is named fcmmapXX.
- A target volume is created. By default, the source name is appended with a _XX suffix.
- A consistency group is created for each mapping, unless multiple volumes were selected. By default, consistency groups are named fccstgrpX.

The newly created consistency group is automatically started.

10.2.5 Single-click clone

The *clone preset* creates a replica of the volume, which can be changed without affecting the original volume. After the copy completes, the mapping that was created by the preset is automatically deleted.

The clone preset uses the following parameters:

- ▶ Background copy rate: 50
- ▶ Incremental: No
- ▶ Delete after completion: Yes
- ▶ Cleaning rate: 50
- ▶ Primary copy source pool: Target pool

To create and start a snapshot, complete the following steps:

1. Open the FlashCopy window from the **Copy Services** → **FlashCopy** menu.
2. Select the volume that you want to create a snapshot of, and right-click it or click **Actions** → **Create Clone**, as shown in Figure 10-34.

Volume Name	Status	Progress	Capacity
Task			500.00 GiB
Test Volume			10.00 GiB
Test-Volume-01	Create Snapshot		10.00 GiB
Test-Volume-02	Create Clone		10.00 GiB
Test-Volume-03	Create Backup		10.00 GiB
Test-Volume-04	Advanced FlashCopy... ▶		10.00 GiB
Test-Volume-05	Show Related Volumes		10.00 GiB
Test-Volume-06	Delete		10.00 GiB
obac-vol0	Properties		1.00 GiB
obac-vol1			1.00 GiB
obac-vol2			1.00 GiB

Showing 27 volumes | Selecting 1 volume (10.00 GiB)

Figure 10-34 Single-click clone creation and start

3. You can select multiple volumes at a time, which creates as many snapshots automatically. The system then automatically groups the FlashCopy mappings in a new consistency group, as shown in Figure 10-35 on page 603.

Volume Name	Status	Progress	Capacity
Task			500.00 GiB
Test Volume			10.00 GiB
Test-Volume-01			10.00 GiB
Test-Volume-02			10.00 GiB
Test-Volume-03			10.00 GiB
Test-Volume-04			10.00 GiB
Test-Volume-05			10.00 GiB
Test-Volume-06			10.00 GiB
obac-vol0			1.00 GiB
obac-vol1			1.00 GiB
obac-vol2			1.00 GiB
Showing 27 volumes Selecting 4 v			

- Create Snapshot as Consistency Group
- Create Clone as Consistency Group
- Create Backup as Consistency Group
- Advanced FlashCopy...
- Show Related Volumes
- Delete
- Properties

Figure 10-35 Selecting single-click clone creation and start

For each selected source volume, the following actions occur:

- A FlashCopy mapping is automatically created. By default, it is named fcmapXX.
- A target volume is created. The source name is appended with an _XX suffix.
- A consistency group is created for each mapping, unless multiple volumes were selected. By default, consistency groups are named fccstgrpX.
- The newly created consistency group is automatically started.

10.2.6 Single-click backup

The backup creates a PiT replica of the production data. After the copy completes, the backup view can be refreshed from the production data, with minimal copying of data from the production volume to the backup volume. The backup preset uses the following parameters:

- ▶ Background Copy rate: 50
- ▶ Incremental: Yes
- ▶ Delete after completion: No
- ▶ Cleaning rate: 50
- ▶ Primary copy source pool: Target pool

To create and start a backup, complete the following steps:

1. Open the FlashCopy window from the **Copy Services** → **FlashCopy** menu.
2. Select the volume that you want to create a backup of, and right-click it or click **Actions** → **Create Backup**, as shown in Figure 10-36.

Volume Name	Status	Progress	Capacity
Task			500.00 GiB
Test Volume			10.00 GiB
Test-Volume-01		Create Snapshot	10.00 GiB
Test-Volume-02		Create Clone	10.00 GiB
Test-Volume-03		Create Backup	10.00 GiB
Test-Volume-04		Advanced FlashCopy... ▶	10.00 GiB
Test-Volume-05		Show Related Volumes	10.00 GiB
Test-Volume-06		Delete	10.00 GiB
obac-vol0		Properties	1.00 GiB
obac-vol1			1.00 GiB
obac-vol2			1.00 GiB

Showing 27 volumes | Selecting 1 volume (10.00 GiB)

Figure 10-36 Single-click backup creation and start

- You can select multiple volumes at a time, which creates as many snapshots automatically. The system then automatically groups the FlashCopy mappings in a new consistency group, as shown Figure 10-37.

Volume Name	Status	Progress	Capacity
Task			500.00 GiB
Test Volume			10.00 GiB
Test-Volume-01			10.00 GiB
Test-Volume-02			10.00 GiB
Test-Volume-03			10.00 GiB
Test-Volume-04			10.00 GiB
Test-Volume-05			10.00 GiB
Test-Volume-06			10.00 GiB
obac-vol0			1.00 GiB
obac-vol1			1.00 GiB
obac-vol2			1.00 GiB
Showing 27 volumes Selecting			

Create Snapshot as Consistency Group

Create Clone as Consistency Group

Create Backup as Consistency Group

Advanced FlashCopy...

Show Related Volumes

Delete

Figure 10-37 Selecting single-click backup creation and start

For each selected source volume, the following actions occur:

- A FlashCopy mapping is automatically created. By default, it is named fcmmapXX.
- A target volume is created. It is named after the source name with a _XX suffix.
- A consistency group is created for each mapping, unless multiple volumes were selected. By default, consistency groups are named fccstgrpX.
- The newly created consistency group is automatically started.

10.2.7 Creating a FlashCopy consistency group

To create a FlashCopy consistency group in the GUI, complete the following steps:

- Open the Consistency Groups window by clicking **Copy Services** → **Consistency Groups**. Click **Create Consistency Group**, as shown in Figure 10-38.

Mapping Name	Status	Source Volume	Target Volume	Progress
Not in a Group				
fcmmap0	✓ Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_01	100%
fcmmap1	✓ Idle	ITSO-VOL0	ITSO-VOL0_02	0%
fcmmap7	⌚ Copying	ITSO-FC-VOL-01	ITSO-FC-VOL-01_04	0%
fcmmap8	⌚ Copying	ITSO-FC-VOL-01	ITSO-FC-VOL-01_05	0%

Figure 10-38 Creating a consistency group

2. Enter the FlashCopy consistency group name that you want to use then, click **Create**, as shown in Figure 10-39.

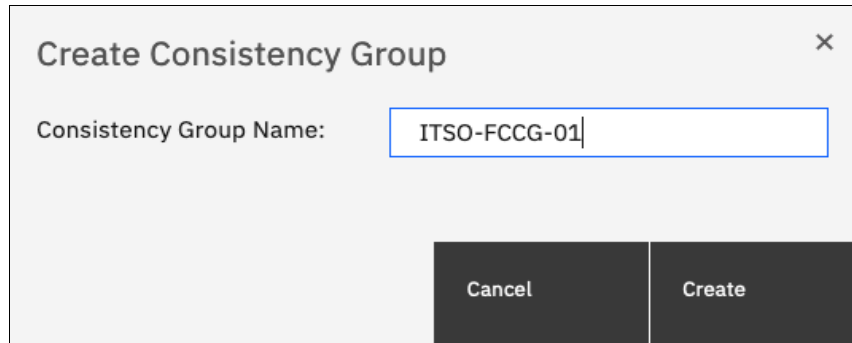


Figure 10-39 Entering the name and ownership group of a new consistency group

Consistency group name: You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The volume name can be 1 - 63 characters.

10.2.8 Creating FlashCopy mappings in a consistency group

To create a FlashCopy consistency group in the GUI, complete the following steps:

1. Open the Consistency Groups window by clicking **Copy Services** → **Consistency Groups**. This example assumes that source and target volumes were previously created.
2. Select the consistency group in which you want to create the FlashCopy mapping. If you prefer not to create a FlashCopy mapping in a consistency group, select **Not in a Group**, and right-click the selected consistency group or click **Actions** → **Create FlashCopy Mapping**, as shown in Figure 10-40.

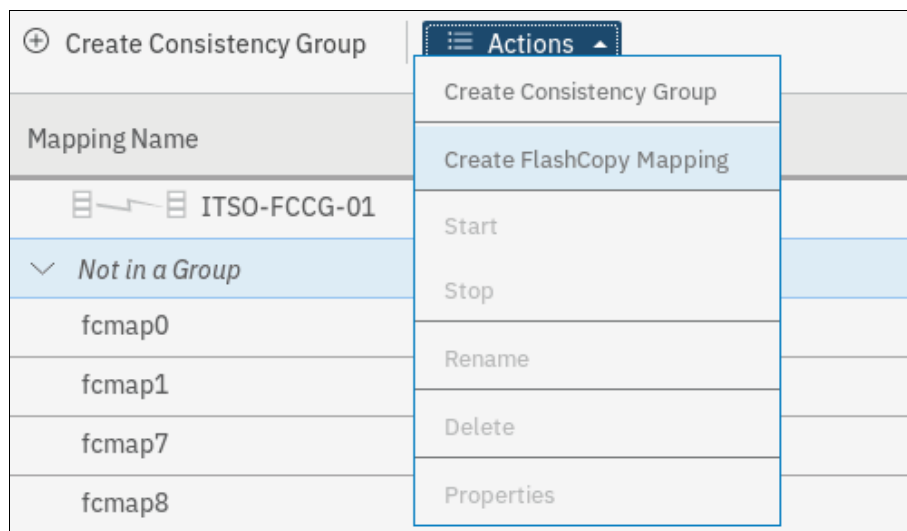


Figure 10-40 Creating a FlashCopy mapping

3. Select a volume in the source volume column by using the drop-down menu. Then, select a volume in the target volume column by using the drop-down menu. Click **Add**, as shown in Figure 10-41.

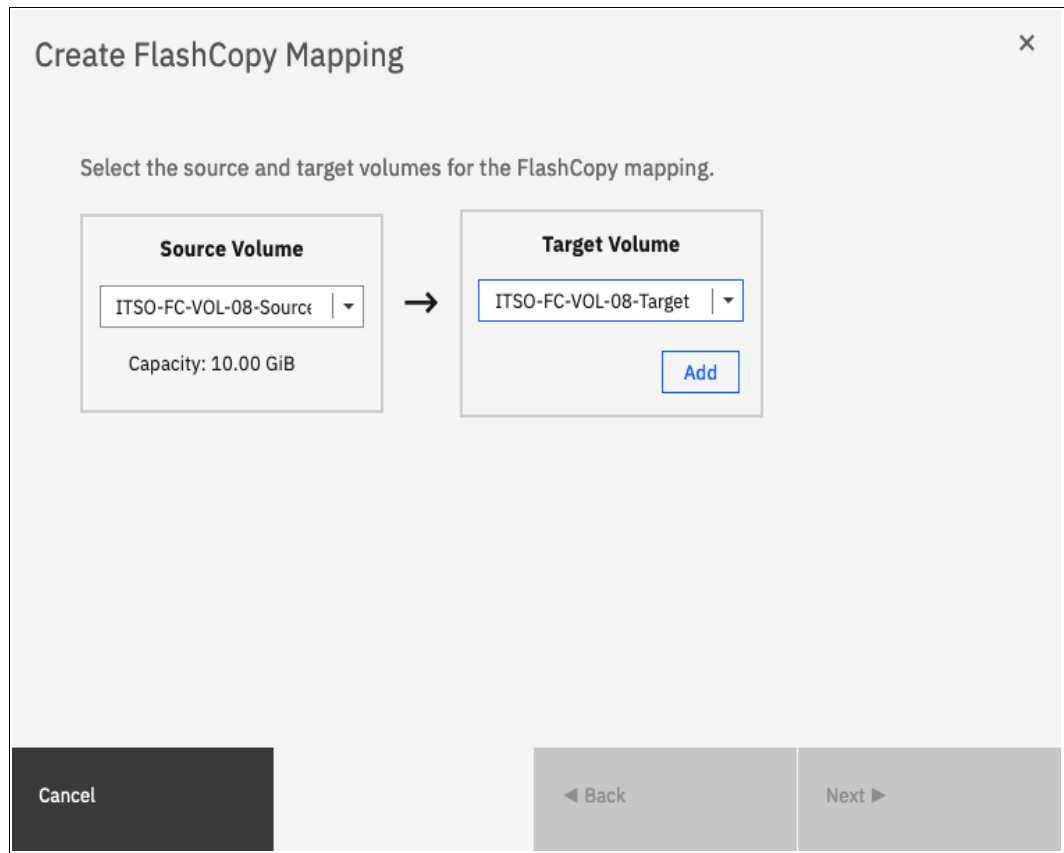


Figure 10-41 Selecting source and target volumes for the FlashCopy mapping

Repeat this step to create other mappings. To remove a mapping that was created, click **X**. Click **Next**.

Important: The source and target volumes must be of equal size. Therefore, only the targets with the suitable size are shown for a source volume.

Volumes that are target volumes in another FlashCopy mapping cannot be target of a new FlashCopy mapping. Therefore, they do not appear in the list.

4. In the next window, select one FlashCopy preset. The GUI provides the following presets to simplify common FlashCopy operations, as shown in Figure 10-42. For more information about the presets, see 10.2.1, “FlashCopy presets” on page 584:
 - Snapshot: Creates a PiT snapshot copy of the source volume.
 - Clone: Creates a PiT replica of the source volume.
 - Backup: Creates an incremental FlashCopy mapping that can be used to recover data or objects if the system experiences data loss. These backups can be copied multiple times from source and target volumes.

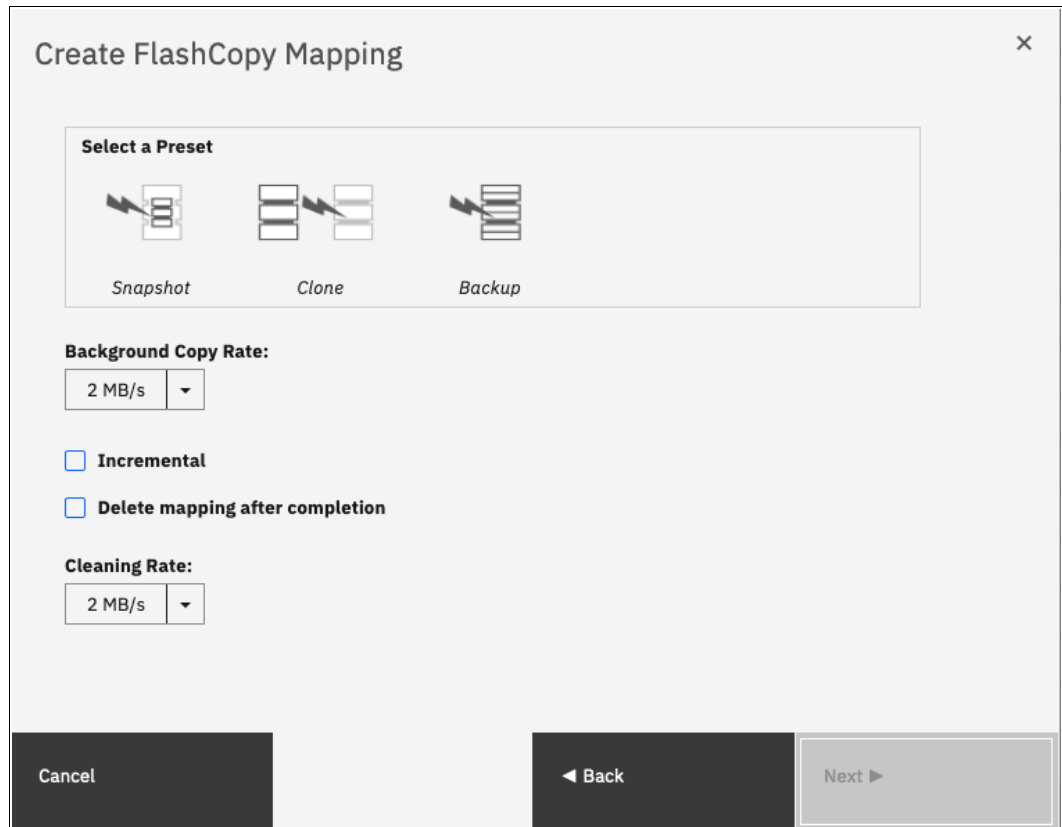


Figure 10-42 FlashCopy mapping preset selection

When selecting a preset, some options, such as Background Copy Rate, Incremental, and Delete mapping after completion, are automatically changed or selected. You can still change the automatic settings, but this is not recommended for the following reasons:

- If you select the **Backup** preset but then clear **Incremental** or select **Delete mapping after completion**, you lose the benefits of the incremental FlashCopy. You must copy the entire source volume each time you start the mapping.
- If you select the **Snapshot** preset but then change the **Background Copy Rate**, you have a full copy of your source volume.

For more information about the Background Copy Rate and the Cleaning Rate, see Table 10-1 on page 566 or Table 10-5 on page 575.

5. When your FlashCopy mapping setup is ready, click **Finish**.

10.2.9 Showing related volumes

To show related volumes for a specific FlashCopy mapping, complete the following steps:

1. Open the Copy Services FlashCopy Mappings window.
2. Right-click a FlashCopy mapping and select **Show Related Volumes**, as shown in Figure 10-43. Also, depending on which window you are inside Copy Services, you can right-click at mappings and select **Show Related Volumes**.

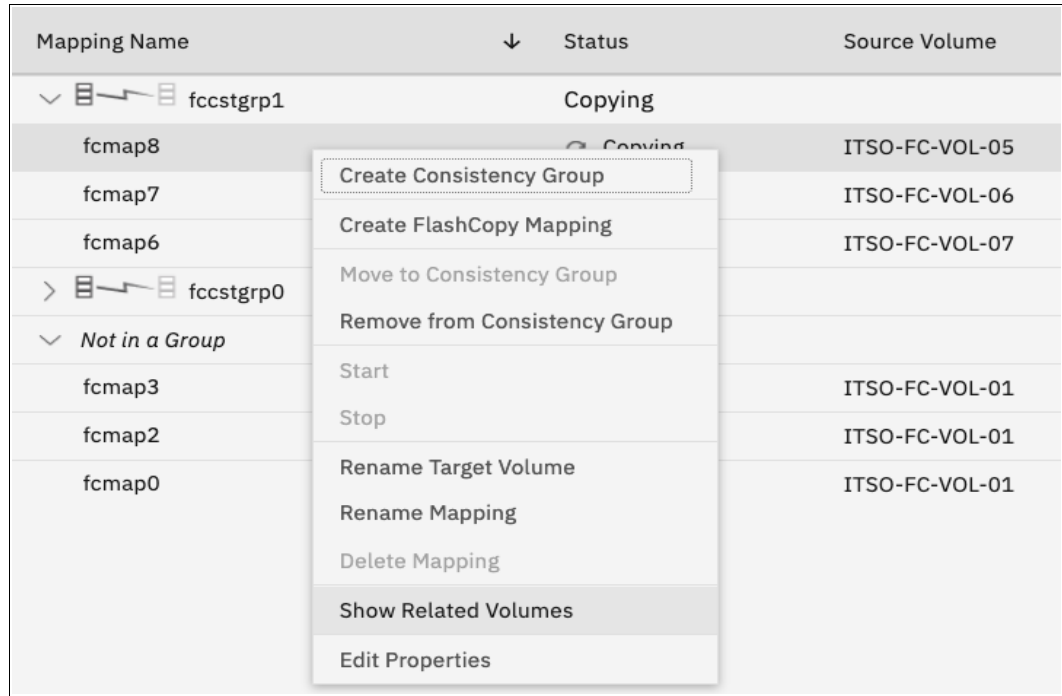


Figure 10-43 Showing related volumes for a mapping, a consistency group, or another volume

3. In the Related Volumes window, you can see the related mapping for a volume, as shown in Figure 10-44. If you click one of these volumes, you can see its properties.

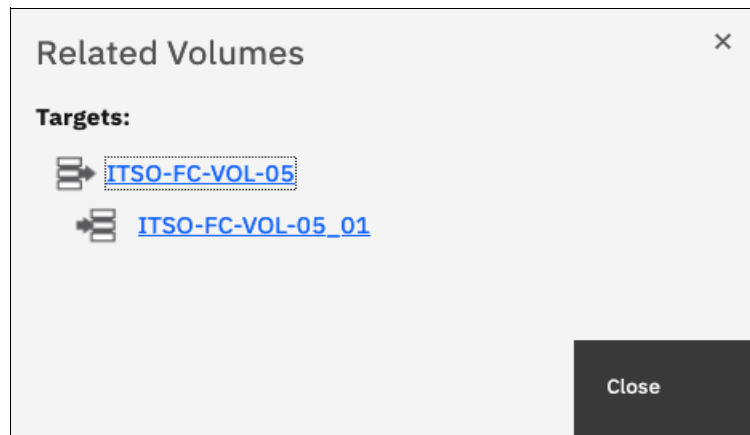


Figure 10-44 Showing related volumes list

10.2.10 Moving FlashCopy mappings across consistency groups

To move one or multiple FlashCopy mappings to a consistency group, complete the following steps:

1. Open the FlashCopy, Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mappings that you want to move and select **Move to Consistency Group**, as shown in Figure 10-45.

Mapping Name	Status	Source Volume	Target Volume	Progress
> fccstgrp1	Copying			
∨ fccstgrp0	Idle or Copied			
fcmap5	✓ Copied	ITSO-FC-VOL-03	ITSO-FC-VOL-03_01	100%
fcmap4		ITSO-FC-VOL-02	ITSO-FC-VOL-02_01	100%
fcmap1		ITSO-FC-VOL-04	ITSO-FC-VOL-04_01	100%
∨ Not in a Group				
fcmap3		ITSO-FC-VOL-01	ITSO-FC-VOL-01_04	0%
fcmap2		ITSO-FC-VOL-01	ITSO-FC-VOL-01_03	100%
fcmap0		ITSO-FC-VOL-01	ITSO-FC-VOL-01_01	100%

- Create Consistency Group
- Create FlashCopy Mapping
- Move to Consistency Group**
- Remove from Consistency Group
- Start
- Stop
- Rename Mapping
- Delete Mapping
- Show Related Volumes
- Edit Properties

Figure 10-45 Moving a FlashCopy mapping to a consistency group

Note: You cannot move a FlashCopy mapping that is in a copying, stopping, or suspended state. The mapping should be idle-or-copied or stopped to be moved.

3. In the Move FlashCopy Mapping to Consistency Group window, select the Consistency Group for the FlashCopy mappings selection by using the drop-down menu, as shown in Figure 10-46.

Move FlashCopy Mapping to Consistency Group ✕

Select the FlashCopy consistency group to which to move FlashCopy mapping **fcmap2**.

Consistency Group ▾

Cancel Move to Consistency Group

Figure 10-46 Selecting the consistency group to which to move the FlashCopy mapping

4. Click **Move to Consistency Group** to confirm your changes.

10.2.11 Removing FlashCopy mappings from consistency groups

To remove one or multiple FlashCopy mappings from a consistency group, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mappings that you want to remove and select **Remove from Consistency Group**, as shown in Figure 10-47.

Note: Only FlashCopy mappings that belong to a consistency group can be removed.

Mapping Name	Status	Source Volume	Target Volume
ITSO-FCCG-01	Idle or Copied		
fcmap1	✓ Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_02
fcmap2	✓ Copied		02-Target
Not in a Group			
fcmap0	✓ Copied		01_01
fcmap3	↻ Copying		01_03_01
fcmap7	✓ Copied		01_04

Selected 1 FlashCopy mapping

Figure 10-47 Removing FlashCopy mappings from a consistency group

3. In the Remove FlashCopy Mapping from Consistency Group window, click **Remove**, as shown in Figure 10-48.

✕

Remove FlashCopy Mapping from Consistency Group

⚠ **Removing FlashCopy mapping(s) from consistency group cannot be undone. Are you sure you want to continue?**

You selected **1** FlashCopy mapping to remove from consistency group **fccstgrp0**.

fcmap5 (ITSO-FC-VOL-03 -> ITSO-FC-VOL-03_01)

Cancel

Remove

Figure 10-48 Confirming the selection of mappings to be removed

10.2.12 Modifying a FlashCopy mapping

To modify a FlashCopy mapping, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mapping that you want to edit and select **Edit Properties**, as shown in Figure 10-49.

Mapping Name ↑	Status	Source Volume	Target Volume	Progress
fcmap0	✓ Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_01	100%
fcmap1	✓ Copied	ITSO-FC-VOL-01_01	ITSO-FC-VOL-01_02	100%
fcmap2	✓ Copied	ITSO-F	et	100%
fcmap3	🔄 Copying	ITSO-F	01	0%
fcmap7	✓ Copied	ITSO-F		100%

Create FlashCopy Mapping

Move to Consistency Group

Remove from Consistency Group

Start

Stop

Rename Mapping

Delete Mapping

Show Related Volumes

Edit Properties

Showing 5 FC mappings | Selecting 1 FC mapping

Figure 10-49 Editing FlashCopy mapping properties

Note: It is not possible to select multiple FlashCopy mappings to edit their properties concurrently.

3. In the Edit FlashCopy Mapping window, you can modify the background copy rate and the cleaning rate for a selected FlashCopy mapping, as shown in Figure 10-50.

Edit FlashCopy Mapping

Background Copy Rate:

128 MB/s ▼

Cleaning Rate:

0 KB/s ▼

Cancel Save

Figure 10-50 Editing copy and cleaning rates of a FlashCopy mapping

For more information about the Background Copy Rate and the Cleaning Rate, see Table 10-1 on page 566 or Table 10-5 on page 575.

4. Click **Save** to confirm your changes.

10.2.13 Renaming FlashCopy mappings

To rename one or multiple FlashCopy mappings, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mappings that you want to rename and select **Rename Mapping**, as shown in Figure 10-51.

Mapping Name	Status	Source Volume	Target Volume	Progress
fcmap0	✓ Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_01	100%
fcmap1	✓ Copied	ITSO-FC-VOL-01_01	ITSO-FC-VOL-01_02	100%
fcmap2	✓ Copied	ITSO-FC-VOL-01_01	ITSO-FC-VOL-01_01	100%
fcmap3	⌛ Copying	ITSO-FC-VOL-01_01	ITSO-FC-VOL-01_01	0%
fcmap7	✓ Copied	ITSO-FC-VOL-01_01	ITSO-FC-VOL-01_01	100%

Showing 5 FC mappings | Selecting 1 FC mapping

Figure 10-51 Renaming FlashCopy mappings

3. In the Rename FlashCopy Mapping window, enter the new name that you want to assign to each FlashCopy mapping and click **Rename**, as shown in Figure 10-52.

FlashCopy mapping name: You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The FlashCopy mapping name can be 1 - 63 characters.

Rename FlashCopy Mapping [X]

*New Name

fcmap1

Cancel Reset Rename

Figure 10-52 Renaming the selected FlashCopy mappings

Renaming a consistency group

To rename a consistency group, complete the following steps:

1. Open the Consistency Groups window.
2. Right-click the consistency group that you want to rename and select **Rename**, as shown in Figure 10-53.

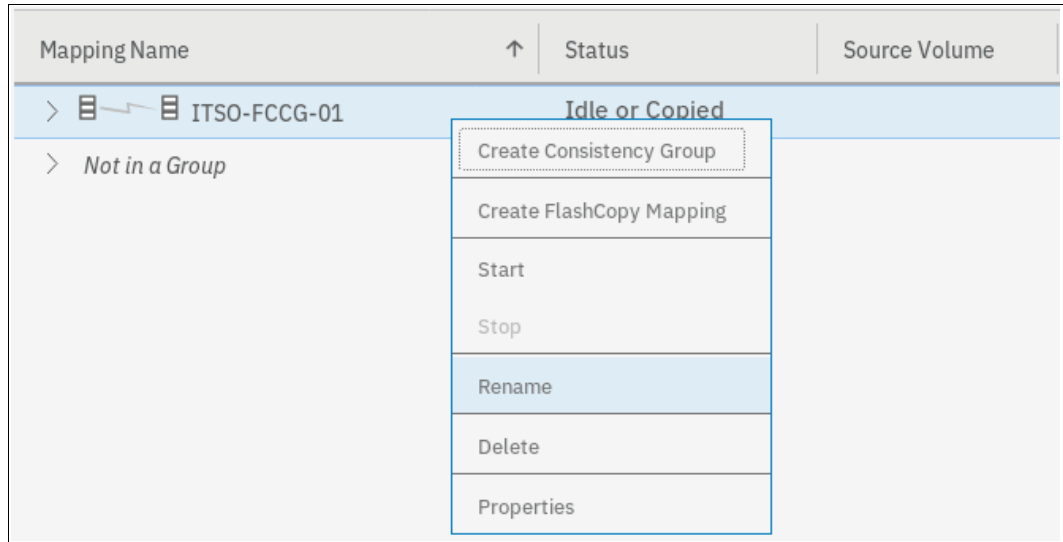


Figure 10-53 Renaming a consistency group

3. Enter the new name that you want to assign to the consistency group and click **Rename**, as shown in Figure 10-54.

Note: It is not possible to select multiple consistency groups to edit their names all at the same time.

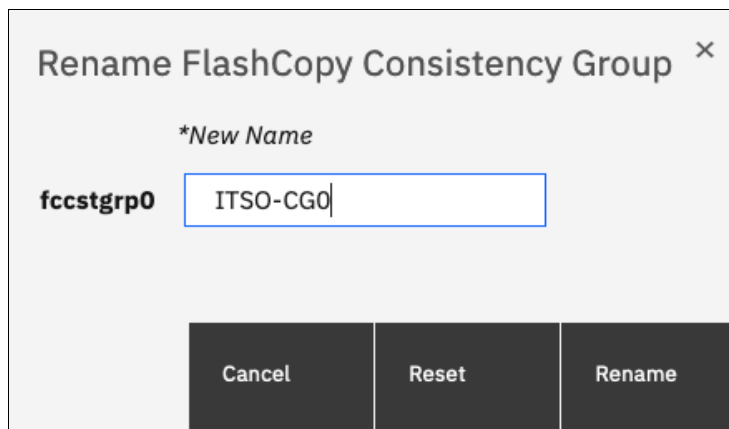


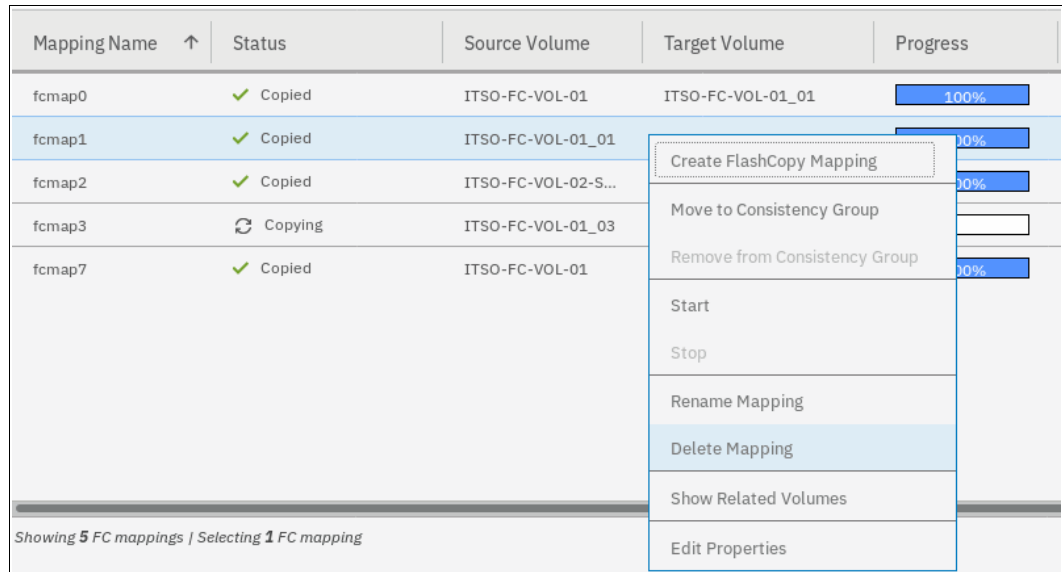
Figure 10-54 Renaming the selected consistency group

Consistency group name: The name can consist of the letters A - Z and a - z, the numbers 0 - 9, the dash (-), and the underscore (_) character. The name can be 1 - 63 characters. However, the name cannot start with a number, a dash, or an underscore.

10.2.14 Deleting FlashCopy mappings

To delete one or multiple FlashCopy mappings, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mappings that you want to delete and select **Delete Mapping**, as shown in Figure 10-55.



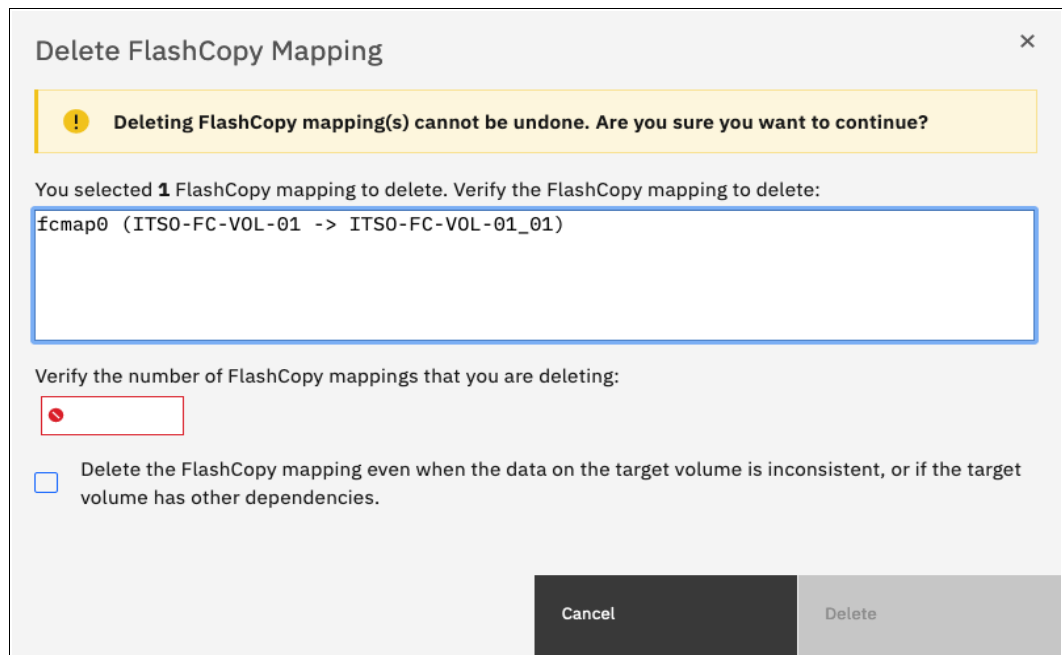
Mapping Name ↑	Status	Source Volume	Target Volume	Progress
fcmap0	✓ Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_01	100%
fcmap1	✓ Copied	ITSO-FC-VOL-01_01		100%
fcmap2	✓ Copied	ITSO-FC-VOL-02-S...		100%
fcmap3	🔄 Copying	ITSO-FC-VOL-01_03		
fcmap7	✓ Copied	ITSO-FC-VOL-01		100%

Showing 5 FC mappings | Selecting 1 FC mapping

- Create FlashCopy Mapping
- Move to Consistency Group
- Remove from Consistency Group
- Start
- Stop
- Rename Mapping
- Delete Mapping
- Show Related Volumes
- Edit Properties

Figure 10-55 Deleting FlashCopy mappings

3. The Delete FlashCopy Mapping window opens, as shown in Figure 10-56. In the **Verify the number of FlashCopy mappings that you are deleting** field, enter the number of volumes that you want to remove. This verification was added to help avoid deleting the wrong mappings.



Delete FlashCopy Mapping [X]

⚠ **Deleting FlashCopy mapping(s) cannot be undone. Are you sure you want to continue?**

You selected **1** FlashCopy mapping to delete. Verify the FlashCopy mapping to delete:

fcmap0 (ITSO-FC-VOL-01 -> ITSO-FC-VOL-01_01)

Verify the number of FlashCopy mappings that you are deleting:

Delete the FlashCopy mapping even when the data on the target volume is inconsistent, or if the target volume has other dependencies.

Cancel Delete

Figure 10-56 Confirming the selection of FlashCopy mappings to be deleted

4. If you still have target volumes that are inconsistent with the source volumes and you want to delete these FlashCopy mappings, select the **Delete the FlashCopy mapping even when the data on the target volume is inconsistent, or if the target volume has other dependencies** option. Click **Delete**.

10.2.15 Deleting a FlashCopy consistency group

Important: Deleting a consistency group does not delete the FlashCopy mappings that it contains.

To delete a consistency group, complete the following steps:

1. Open the Consistency Groups window.
2. Right-click the consistency group that you want to delete and select **Delete**, as shown in Figure 10-57.

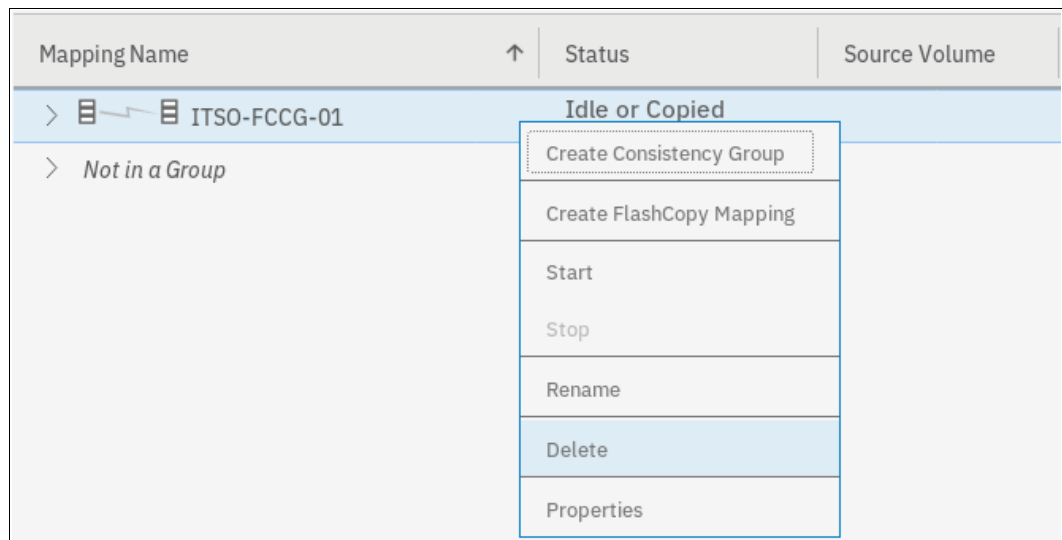


Figure 10-57 Deleting a consistency group

3. A warning message is displayed, as shown in Figure 10-58. Click **Yes**.

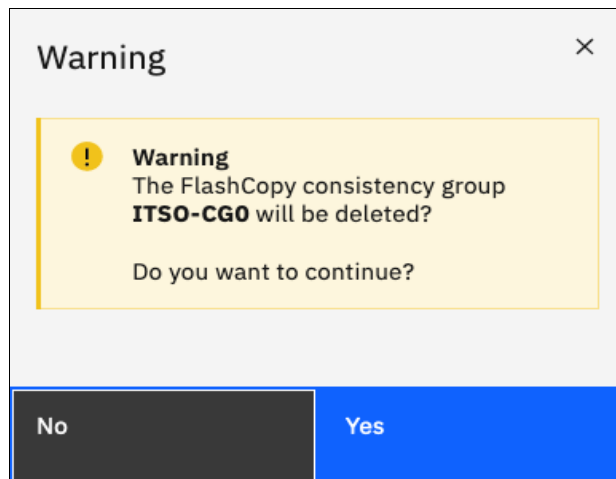


Figure 10-58 Confirming the consistency group deletion

10.2.16 Starting FlashCopy mappings

Important: Only FlashCopy mappings that do not belong to a consistency group can be started individually. If FlashCopy mappings are part of a consistency group, they can be started only all together by using the consistency group **start** command.

It is the **start** command that defines the “PiT”. It is the moment that is used as a reference (T0) for all subsequent operations on the source and the target volumes. To start one or multiple FlashCopy mappings that do not belong to a consistency group, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mappings that you want to start and select **Start**, as shown in Figure 10-59.

Mapping Name ↑	Status	Source Volume	Target Volume	Progress
fcmap0	✓ Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_01	100%
fcmap1	✓ Copied	ITSO-FC-VOL-01_01	ITSO-FC-VOL-01_02	100%
fcmap2	✓ Copied		D-FC-VOL-02-Target	100%
fcmap3	🔄 Copying		D-FC-VOL-01_03_01	0%
fcmap7	✓ Copied		D-FC-VOL-01_04	100%

Create FlashCopy Mapping

Move to Consistency Group

Remove from Consistency Group

Start

Stop

Rename Mapping

Delete Mapping

Show Related Volumes

Edit Properties

Showing 5 FC mappings | Selecting 1 FC mapping

Figure 10-59 Starting FlashCopy mappings

You can check the FlashCopy state and the progress of the mappings in the Status and Progress columns of the table, as shown in Figure 10-60.

Mapping Name ↑	Status	Source Volume	Target Volume	Progress	Group
fcmap0	✓ Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_01	100%	
fcmap1	🔄 Copying	ITSO-FC-VOL-01_01	ITSO-FC-VOL-01_02	3%	
fcmap2	✓ Copied	ITSO-FC-VOL-02-S...	ITSO-FC-VOL-02-Target	100%	ITSO-FCCG-01
fcmap3	🔄 Copying	ITSO-FC-VOL-01_03	ITSO-FC-VOL-01_03_01	0%	
fcmap7	✓ Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_04	100%	

Figure 10-60 FlashCopy mappings status and progress examples

FlashCopy Snapshots depend on the source volume and should be in a “copying” state if the mapping is started.

FlashCopy clones and the first occurrence of FlashCopy backup can take some time to complete, depending on the copyrate value and the size of the source volume. The next occurrences of FlashCopy backups are faster because only the changes that were made during two occurrences are copied.

For more information about FlashCopy starting operations and states, see 10.1.10, “Starting FlashCopy mappings and consistency groups” on page 568.

10.2.17 Stopping FlashCopy mappings

Important: Only FlashCopy mappings that do not belong to a consistency group can be stopped individually. If FlashCopy mappings are part of a consistency group, they can be stopped all together only by using the consistency group **stop** command.

The only reason to stop a FlashCopy mapping is for incremental FlashCopy. When the first occurrence of an incremental FlashCopy is started, a full copy of the source volume is made. When 100% of the source volume is copied, the FlashCopy mapping does not stop automatically and a manual stop can be performed. The target volume is available for read and write operations, during the copy, and after the mapping is stopped.

In any other case, stopping a FlashCopy mapping interrupts the copy and resets the bitmap table. Because only part of the data from the source volume was copied, the copied grains might be meaningless without the remaining grains. Therefore, the target volumes are placed offline and are unusable, as shown in Figure 10-61.

Mapping Name	Flash Time	Status	Source Volume	Target Volume	Progress	Group
fcmap0		Stopped	ITSO-FC-VOL-01	ITSO-FC-VOL-01_01	0%	
fcmap2	11/3/2020 5:43:07 PM	Copied	ITSO-FC-VOL-01	ITSO-FC-VOL-01_03	100%	
fcmap3	11/3/2020 5:43:22 PM	Copying	ITSO-FC-VOL-01	ITSO-FC-VOL-01_04	0%	
fcmap4	11/3/2020 5:49:12 PM	Copied	ITSO-FC-VOL-02	ITSO-FC-VOL-02_01	100%	ITSO-CG0
fcmap5	11/3/2020 5:49:12 PM	Copied	ITSO-FC-VOL-03	ITSO-FC-VOL-03_01	100%	ITSO-CG0

Figure 10-61 Showing the target volumes state and FlashCopy mappings status

To stop one or multiple FlashCopy mappings that do not belong to a consistency group, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.
2. Right-click the FlashCopy mappings that you want to stop and select **Stop**, as shown in Figure 10-62 on page 619.

Mapping Name	Status	Source Volume	Target Volume	Progress	Group
fcmap0	Copying	ITSO-FC-VOL-01	ITSO-FC-VOL-01_01	51%	
fcmap2	Copied		C-VOL-01_03	100%	
fcmap3	Copying		C-VOL-01_04	0%	
fcmap4	Copied		C-VOL-02_01	100%	ITSO-CG0
fcmap5	Copied		C-VOL-03_01	100%	ITSO-CG0
fcmap6	Copying		C-VOL-07_01	0%	fccstgrp1
fcmap7	Copying		C-VOL-06_01	0%	fccstgrp1
fcmap8	Copying		C-VOL-05_01	0%	fccstgrp1

Figure 10-62 Stopping FlashCopy mappings

Note: FlashCopy mappings can be in a stopping state for some time if you created dependencies between several targets. It is in a cleaning mode. For more information about dependencies and stopping process, see “Stopping process in a multiple target FlashCopy: Cleaning Mode” on page 574.

10.2.18 Memory allocation for FlashCopy

Copy Services features require that small amounts of volume cache be converted from cache memory into bitmap memory to allow the functions to operate at an I/O group level. If not enough bitmap space is allocated when you try to use one of the functions, you cannot complete the configuration. The total memory that can be dedicated to these functions is not defined by the physical memory in the system. The memory is constrained by the software functions that use the memory.

For every FlashCopy mapping that is created on an IBM Spectrum Virtualize system, a bitmap table is created to track the copied grains. By default, the system allocates 20 MiB of memory for a minimum of 10 TiB of FlashCopy source volume capacity and 5 TiB of incremental FlashCopy source volume capacity.

Depending on the grain size of the FlashCopy mapping, the memory capacity usage is different. 1 MiB of memory provides the following volume capacity for the specified I/O group:

- ▶ For clones and snapshots FlashCopy with 256 KiB grains size, 2 TiB of total FlashCopy source volume capacity
- ▶ For clones and snapshots FlashCopy with 64 KiB grains size, 512 GiB of total FlashCopy source volume capacity
- ▶ For incremental FlashCopy, with 256 KiB grains size, 1 TiB of total incremental FlashCopy source volume capacity
- ▶ For incremental FlashCopy, with 64 KiB grains size, 256 GiB of total incremental FlashCopy source volume capacity

To calculate the memory requirements and confirm that your system can accommodate the total installation size, see Table 10-9.

Table 10-9 Memory allocation for FlashCopy services

Minimum allocated bitmap space	Default allocated bitmap space	Maximum allocated bitmap space	Minimum ^a functionality when using the default values
0	20 MiB	2 GiB	10 TiB of FlashCopy source volume capacity 5 TiB of incremental FlashCopy source volume capacity

a. The actual amount of functionality might increase based on settings, such as grain size and strip size.

FlashCopy includes the FlashCopy function, Global Mirror with Change Volumes (GMCV), and active-active (HyperSwap) relationships.

For multiple FlashCopy targets, you must consider the number of mappings. For example, for a mapping with a grain size of 256 KiB, 8 KiB of memory allows one mapping between a 16 GiB source volume and a 16 GiB target volume. Alternatively, for a mapping with a 256 KiB grain size, 8 KiB of memory allows two mappings between one 8 GiB source volume and two 8 GiB target volumes.

When creating a FlashCopy mapping, if you specify an I/O group other than the I/O group of the source volume, the memory accounting goes toward the specified I/O group, not toward the I/O group of the source volume.

When creating FlashCopy relationships or mirrored volumes, more bitmap space is allocated automatically by the system, if required.

For FlashCopy mappings, only one I/O group uses bitmap space. By default, the I/O group of the source volume is used.

When you create a reverse mapping, such as when you run a restore operation from a snapshot to its source volume, a bitmap is created.

When you configure change volumes for use with GM, two internal FlashCopy mappings are created for each change volume.

You can modify the resource allocation for each I/O group of an IBM Spectrum Virtualize system by selecting **Settings** → **System** and clicking the **Resources** menu, as shown in Figure 10-63 on page 621. At the time of writing, this GUI option is not available for other IBM Spectrum Virtualize based systems, so resource allocation can be adjusted by running the `chiogrp` command. For more information about the command's syntax, see [IBM Documentation](#).

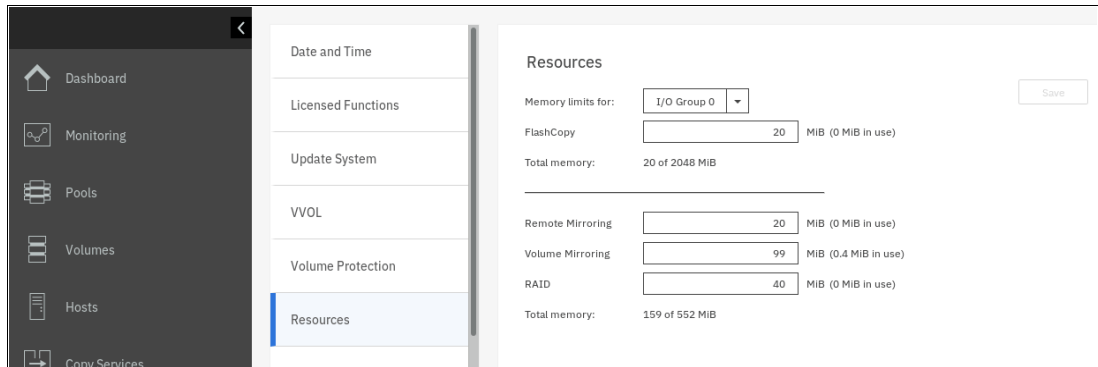


Figure 10-63 Modifying resources allocation per I/O group

10.3 Transparent Cloud Tiering

Introduced in Version 7.8, TCT is a function of IBM Spectrum Virtualize that uses IBM FlashCopy mechanisms to produce a PiT snapshot of the data. TCT helps to increase the flexibility to protect and transport data to public or private cloud infrastructure. This technology is built on top of IBM Spectrum Virtualize software capabilities. TCT uses the cloud to store snapshot targets and provides alternatives to restore snapshots from the private and public cloud of an entire volume or set of volumes.

TCT can help to solve business needs that require duplication of data of your source volume. Volumes can remain online and active while you create snapshot copies of the data sets. TCT operates below the host OS and its cache. Therefore, the copy is not apparent to the host.

IBM Spectrum Virtualize features built-in software algorithms that allow the TCT function to securely interact; for example, with Information Dispersal Algorithms (IDAs), which is essentially the interface to IBM Cloud Object Storage.

Object Storage is a general term that refers to the entity in which IBM Cloud Object Storage organizes, manages, and stores units of data. To transform these snapshots of traditional data into Object Storage, the storage nodes and the IDA import the data and transform it into several metadata and slices. The object can be read by using a subset of those slices. When an Object Storage entity is stored as IBM Cloud Object Storage, the objects must be manipulated or managed as a whole unit. Therefore, objects cannot be accessed or updated partially.

IBM Spectrum Virtualize uses internal software components to support HTTP-based Representational State Transfer (REST) application programming interface (API) to interact with an external cloud service provider (CSP) or private cloud.

For more information about the IBM Cloud Object Storage portfolio, see this [web page](#).

Demonstration: The IBM Client Demonstration Center has a demonstration available at this [web page](#) (log in required).

10.3.1 Considerations for using Transparent Cloud Tiering

TCT can help to address certain business needs. When considering whether to use TCT, adopt a combination of business and technical views of the challenges and determine whether TCT can solve both of those needs.

The use of TCT can help businesses to manipulate data as shown in the following examples:

- ▶ Creating a consistent snapshot of dynamically changing data
- ▶ Creating a consistent snapshot of production data to facilitate data movement or migration between systems that are running at different locations
- ▶ Creating a snapshot of production data sets for application development and testing
- ▶ Creating a snapshot of production data sets for quality assurance
- ▶ Using secure data tiering to off-premises cloud providers

From a technical standpoint, ensure that you evaluate the network capacity and bandwidth requirements to support your data migration to off-premises infrastructure. To maximize productivity, you must match the amount of data that must be transmitted to the cloud plus your network capacity.

From a security standpoint, ensure that your on-premises or off-premises cloud infrastructure supports your requirements in terms of methods and level of encryption.

Regardless of your business needs, TCT within the IBM Spectrum Virtualize can provide opportunities to manage the exponential data growth and to manipulate data at low cost.

Today, many CSPs offers several *storage-as-services* solutions, such as content repository, backup, and archive. Combining all of these services, your IBM Spectrum Virtualize can help you solve many challenges that are related to rapid data growth, scalability, and manageability at attractive costs.

10.3.2 Transparent Cloud Tiering as backup solution and data migration

TCT can also be used as backup and data migration solution. In certain conditions, can be easily applied to eliminate the downtime that is associated with the needs to import and export data.

When TCT is applied as your backup strategy, IBM Spectrum Virtualize uses the same FlashCopy functions to produce *PiT* snapshot of an entire volume or set of volumes.

To ensure the integrity of the snapshot, it might be necessary to flush the host OS and application cache of any outstanding reads or writes before the snapshot is performed. Failing to flush the host OS and application cache can produce inconsistent and useless data.

Many OSs and applications provide mechanism to stop I/O operations and ensure that all data is flushed from host cache. If these mechanisms are available, they can be used in combination with snapshot operations. When these mechanisms are not available, it might be necessary to flush the cache manually by quiescing the application and unmounting the file system or logical drives.

When choosing cloud object storage as a backup solution, be aware that the object storage must be managed as a whole. Backup and restore of individual files, folders, and partitions, are not possible.

To interact with external CSPs or a private cloud, IBM Spectrum Virtualize requires interaction with the correct architecture and specific properties. Conversely, CSPs offer attractive prices for Object Storage in cloud and deliver an easy-to-use interface. Normally, cloud providers offer low-cost prices for Object Storage space, and charges are applied for the cloud outbound traffic only.

10.3.3 Restoring data by using Transparent Cloud Tiering

TCT can also be used to restore data from any snapshot that is stored in cloud providers. When the cloud accounts' technical and security requirements are met, the storage objects in the cloud can be used as a data recovery solution. The recovery method is similar to backup, except that the reverse direction is applied.

TCT running on IBM Spectrum Virtualize queries for Object Storage stored in a cloud infrastructure. It enables users to restore the objects into a new volume or set of volumes.

This approach can be used for various applications, such as recovering your production database application after an errant batch process that caused extensive damage.

Note: Always consider the bandwidth characteristics and network capabilities when choosing to use TCT.

Restoring individual files by using TCT is not possible. Object Storage is unlike a file or a block; therefore, Object Storage must be managed as a whole unit piece of storage, and not partially. Cloud Object Storage is accessible by using an HTTP-based REST API.

10.3.4 Transparent Cloud Tiering restrictions

The following restrictions must be considered before TCT is used:

- ▶ Because the Object Storage is normally accessed by using the HTTP protocol on top of a TCP/IP stack, all traffic that is associated with cloud service flows through the node management ports.
- ▶ The size of cloud-enabled volumes cannot change. If the size of the volume changes, a new snapshot must be created, so new Object Storage is constructed.
- ▶ TCT cannot be applied to volumes that are part of traditional copy services, such as FlashCopy, MM, GM, and HyperSwap.
- ▶ Volume containing two physical copies in two different storage pools cannot be part of TCT.
- ▶ Cloud Tiering snapshots cannot be taken from a volume that is part of migration activity across storage pools.
- ▶ Because VMware vSphere Virtual Volumes (VVOLs) are managed by a specific VMware application, these volumes are not candidates for TCT.
- ▶ File system volumes are not qualified for TCT.

10.4 Implementing Transparent Cloud Tiering

This section describes the steps and requirements to implement TCT by using IBM Spectrum Virtualize.

10.4.1 Domain Name System configuration

Because most of IBM Cloud Object Storage is managed and accessible by using the HTTP protocol, the Domain Name System (DNS) setting is an important requirement to ensure consistent resolution of domain names to internet resources.

Using your IBM Spectrum Virtualize management GUI, click **Settings** → **Network** → **DNS** and insert your DNS IPv4 or IPv6. The DNS name can be anything that you want, and is used as a reference. Click **Save** after you complete the choices, as shown in Figure 10-64.

IP Address:	Name:
0.0.0.0	dnsserver0

Figure 10-64 DNS settings

10.4.2 Enabling Transparent Cloud Tiering

After you complete the DNS settings, you can enable the TCT function in your IBM Spectrum Virtualize system by completing the following steps:

1. Using the IBM Spectrum Virtualize GUI, click **Settings** → **System** → **Transparent Cloud Tiering** and then, click **Enable Cloud Connection**, as shown in Figure 10-65. The TCT wizard starts and shows the welcome warning.

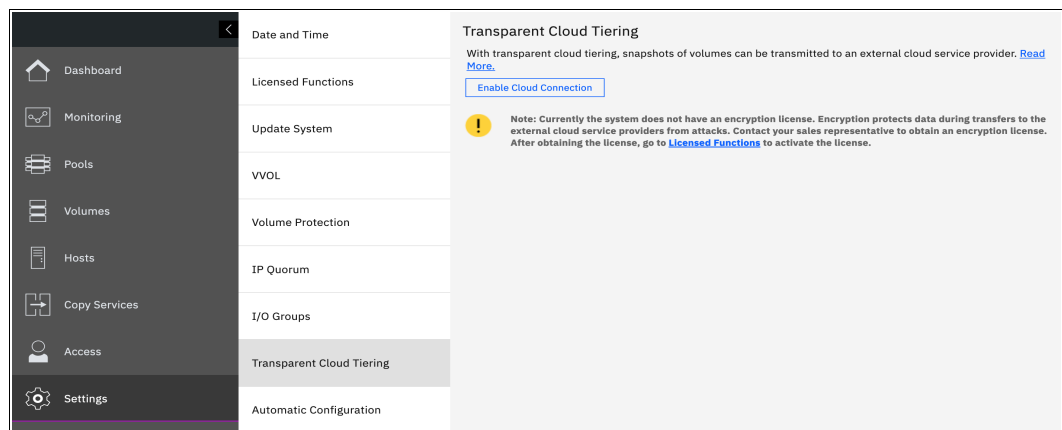


Figure 10-65 Enabling Cloud Tiering

Note: It is important to implement encryption before enabling cloud connecting. Encryption protects your data from attacks during the transfer to the external cloud service. Because the HTTP protocol is used to connect to cloud infrastructure, it is likely to start transactions by using the internet. For purposes of this writing, our system does not have encryption enabled.

2. Click **Next** to continue. You must select one of three CSPs:
 - IBM Cloud
 - OpenStack Swift
 - Amazon S3

Figure 10-66 shows the available options.

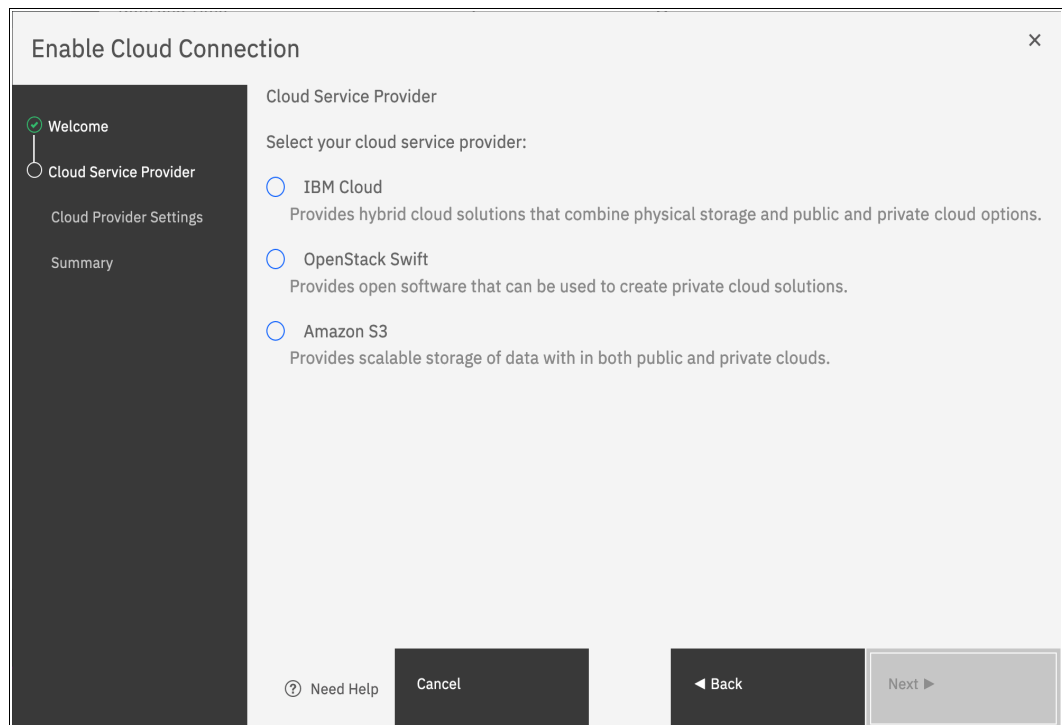


Figure 10-66 Selecting the cloud service provider

- In the next window, you must complete the settings of the cloud provider, credentials, and security access keys. The required settings can change depending on your CSP. An example of an empty form for an IBM Cloud connection is shown in Figure 10-67.

The screenshot shows a window titled "Enable Cloud Connection" with a sidebar on the left containing a progress indicator with four steps: "Welcome", "Cloud Service Provider", "Cloud Provider Settings" (which is currently selected), and "Summary". The main area is titled "Cloud Provider Settings" and contains two sections. The first section, "IBM Cloud account", includes input fields for "Object Storage URL:", "Tenant:", "User name:", and "API key:". Below the "API key" field is a checkbox labeled "Show characters". There is also a "Container prefix:" field and an "Encryption" section with an "Enable" checkbox. The second section, "Bandwidth:", has "Upload:" and "Download:" sub-sections. Each sub-section has a radio button for "No limit" (which is selected) and a radio button for "Limit to:" followed by an input field and the unit "Mbps". At the bottom of the window are three buttons: a "Cancel" button with a help icon, a "Back" button, and a "Next" button.

Figure 10-67 Entering the cloud service provider information

- Review your settings and click **Finish**, as shown in Figure 10-68.

The screenshot shows the same "Enable Cloud Connection" window, but now the "Summary" step is selected in the sidebar. The main area displays a summary of the configured settings in a table-like format. The settings are: Provider: OpenStack Swift; Endpoint: http://9.71.48.122:8080/auth/v1.0; Keystone: Disabled; Encryption: Disabled; Max Upload bandwidth: No limit; and Max Download bandwidth: No limit. At the bottom right, there are two buttons: "Back" and "Finish".

Figure 10-68 Cloud Connection summary

- The cloud credentials can be viewed and updated at any time by using the function icons in left side of the GUI and clicking **Settings** → **Systems** → **Transparent Cloud Tiering**. From this window, you can also verify the status, the data usage statistics, and the upload and download bandwidth limits set to support this function.

In the account information window, you can visualize your cloud account information. This window also enables you to remove the account.

An example of visualizing your cloud account information is shown in Figure 10-69.

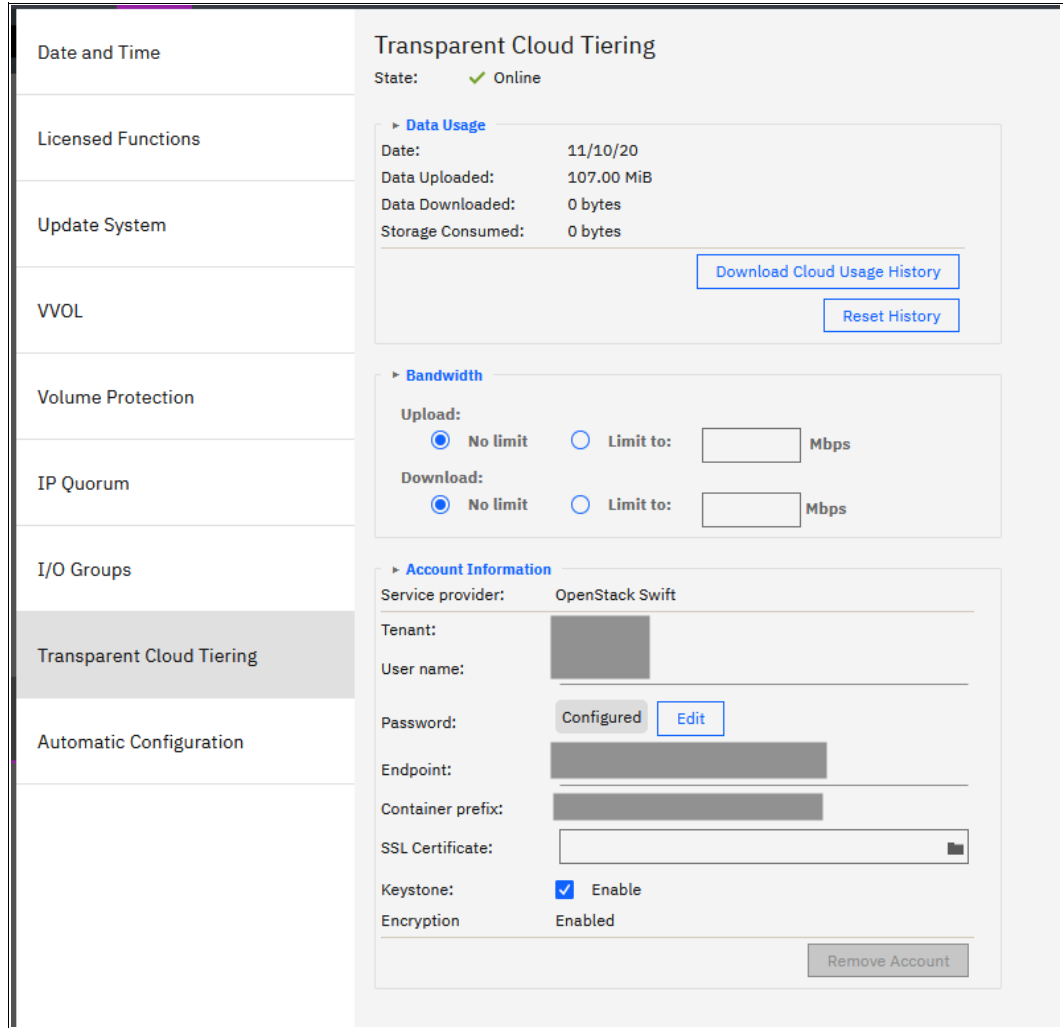


Figure 10-69 Enabled Transparent Cloud Tiering window

10.4.3 Creating cloud snapshots

To manage the cloud snapshots, the IBM Spectrum Virtualize provides a section in the GUI named Cloud Volumes. This section shows you how to add the volumes that are going to be part of the TCT. As described in 10.3.4, “Transparent Cloud Tiering restrictions” on page 623, cloud snapshot is available only for volumes that do not have a relationship to the list of restrictions previously mentioned.

Any volume can be added to the cloud volumes. However, snapshots work only for volumes that are not related to any other copy service.

To create and cloud snapshots, complete the following steps:

1. Click **Volumes** → **Cloud Volumes**, as shown in Figure 10-70.

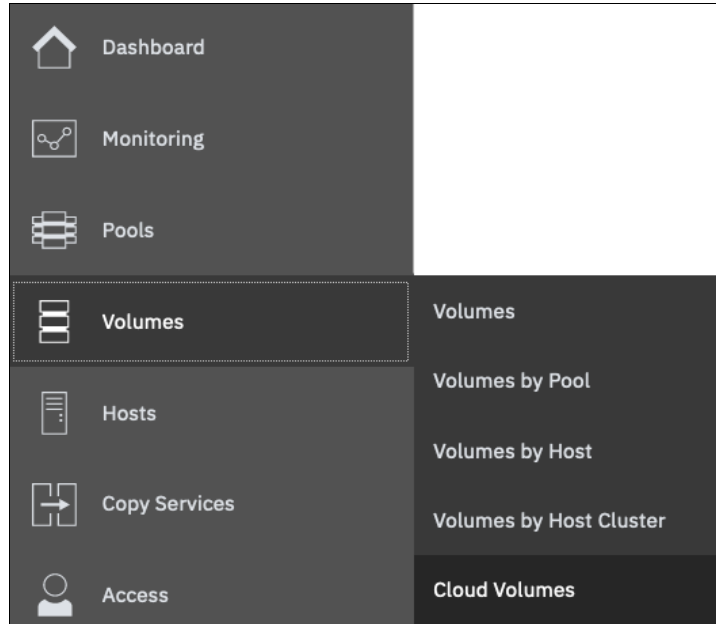


Figure 10-70 Cloud Volumes menu

2. A new window opens, and you can use the GUI to select one or more volumes that you need to enable a cloud snapshot or you can add volumes to the list, as shown in Figure 10-71.

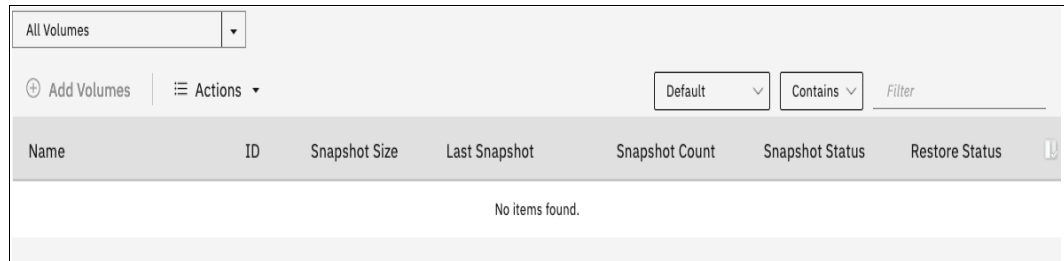


Figure 10-71 Cloud Volumes window

3. Click **Add Volumes** to enable cloud-snapshot on volumes. A new window opens, as shown in Figure 10-72 on page 629. Select the volumes that you want to enable TCT for and click **Next**.

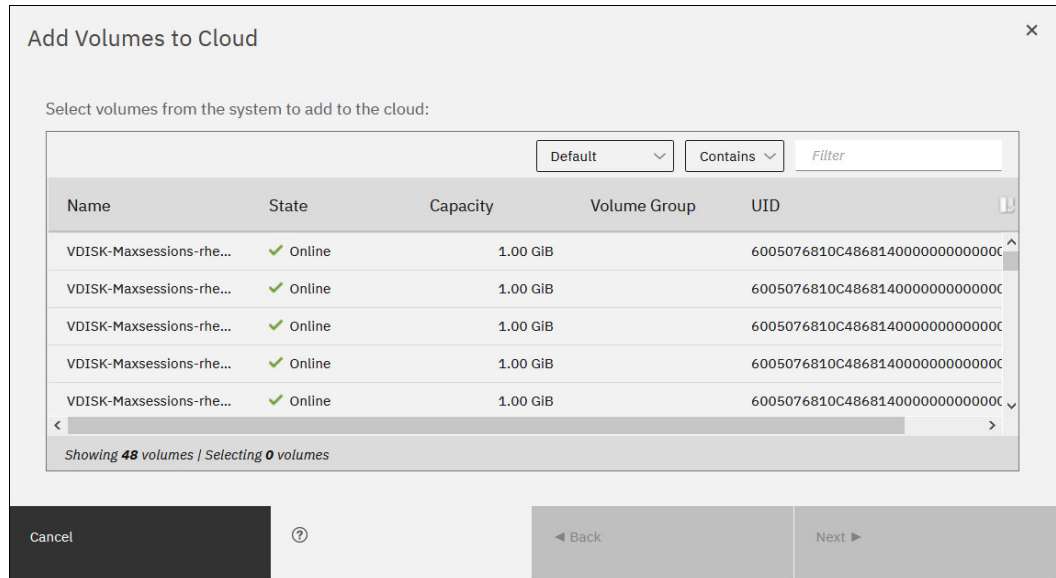


Figure 10-72 Add Volumes to Cloud

- IBM Spectrum Virtualize GUI provides two options for you to select. If the first option is selected, the system decides what type of snapshot is created based on previous objects for each selected volume. If a full copy (full snapshot) of a volume was created, the system makes an incremental copy of the volume.

The second option creates a full snapshot of one or more selected volumes. You can select the second option for a first occurrence of a snapshot and click **Finish**, as shown in Figure 10-73. You can also select the second option, even if another full copy of the volume exists.

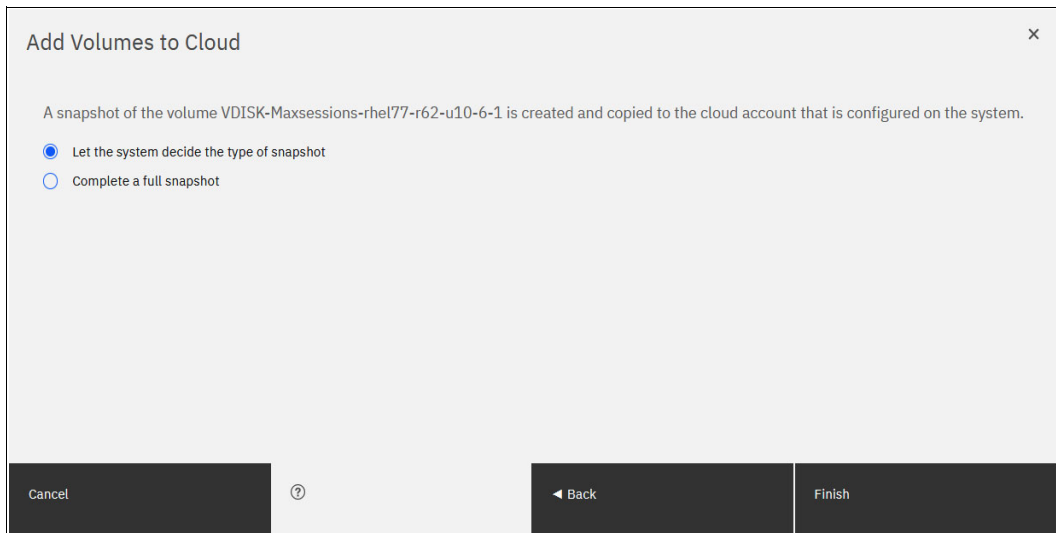


Figure 10-73 Selecting whether a full copy is made or whether the system decides

The **Cloud Volumes** window shows complete information about the volumes and their snapshots. The GUI shows the following information:

- Name of the volume
- ID of the volume assigned by the IBM Spectrum Virtualize
- Snapshot size
- Date and time that the last snapshot was created
- Number of snapshots that are taken for every volume
- Snapshot status
- Restore status
- Volume group for a set of volumes
- Volume unique identifier (UID)

Figure 10-74 shows an example of a Cloud Volumes list.

Name	ID	Snapshot Size ↓	Last Snapshot	Snapshot Count	Snapshot Status	Restore Status
VDISK-Maxsessions-rhel7...	66	2.98 MIB	11/2/2020 9:21:46 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	68	2.95 MIB	11/2/2020 9:21:51 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	34	2.87 MIB	11/2/2020 9:21:01 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	62	2.75 MIB	11/2/2020 9:21:26 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	60	2.66 MIB	11/2/2020 9:21:21 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	58	2.54 MIB	11/2/2020 9:21:11 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	33	2.40 MIB	11/2/2020 9:20:51 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	63	2.33 MIB	11/2/2020 9:21:36 PM	1	Ready	Available

Figure 10-74 Cloud Volumes list example

5. Click the **Actions** menu in the Cloud Volumes window to create and manage snapshots. Also, you can use the menu to cancel, disable, and restore snapshots to volumes, as shown in Figure 10-75.

Name	ID	Snapshot Size ↓	Last Snapshot	Snapshot Count	Snapshot Status	Restore Status
VDISK-Maxsessions-rhel7...	66	2.98 MIB	11/2/2020 9:21:46 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	68	2.95 MIB	11/2/2020 9:21:51 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	34	2.87 MIB	11/2/2020 9:21:01 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	62	2.75 MIB	11/2/2020 9:21:26 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	60	2.66 MIB	11/2/2020 9:21:21 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	58	2.54 MIB	11/2/2020 9:21:11 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	33	2.40 MIB	11/2/2020 9:20:51 PM	1	Ready	Available
VDISK-Maxsessions-rhel7...	63	2.33 MIB	11/2/2020 9:21:36 PM	1	Ready	Available

Figure 10-75 Available actions in Cloud Volumes window

10.4.4 Managing cloud snapshots

To manage volume cloud snapshots, open the Cloud Volumes window, right-click the volume that you want to manage the snapshots from, and select **Manage Cloud Snapshot**.

“Managing” a snapshot is deleting one or multiple versions. The list of PiT copies list of PiT copies appears and provide details about their status, type, and snapshot date, as shown in Figure 10-76 on page 631.

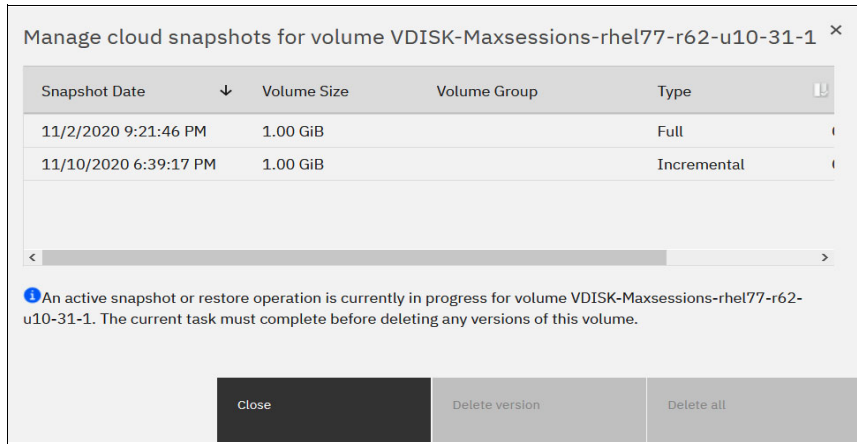


Figure 10-76 Deleting versions of a volume's snapshots

From this window, an administrator can delete old snapshots (old PiT copies) if they are no longer needed. The most recent copy cannot be deleted. If you want to delete the most recent copy, you must first disable Cloud Tiering for the specified volume.

10.4.5 Restoring cloud snapshots

This option allows IBM Spectrum Virtualize to restore snapshots from the cloud to the selected volumes or to create volumes with the restored data.

If the cloud account is shared among systems, IBM Spectrum Virtualize queries the snapshots that are stored in the cloud, and enables you to restore to a new volume. To restore a volume's snapshot, complete the following steps:

1. Open the Cloud Volumes window.
2. Right-click a volume and select **Restore**, as shown in Figure 10-77.

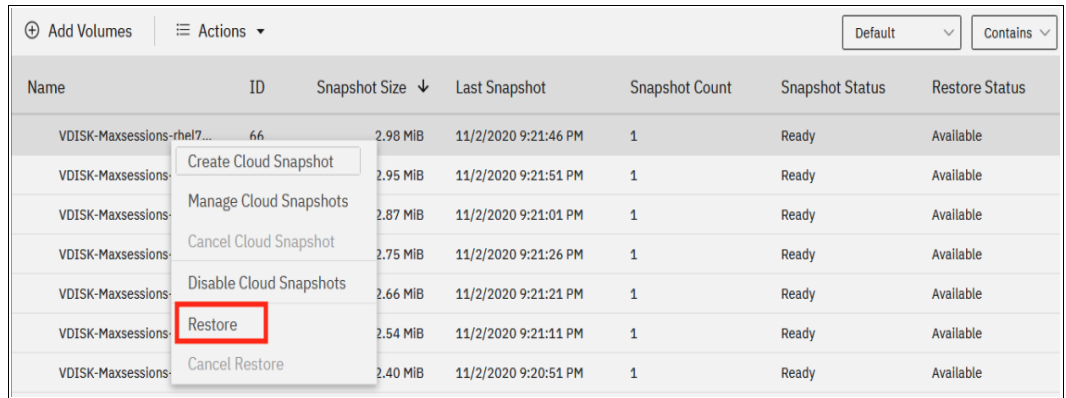


Figure 10-77 Selecting a volume to restore a snapshot from

3. A list of available snapshots is displayed. The snapshots date (PiT), their type (full or incremental), their state, and their size are shown (see Figure 10-78). Select the version that you want to restore and click **Next**.

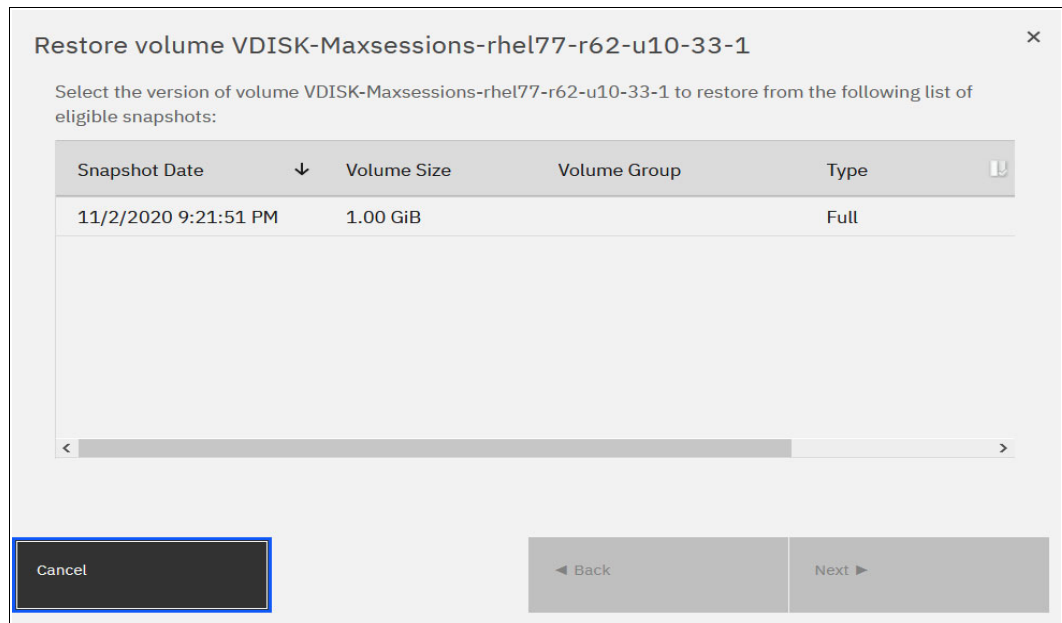


Figure 10-78 Selecting a snapshot version to restore

If the snapshot version that you selected has later generations (more recent Snapshot dates), the newer copies are removed from the cloud.

4. The IBM Spectrum Virtualize GUI provides two options to restore the snapshot from cloud. You can restore the snapshot from cloud directly to the selected volume, or create a volume to restore the data on, as shown in Figure 10-79. Make a selection and click **Next**.

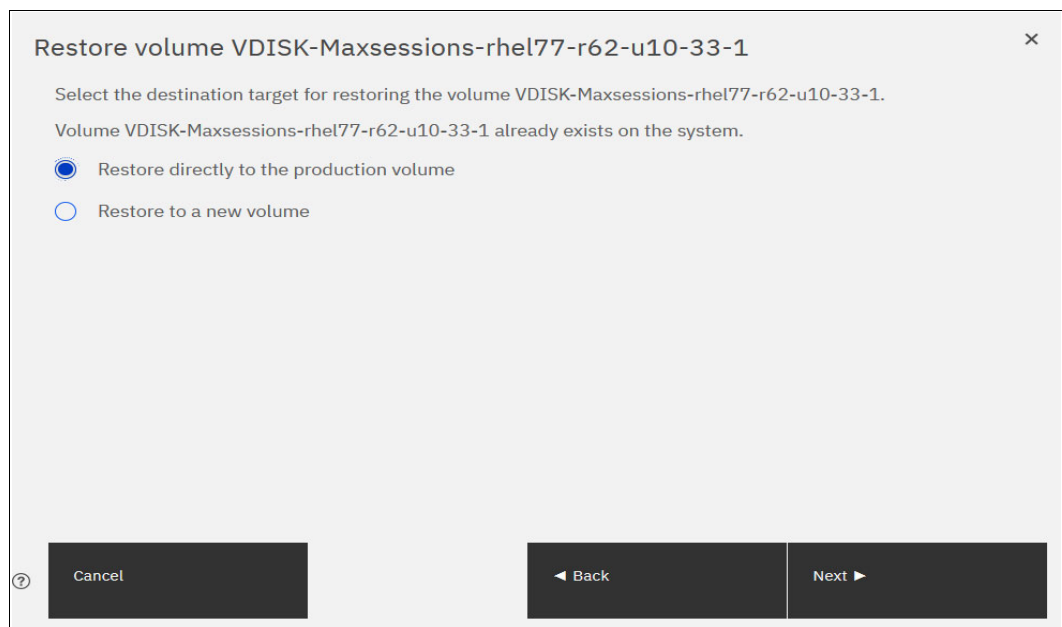


Figure 10-79 Restoring a snapshot on an existing volume or on a new volume

Note: Restoring a snapshot on the volume overwrites the data on the volume. The volume is taken offline (no read or write access) and the data from the PiT copy of the volume are written. The volume returns back online when all data is restored from the cloud.

5. If you selected the **Restore to a new Volume** option, you must enter the following information for the volume to be created with the snapshot data, as shown in Figure 10-80:
- Name
 - Storage Pool
 - Capacity Savings (None, Compressed or Thin-provisioned)
 - I/O group

You are not asked to enter the volume size because the new volume's size is identical to the snapshot copy size.

Enter the settings for the new volume and click **Next**.

Restore volume VDISK-Maxsessions-rhel77-r62-u10-33-1

Restore to a new volume

A new volume is created on the system to restore the data of this volume from the cloud.

Specify the settings for the new volume:

Name: Capacity savings: Deduplicated

Pool:

I/O group:

Cancel < Back Next >

Figure 10-80 Restoring a snapshot to a new volume

6. A Summary window is displayed so you can review the restoration settings, as shown in Figure 10-81. Click **Finish**. The system creates a volume or overwrites the selected volume. The more recent snapshots (later versions) of the volume are deleted from the cloud.

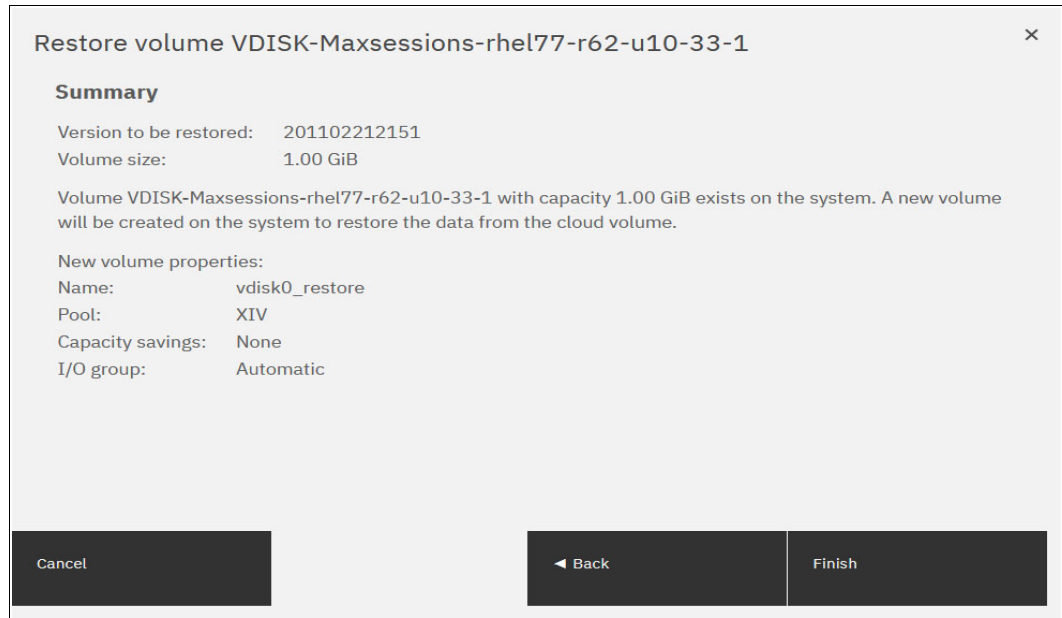


Figure 10-81 Restoring a snapshot summary

If you chose to restore the data from the cloud to a new volume, the new volume appears immediately in the volumes window. However, it is taken offline until all the data from the snapshot is written. The new volume is independent. It is not defined as a target in a FlashCopy mapping with the selected volume, for example. It also is not mapped to a host.

10.5 Volume mirroring and migration options

Volume mirroring is a simple redundant array of independent disks (RAID) 1 type function that enables a volume to remain online, even when the storage pool that is backing it becomes inaccessible. Volume mirroring is designed to protect the volume from storage infrastructure failures by seamless mirroring between storage pools.

Volume mirroring is provided by a specific volume mirroring function in the I/O stack. It cannot be manipulated like a FlashCopy or other types of copy volumes. However, this feature provides migration functions, which can be obtained by splitting the mirrored copy from the source or by using the *migrate to* function. Volume mirroring cannot control backend storage mirroring or replication.

With volume mirroring, host I/O completes when both copies are written. This feature is enhanced with a tunable latency tolerance. This tolerance provides an option to give preference to losing the redundancy between the two copies. This tunable timeout value is Latency or Redundancy.

The Latency tuning option, which is set by running the `chvdisk -mirrorwritepriority Latency` command, is the default. It prioritizes host I/O latency, which yields a preference to host I/O over availability. However, you might need to give preference to redundancy in your environment when availability is more important than I/O response time. Run the `chvdisk -mirrorwritepriority redundancy` command to set the redundancy option.

Regardless of which option you choose, volume mirroring can provide extra protection for your environment.

Migration offers the following options:

- ▶ Export to Image mode

By using this option, you can move storage from managed mode to image mode, which is useful if you use the IBM Spectrum Virtualize system as a migration device. For example, vendor A's product cannot communicate with vendor B's product, but you must migrate data from vendor A to vendor B. By using Export to Image mode, you can migrate data by using Copy Services functions and then return control to the native array while maintaining access to the hosts.

- ▶ Import to Image mode

By using this option, you can import a storage MDisk or logical unit number (LUN) with its data from an external storage system without putting metadata on it so that the data remains intact. After you import it, all copy services functions can be used to migrate the storage to other locations while the data remains accessible to your hosts.

- ▶ Volume migration by using volume mirroring and then by using Split into New Volume

By using this option, you can use the available RAID 1 functions. You create two copies of data that initially has a set relationship (one volume with two copies, one primary and one secondary) but then break the relationship (two volumes, both primary and no relationship between them) to make them independent copies of data.

You can use this option to migrate data between storage pools and devices. You might use this option if you want to move volumes to multiple storage pools. Each volume can have two copies at a time, which means that you can add only one copy to the original volume, and then you must split those copies to create another copy of the volume.

- ▶ Volume migration by using move to another pool

By using this option, you can move any volume between storage pools without any interruption to the host access. This option is a quicker version of the Volume Mirroring and Split into New Volume option. You might use this option if you want to move volumes in a single step, or you do not have a volume mirror copy.

Migration: Although these migration methods do not disrupt access, a brief outage does occur to install the host drivers for your IBM Spectrum Virtualize system if they are not yet installed.

With volume mirroring, you can move data to different MDisks within the same storage pool or move data between different storage pools. The use of volume mirroring over volume migration is beneficial because with volume mirroring, storage pools do not need to have the same extent size as is the case with volume migration.

Note: Volume mirroring does not create a second volume before you split copies. Volume mirroring adds a second copy of the data under the same volume. Therefore, you have one volume that is presented to the host with two copies of data that are connected to this volume. Only splitting copies creates another volume, and then both volumes have only one copy of the data.

Starting with Version 7.3 and the introduction of the dual-layer cache architecture, mirrored volume performance was improved. The lower cache is beneath the volume mirroring layer, which means that both copies have their own cache. This approach helps when you have copies of different types, for example, generic and compressed, because both copies use their independent cache and perform their own read prefetch. Destaging of the cache can be done independently for each copy, so one copy does not affect the performance of a second copy.

Also, because the IBM Spectrum Virtualize destage algorithm is MDisk aware, it can tune or adapt the destaging process, depending on MDisk type and usage, for each copy independently.

For more information about Volume Mirroring, see Chapter 6, “Volumes” on page 299.

10.6 Remote Copy

This section describes the Remote Copy (RC) services, which are a synchronous RC that is called MM, and two asynchronous RC options that are called GM and GMCV. RC in an IBM Spectrum Virtualize system is like RC in the IBM System Storage DS8000 family at a functional level, but the implementation differs.

IBM Spectrum Virtualize provides a single point of control when RC is enabled in your cluster (regardless of the disk subsystems that are used as underlying storage, if those disk subsystems are supported).

The general application of RC services is to maintain two real-time synchronized copies of a volume. Often, the two copies are geographically dispersed between two IBM Spectrum Virtualize systems. However, it is possible to use MM or GM within a single system (within an I/O group). If the master copy fails, you can enable an auxiliary copy for I/O operations.

Tips: Intracluster MM/GM uses more resources within the system when compared to an intercluster MM/GM relationship, where resource allocation is shared between the systems. Use intercluster MM/GM when possible. For mirroring volumes in the same system, it is better to use volume mirroring or the FlashCopy feature.

A typical application of this function is to set up a dual-site solution that uses two IBM Spectrum Virtualize systems. The first site is considered the *primary site* or *production site*, and the second site is considered the *backup site* or *failover site*. The failover site is activated when a failure at the first site is detected.

When MM or GM is used, a certain amount of bandwidth is required for the system intercluster heartbeat traffic. The amount of traffic depends on how many nodes are in each of the two clustered systems.

Table 10-10 on page 637 lists the amount of heartbeat traffic (in megabits per second (Mbps)) that is generated by various sizes of clustered systems.

Table 10-10 Intersystem heartbeat traffic in Mbps

IBM Spectrum Virtualize system 1	IBM Spectrum Virtualize system 2			
	2 nodes	4 nodes	6 nodes	8 nodes
2 nodes	5	6	6	6
4 nodes	6	10	11	12
6 nodes	6	11	16	17
8 nodes	6	12	17	21

10.6.1 IBM SAN Volume Controller and IBM FlashSystem system layers

An IBM Spectrum Virtualize based system can be in one of two layers: the *replication* layer or the *storage* layer. The layer that the system is in affects how the system interacts with other IBM Spectrum Virtualize based systems. IBM SAN Volume Controller (SVC) is always set to the replication layer. This parameter *cannot* be changed.

In the storage layer, an IBM FlashSystem system has the following characteristics and requirements:

- ▶ The system can perform MM and GM replication with other storage layer systems.
- ▶ The system can provide external storage for replication layer systems or SVC.
- ▶ The system cannot use a storage layer system as external storage.

In the replication layer, an SVC or an IBM FlashSystem system has the following characteristics and requirements:

- ▶ Can perform MM and GM replication with other replication layer systems.
- ▶ Cannot provide external storage for a replication layer system.
- ▶ Can use a storage layer system as external storage.

An IBM FlashSystem family system is in the storage layer by default, but the layer can be changed. For example, you might want to change an IBM FlashSystem 7200 to the replication layer if you want to virtualize other IBM FlashSystem systems or replicate to an SVC system.

Note: Before you change the system layer, the following conditions must be met on the system at the time of the layer change:

- ▶ No other IBM Spectrum Virtualize based system can exist as a back-end or host entity.
- ▶ No system partnerships can exist.
- ▶ No other IBM Spectrum Virtualize based system can be visible on the SAN fabric.

The layer can be changed during normal host I/O.

In your IBM FlashSystem system, run the `lssystem` command to check the current system layer, as shown in Example 10-2.

Example 10-2 Output from the `lssystem` command showing the system layer

```
IBM_IBM FlashSystem:GLTLoaner:superuser>lssystem
id 000002042160049E
name GLTLoaner
...
```

```
code_level 8.4.0.0 (build 152.19.2009281534000)
...
layer storage
...
```

Note: Consider the following rules for creating remote partnerships between the SVC and IBM FlashSystem systems:

- ▶ An SVC is always in the replication layer.
- ▶ By default, the IBM FlashSystem systems are in the storage layer, but can be changed to the replication layer.
- ▶ A system can form partnerships only with systems in the same layer.
- ▶ Starting in Version 6.4, any IBM Spectrum Virtualize based system in the replication layer can virtualize an IBM FlashSystem system in the storage layer.

10.6.2 Multiple IBM Spectrum Virtualize systems replication

Each IBM Spectrum Virtualize system can maintain up to three partner system relationships, which enables as many as four systems to be directly associated with each other. This system partnership capability enables the implementation of disaster recovery (DR) solutions.

Note: For more information about restrictions and limitations of native IP replication, see 10.8.2, “IP partnership limitations” on page 674.

Figure 10-82 on page 639 shows an example of a multiple-system mirroring configuration.

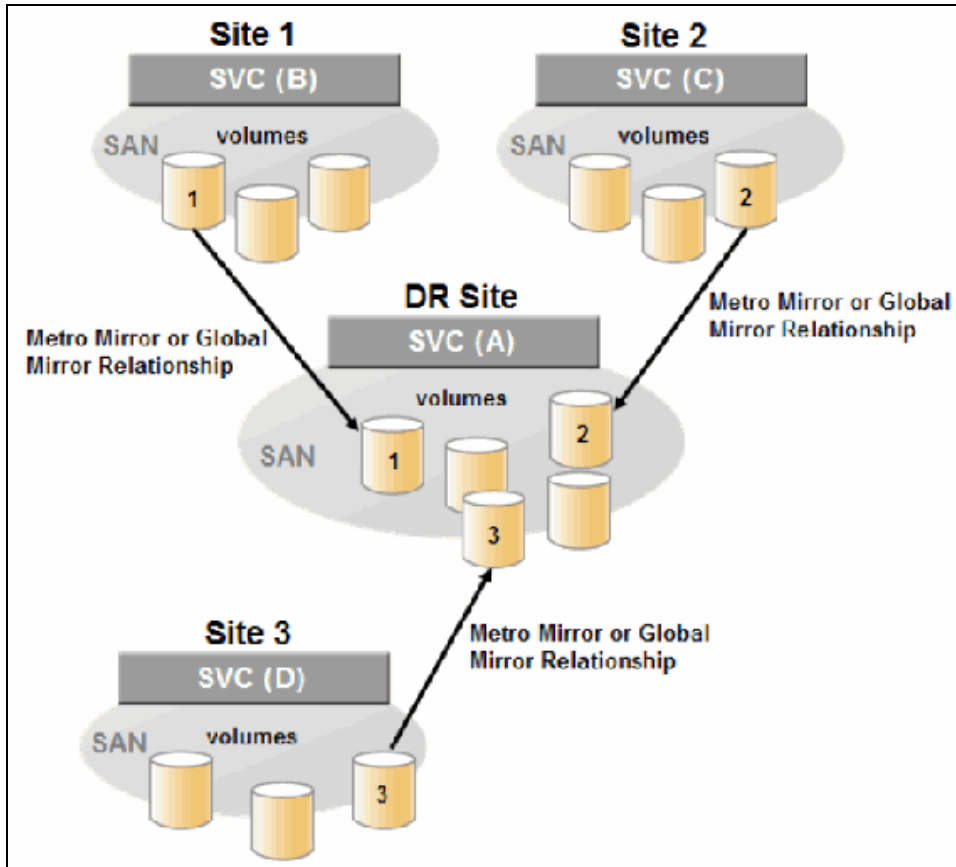


Figure 10-82 Multiple-system mirroring configuration example

Supported multiple-system mirroring topologies

Multiple-system mirroring supports various partnership topologies, as shown in the example in Figure 10-83. This example is a star topology (A → B, A → C, and A → D).

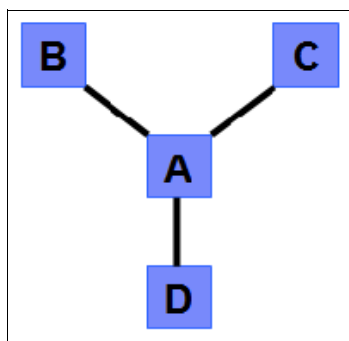


Figure 10-83 Star topology

Figure 10-83 shows four systems in a star topology, with System A at the center. System A can be a central DR site for the three other locations.

By using a star topology, you can migrate applications by using a process, such as the one described in the following example:

1. Suspend application at A.
2. Remove the A → B relationship.

3. Create the $A \rightarrow C$ relationship (or the $B \rightarrow C$ relationship).
4. Synchronize to system C, and ensure that $A \rightarrow C$ is established:
 - $A \rightarrow B$, $A \rightarrow C$, $A \rightarrow D$, $B \rightarrow C$, $B \rightarrow D$, and $C \rightarrow D$
 - $A \rightarrow B$, $A \rightarrow C$, and $B \rightarrow C$

Figure 10-84 shows an example of a triangle topology ($A \rightarrow B$, $A \rightarrow C$, and $B \rightarrow C$).

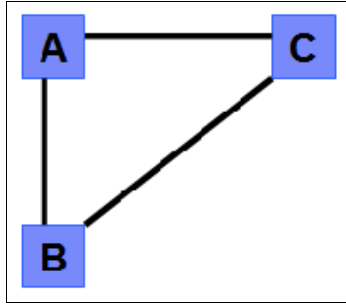


Figure 10-84 Triangle topology

Figure 10-85 shows an example of an IBM Spectrum Virtualize system fully connected topology ($A \rightarrow B$, $A \rightarrow C$, $A \rightarrow D$, $B \rightarrow D$, and $C \rightarrow D$).

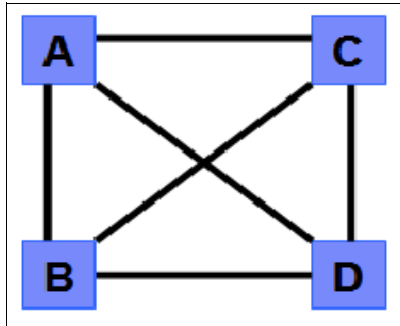


Figure 10-85 Fully connected topology

Figure 10-85 shows a fully connected mesh in which every system has a partnership to each of the three other systems. This topology enables volumes to be replicated between any pair of systems; for example, $A \rightarrow B$, $A \rightarrow C$, and $B \rightarrow C$.

Figure 10-86 shows a daisy-chain topology.

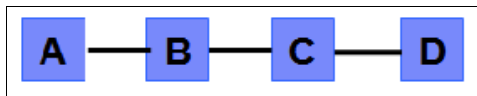


Figure 10-86 Daisy-chain topology

Although systems can have up to three partnerships, volumes can be part of only one RC relationship, for example, $A \rightarrow B$.

System partnership intermix: All these topologies are valid for the intermix of SVC with IBM FlashSystem if the IBM FlashSystem system is set to the replication layer.

IBM Spectrum Virtualize V8.3.1 introduced a three-site replication solution option that was expanded in Version 8.4. The solution enables active-active implementations while replicating to a third site. For more information, see 1.16.2, “Business continuity with three-site replication” on page 65, or for a detailed overview and configuration steps, see *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597.

10.6.3 Importance of write ordering

Many applications that use block storage are required to survive failures, such as loss of power or a software crash, and to not lose data that existed before the failure. Because many applications must perform many update operations in parallel, maintaining write ordering is key to ensure the correct operation of applications after a disruption.

An application that performs many database updates is designed with the concept of dependent writes. With dependent writes, it is important to ensure that an earlier write completed before a later write is started. Reversing or performing the order of writes differently than the application intended can undermine the application’s algorithms and can lead to problems, such as detected or undetected data corruption.

The IBM Spectrum Virtualize MM and GM implementations keep a consistent image at the secondary site. The GM implementation uses complex algorithms that identify sets of data and number those sets of data in sequence. Then, the data is applied at the secondary site in this same defined sequence.

Operating in this manner ensures that if the relationship is in a `Consistent_Synchronized` state, the GM target data is at least crash consistent and supports quick recovery through application crash recovery facilities.

For more information about dependent writes, see 10.1.13, “FlashCopy and image mode volumes” on page 577.

Remote Copy consistency groups

An RC consistency group can contain an arbitrary number of relationships up to the maximum number of MM/GM relationships that is supported by the IBM Spectrum Virtualize system. MM/GM commands can be issued to an RC consistency group.

Therefore, these commands can be issued simultaneously for all MM/GM relationships that are defined within that consistency group, or to a single MM/GM relationship that is not part of an RC consistency group. For example, when a `startrcconsistgrp` command is issued to the consistency group, all of the MM/GM relationships in the consistency group are started at the same time.

Figure 10-87 shows the concept of RC consistency groups.

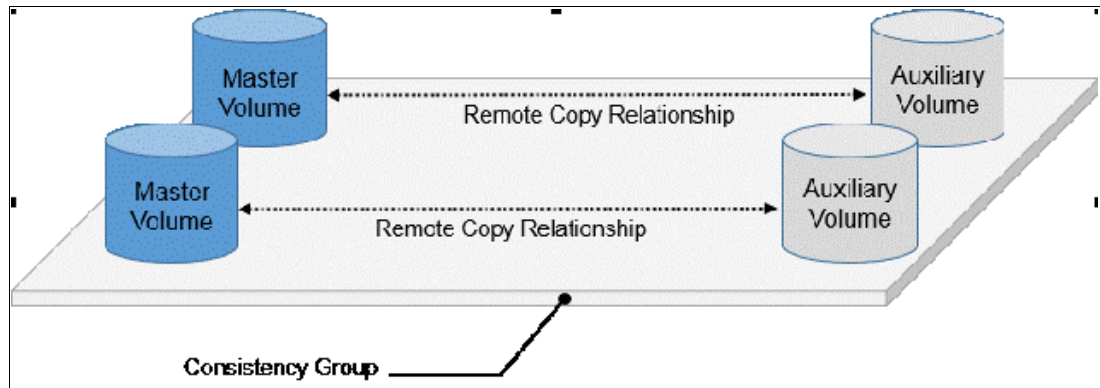


Figure 10-87 Remote Copy consistency group

Certain uses of MM/GM require the manipulation of more than one relationship. RC consistency groups can group relationships so that they are manipulated in unison.

Consider the following points:

- ▶ MM/GM relationships can be part of a consistency group, or they can be stand-alone and, therefore, are handled as single instances.
- ▶ A consistency group can contain zero or more relationships. An empty consistency group with zero relationships in it has little purpose until it is assigned its first relationship, except that it has a name.
- ▶ All relationships in a consistency group must have corresponding master and auxiliary volumes.
- ▶ All relationships in one consistency group must be the same type, for example, only MM or only GM.

Although consistency groups can be used to manipulate sets of relationships that do not need to satisfy these strict rules, this manipulation can lead to unwanted side effects. The rules behind a consistency group mean that certain configuration commands are prohibited. These configuration commands are not prohibited if the relationship is not part of a consistency group.

For example, consider the case of two applications that are independent, yet they are placed into a single consistency group. If an error occurs, synchronization is lost and a background copy process is required to recover synchronization. While this process is progressing, MM/GM rejects attempts to enable access to the auxiliary volumes of either application.

If one application finishes its background copy more quickly than the other application, MM/GM still refuses to grant access to its auxiliary volumes, even though it is safe in this case. The MM/GM policy is to refuse access to the entire consistency group if any part of it is inconsistent. Stand-alone relationships and consistency groups share a common configuration and state models. All of the relationships in a non-empty consistency group feature the same state as the consistency group.

10.6.4 Remote Copy intercluster communication

In the traditional Fibre Channel (FC), the intercluster communication between systems in a MM/GM partnership is performed over the SAN. This section describes this communication path.

For more information about intercluster communication between systems in an IP partnership, see 10.8.6, “States of IP partnership” on page 678.

Zoning

At least two FC ports of every node of each system must communicate with each other to create the partnership. Switch zoning is critical to facilitate intercluster communication.

Intercluster communication channels

When an IBM Spectrum Virtualize system partnership is defined on a pair of systems, the following intercluster communication channels are established:

- ▶ A single control channel, which is used to exchange and coordinate configuration information
- ▶ I/O channels between each of these nodes in the systems

These channels are maintained and updated as nodes and links appear and disappear from the fabric, and are repaired to maintain operation where possible. If communication between the systems is interrupted or lost, an event is logged (and the MM/GM relationships stop).

Alerts: You can configure the system to raise SNMP traps to the enterprise monitoring system to alert on events that indicate an interruption in internode communication occurred.

Intercluster links

All IBM Spectrum Virtualize nodes maintain a database of other devices that is visible on the fabric. This database is updated as devices appear and disappear.

Devices that advertise themselves as SVC or IBM FlashSystem nodes are categorized according to the system to which they belong. Nodes that belong to the same system establish communication channels between themselves and exchange messages to implement clustering and the functional protocols of IBM Spectrum Virtualize.

Nodes that are in separate systems do not exchange messages after initial discovery is complete, unless they are configured together to perform an RC relationship.

The intercluster link carries control traffic to coordinate activity between two systems. The link is formed between one node in each system. The traffic between the designated nodes is distributed among logins that exist between those nodes.

If the designated node fails (or all of its logins to the remote system fail), a new node is chosen to carry control traffic. This node change causes the I/O to pause, but it does not put the relationships in a `ConsistentStopped` state.

Note: Run the `chsystem` command with `-partnerfcportmask` to dedicate several FC ports only to system-to-system traffic to ensure that RC is not affected by other traffic, such as host-to-node traffic or node-to-node traffic within the same system.

10.6.5 Metro Mirror overview

MM establishes a synchronous relationship between two volumes of equal size. The volumes in an MM relationship are referred to as the *master* (primary) volume and the *auxiliary* (secondary) volume. Traditional FC MM is primarily used in a metropolitan area or geographical area, up to a maximum distance of 300 km (186.4 miles) to provide synchronous replication of data.

With synchronous copies, host applications write to the master volume, but they do not receive confirmation that the write operation completed until the data is written to the auxiliary volume. This action ensures that both the volumes have identical data when the copy completes. After the initial copy completes, the MM function always maintains a fully synchronized copy of the source data at the target site.

MM has the following characteristics:

- ▶ Zero recovery point objective (RPO)
- ▶ Synchronous
- ▶ Production application performance that is affected by round-trip latency

Increased distance directly affects host I/O performance because the writes are synchronous. Use the requirements for application performance when you are selecting your MM auxiliary location.

Consistency groups can be used to maintain data integrity for dependent writes, which is similar to FlashCopy consistency groups.

IBM Spectrum Virtualize provides intracluster and intercluster MM, which are described next.

Intracluster Metro Mirror

Intracluster MM performs the intracluster copying of a volume, in which both volumes belong to the same system and I/O group within the system. Because it is within the same I/O group, sufficient bitmap space must exist within the I/O group for both sets of volumes and licensing on the system.

Important: Performing MM across I/O groups within a system is not supported.

Intercluster Metro Mirror

Intercluster MM performs intercluster copying of a volume, in which one volume belongs to a system, and the other volume belongs to a separate system.

Two IBM Spectrum Virtualize systems must be defined in a partnership, which must be performed on both systems to establish a fully functional MM partnership.

By using standard single-mode connections, the supported distance between two systems in an MM partnership is 10 km (6.2 miles), although greater distances can be achieved by using extenders. For extended distance solutions, contact your IBM representative.

Limit: When a local fabric and a remote fabric are connected for MM purposes, the inter-switch link (ISL) hop count between a local node and a remote node cannot exceed seven.

10.6.6 Synchronous Remote Copy

MM is a fully synchronous RC technique that ensures that writes are committed at the master and auxiliary volumes before write completion is acknowledged to the host, but only if writes to the auxiliary volumes are possible.

Events, such as a loss of connectivity between systems, can cause mirrored writes from the master volume and the auxiliary volume to fail. In that case, MM suspends writes to the auxiliary volume and enables I/O to the master volume to continue to avoid affecting the operation of the master volumes.

Figure 10-88 shows how a write to the master volume is mirrored to the cache of the auxiliary volume before an acknowledgment of the write is sent back to the host that issued the write. This process ensures that the auxiliary is synchronized in real time if it is needed in a failover situation.

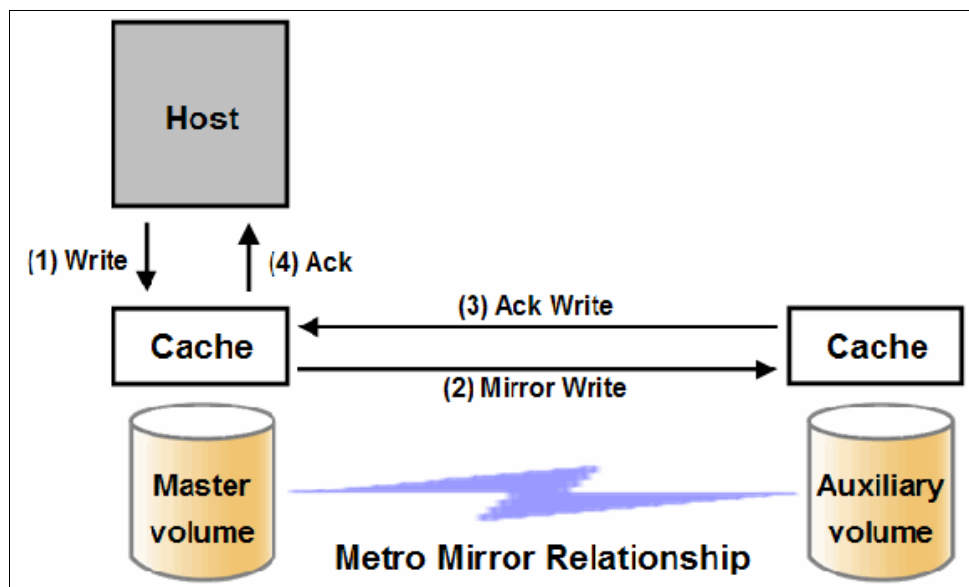


Figure 10-88 Write on volume in a Metro Mirror relationship

However, this process also means that the application is exposed to the latency and bandwidth limitations (if any) of the communication link between the master and auxiliary volumes. This process might lead to unacceptable application performance, particularly when placed under peak load. Therefore, the use of traditional FC MM has distance limitations that are based on your performance requirements. IBM Spectrum Virtualize does not support more than 300 km (186.4 miles).

10.6.7 Metro Mirror features

The IBM Spectrum Virtualize MM function supports the following features:

- ▶ Synchronous RC of volumes that are dispersed over metropolitan distances.
- ▶ MM relationships between volume pairs, with each volume in a pair that is managed by an IBM Spectrum Virtualize based system (requires Version 6.3.0 or later).
- ▶ Supports intracluster MM where both volumes belong to the same system (and I/O group).
- ▶ IBM Spectrum Virtualize supports intercluster MM where each volume belongs to a separate system. You can configure a specific system for partnership with another system.

All intercluster MM processing occurs between two IBM Spectrum Virtualize systems that are configured in a partnership.

- ▶ Intercluster and intracluster MM can be used concurrently.
- ▶ IBM Spectrum Virtualize does not require that a control network or fabric is installed to manage MM. For intercluster MM, the system maintains a control link between two systems. This control link is used to control the state and coordinate updates at either end. The control link is implemented on top of the same FC fabric connection that the system uses for MM I/O.
- ▶ IBM Spectrum Virtualize implements a configuration model that maintains the MM configuration and state through major events, such as failover, recovery, and resynchronization, to minimize user configuration action through these events.

IBM Spectrum Virtualize supports the resynchronization of changed data so that write failures that occur on the master or auxiliary volumes do not require a complete resynchronization of the relationship.

10.6.8 Metro Mirror attributes

The MM function in IBM Spectrum Virtualize features the following attributes:

- ▶ A partnership is created between two IBM Spectrum Virtualize systems that are operating in the replication layer (for intercluster MM).
- ▶ An MM relationship is created between two volumes of the same size.
- ▶ To manage multiple MM relationships as one entity, relationships can be made part of an MM consistency group, which ensures data consistency across multiple MM relationships and provides ease of management.
- ▶ When an MM relationship is started and when the background copy completes, the relationship becomes consistent and synchronized.
- ▶ After the relationship is synchronized, the auxiliary volume holds a copy of the production data at the primary, which can be used for DR.
- ▶ The auxiliary volume is in read-only mode when relationship is active.
- ▶ To access the auxiliary volume, the MM relationship must be stopped with the access option enabled before write I/O is allowed to the auxiliary.
- ▶ The remote host server is mapped to the auxiliary volume, and the disk is available for I/O.

10.6.9 Practical use of Metro Mirror

The master volume is the production volume, and updates to this copy are mirrored in real time to the auxiliary volume. The contents of the auxiliary volume that existed when the relationship was created are deleted.

Switching copy direction: The copy direction for an MM relationship can be switched so that the auxiliary volume becomes the master, and the master volume becomes the auxiliary, which is similar to the FlashCopy restore option. However, although the FlashCopy target volume can operate in read/write mode, the target volume of the started RC is always in read-only mode.

While the MM relationship is active, the auxiliary volume is not accessible for host application write I/O. The IBM Spectrum Virtualize based systems enable read-only access to the auxiliary volume when it contains a consistent image. They also allow boot time OS discovery to complete without an error so that any hosts at the secondary site can be ready to start the applications with a minimal delay if required.

For example, many OSs must read logical block address (LBA) zero to configure a logical unit (LU). Although read access is allowed at the auxiliary in practice, the data on the auxiliary volumes cannot be read by a host because most OSs write a “dirty bit” to the file system when it is mounted. Because this write operation is not allowed on the auxiliary volume, the volume cannot be mounted.

This access is provided only where consistency can be ensured. However, coherency cannot be maintained between reads that are performed at the auxiliary and later write I/Os that are performed at the master.

To enable access to the auxiliary volume for host operations, you must stop the MM relationship by specifying the `-access` parameter. While access to the auxiliary volume for host operations is enabled, the host must be instructed to mount the volume before the application can be started, or instructed to perform a recovery process.

For example, the MM requirement to enable the auxiliary copy for access differentiates it from third-party mirroring software on the host, which aims to emulate a single, reliable disk regardless of what system is accessing it. MM retains the property that there are two volumes in existence, but it suppresses one volume while the copy is being maintained.

The use of an auxiliary copy demands a conscious policy decision by the administrator that a failover is required, and that the tasks to be performed on the host that is involved in establishing the operation on the auxiliary copy are substantial. The goal is to make this copy rapid (much faster when compared to recovering from a backup copy) but not seamless.

The failover process can be automated by using failover management software. The IBM Spectrum Virtualize software provides SNMP traps and programming (or scripting) commands for the CLI to enable this automation.

10.6.10 Global Mirror overview

This section describes the GM copy service, which is an asynchronous RC service. This service provides and maintains a consistent mirrored copy of a source volume to a target volume.

GM function establishes a GM relationship between two volumes of equal size. The volumes in a GM relationship are referred to as the *master* (source) volume and the *auxiliary* (target) volume, which is the same as MM. Consistency groups can be used to maintain data integrity for dependent writes, which is similar to FlashCopy consistency groups.

GM writes data to the auxiliary volume asynchronously, which means that host writes to the master volume provide the host with confirmation that the write is complete before the I/O completes on the auxiliary volume.

GM has the following characteristics:

- ▶ Near-zero RPO
- ▶ Asynchronous
- ▶ Production application performance that is affected by I/O sequencing preparation time

Intracuster Global Mirror

Although GM is available for intracuster, it has no functional value for production use. Intracuster MM provides the same capability with less processor use. However, leaving this function in place simplifies testing and supports client experimentation and testing (for example, to validate server failover on a single test system). As with Intracuster MM, you must consider the increase in the license requirement because source and target exist on the same IBM Spectrum Virtualize system.

Intercluster Global Mirror

Intercluster GM operations require a pair of IBM Spectrum Virtualize systems that are connected by several intercluster links. The two systems must be defined in a partnership to establish a fully functional GM relationship.

Limit: When a local fabric and a remote fabric are connected for GM purposes, the ISL hop count between a local node and a remote node must not exceed seven hops.

10.6.11 Asynchronous Remote Copy

GM is an asynchronous RC technique. In asynchronous RC, the write operations are completed on the primary site and the write acknowledgment is sent to the host before it is received at the secondary site. An update of this write operation is sent to the secondary site at a later stage, which provides the capability to perform RC over distances that exceed the limitations of synchronous RC.

The GM function provides the same function as MM RC, but over long-distance links with higher latency without requiring the hosts to wait for the full round-trip delay of the long-distance link.

Figure 10-89 shows that a write operation to the master volume is acknowledged back to the host that is issuing the write before the write operation is mirrored to the cache for the auxiliary volume.

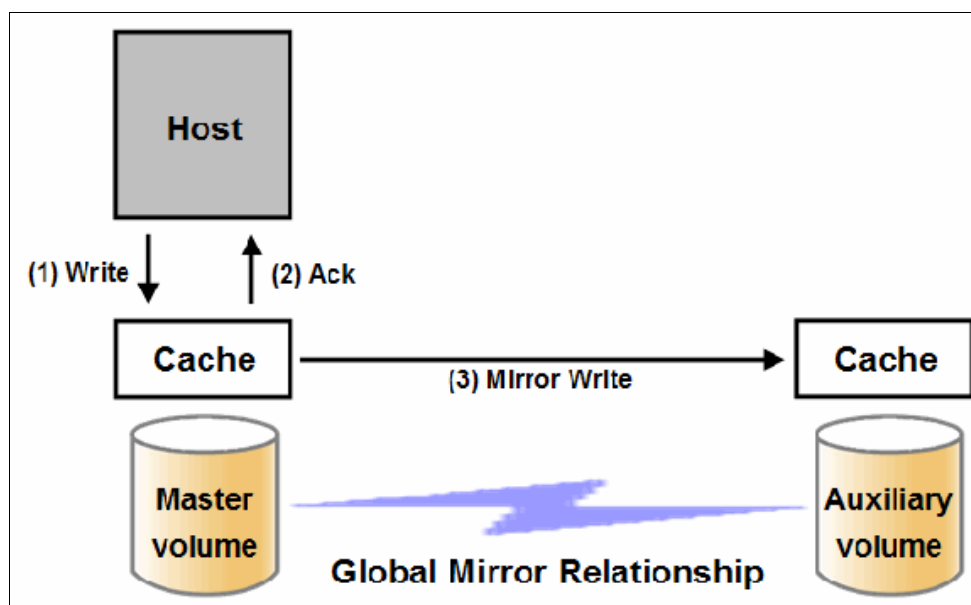


Figure 10-89 Global Mirror write sequence

The GM algorithms maintain a consistent image on the auxiliary. They achieve this consistent image by identifying sets of I/Os that are active concurrently at the master, assigning an order to those sets, and applying those sets of I/Os in the assigned order at the secondary. As a result, GM maintains the features of Write Ordering and Read Stability.

The multiple I/Os within a single set are applied concurrently. The process that marshals the sequential sets of I/Os operates at the secondary system. Therefore, the process is not subject to the latency of the long-distance link. These two elements of the protocol ensure that the throughput of the total system can be grown by increasing system size while maintaining consistency across a growing data set.

GM write I/O from production system to a secondary system requires serialization and sequence-tagging before being sent across the network to a remote site (to maintain a write-order consistent copy of data).

To avoid affecting the production site, IBM Spectrum Virtualize supports more parallelism in processing and managing GM writes on the secondary system by using the following methods:

- ▶ Secondary system nodes store replication writes in new redundant non-volatile cache
- ▶ Cache content details are shared between nodes
- ▶ Cache content details are batched together to make node-to-node latency less of an issue
- ▶ Nodes intelligently apply these batches in parallel as soon as possible
- ▶ Nodes internally manage and optimize GM secondary write I/O processing

In a failover scenario where the secondary site must become the master source of data, specific updates might be missing at the secondary site. Therefore, any applications that use this data must have an external mechanism for recovering the missing updates and reapplying them, such as a transaction log replay.

GM is supported over FC, Fibre Channel over IP (FCIP), Fibre Channel over Ethernet (FCoE), and native IP connections. The maximum distance cannot exceed 80 ms round trip, which is approximately 4000 km (2485.48 miles) between mirrored systems. However, starting with IBM Spectrum Virtualize V7.4, this distance was increased to 250 ms for certain configurations. Figure 10-90 shows the supported round-trip distances for GM RC.

System Hardware	Partnership		
	FC	1Gbps - IP	10 Gbps - IP
SVC DH8	250ms	80ms	10ms
SVC SV1	250ms	80ms	10ms

Figure 10-90 Supported Global Mirror link latencies

10.6.12 Global Mirror features

IBM Spectrum Virtualize GM supports the following features:

- ▶ Asynchronous RC of volumes that are dispersed over metropolitan-scale distances.
- ▶ IBM Spectrum Virtualize implements the GM relationship between a volume pair, with each volume in the pair being managed by an IBM Spectrum Virtualize system.
- ▶ IBM Spectrum Virtualize supports intracluster GM where both volumes belong to the same system (and I/O group).

- ▶ An IBM Spectrum Virtualize system can be configured for partnership with 1 - 3 other systems. For more information about IP partnership restrictions, see 10.8.2, “IP partnership limitations” on page 674.
- ▶ Intercluster and intracluster GM can be used concurrently, but not for the same volume.
- ▶ IBM Spectrum Virtualize does not require a control network or fabric to be installed to manage GM. For intercluster GM, the system maintains a control link between the two systems. This control link is used to control the state and to coordinate the updates at either end. The control link is implemented on top of the same FC fabric connection that the system uses for GM I/O.
- ▶ IBM Spectrum Virtualize implements a configuration model that maintains the GM configuration and state through major events, such as failover, recovery, and resynchronization, to minimize user configuration action through these events.
- ▶ IBM Spectrum Virtualize implements flexible resynchronization support, enabling it to resynchronize volume pairs that experienced write I/Os to both disks, and to resynchronize only those regions that changed.
- ▶ An optional feature for GM is a delay simulation to be applied on writes that are sent to auxiliary volumes. It is useful in intracluster scenarios for testing purposes.

Colliding writes

The GM algorithm requires that only a single write is active on a volume. I/Os that overlap an active I/O are sequential, which is called *colliding writes*. If another write is received from a host while the auxiliary write is still active, the new host write is delayed until the auxiliary write is complete. This rule is needed if a series of writes to the auxiliary must be tried again and is called *reconstruction*. Conceptually, the data for reconstruction comes from the master volume.

If multiple writes are allowed to be applied to the master for a sector, only the most recent write gets the correct data during reconstruction. If reconstruction is interrupted for any reason, the intermediate state of the auxiliary is inconsistent. Applications that deliver such write activity do not achieve the performance that GM is intended to support. A volume statistic is maintained about the frequency of these collisions.

An attempt is made to allow multiple writes to a single location to be outstanding in the GM algorithm. Master writes must still be sequential, and the intermediate states of the master data must be kept in a non-volatile journal while the writes are outstanding to maintain the correct write ordering during reconstruction. Reconstruction must never overwrite data on the auxiliary with an earlier version. The volume statistic that is monitoring colliding writes is now limited to those writes that are not affected by this change.

Figure 10-91 on page 651 shows a colliding write sequence example.

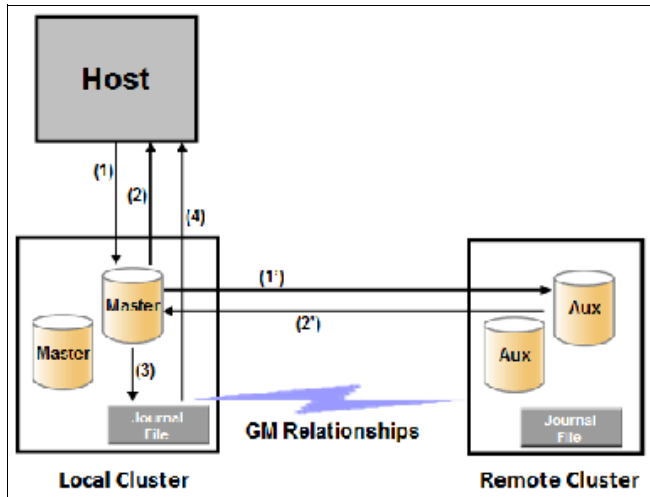


Figure 10-91 Colliding writes example

The following numbers correspond to the numbers that are shown in Figure 10-91:

- ▶ (1) The first write is performed from the host to LBA X.
- ▶ (2) The host is provided acknowledgment that the write completed, even though the mirrored write to the auxiliary volume is not yet complete.
- ▶ (1') and (2') occur asynchronously with the first write.
- ▶ (3) The second write is performed from the host also to LBA X. If this write occurs before (2'), the write is written to the journal file.
- ▶ (4) The host is provided acknowledgment that the second write is complete.

Delay simulation

GM provides a feature that enables a delay simulation to be applied on writes that are sent to the auxiliary volumes. With this feature, tests can be done to detect colliding writes. It also provides the capability to test an application before the full deployment. The feature can be enabled separately for each of the intracluster or intercluster GMs.

By running the `chsystem` command, the delay setting can be set up and the delay can be checked by running the `lssystem` command. The `gm_intra_cluster_delay_simulation` field expresses the amount of time that intracluster auxiliary I/Os are delayed. The `gm_inter_cluster_delay_simulation` field expresses the amount of time that intercluster auxiliary I/Os are delayed. A value of zero disables the feature.

Tip: If you are experiencing repeated problems with the delay on your link, ensure that the delay simulator was correctly disabled.

10.6.13 Using Global Mirror with Change Volumes

GM is designed to achieve an RPO as low as possible so that data is as up to date as possible. This design places several strict requirements on your infrastructure. In certain situations with low network link quality, congested hosts, or overloaded hosts, you might be affected by multiple 1920 congestion errors.

Congestion errors occur in the following primary situations:

- ▶ At the source site through the host or network
- ▶ In the network link or network path
- ▶ At the target site through the host or network

GM has functions that are designed to address the following conditions, which might negatively affect certain GM implementations:

- ▶ The estimation of the bandwidth requirements tends to be complex.
- ▶ Ensuring that the latency and bandwidth requirements can be met is often difficult.
- ▶ Congested hosts on the source or target site can cause disruption.
- ▶ Congested network links can cause disruption with only intermittent peaks.

To address these issues, change volumes were added as an option for GM relationships. Change volumes use the FlashCopy function, but they cannot be manipulated as FlashCopy volumes because they are for a special purpose only. Change volumes replicate PiT images on a cycling period. The default is 300 seconds.

Your change rate must include only the condition of the data at the PiT that the image was taken, rather than all the updates during the period. The use of this function can provide significant reductions in replication volume.

GMCV has the following characteristics:

- ▶ Larger RPO
- ▶ PiT copies
- ▶ Asynchronous
- ▶ Possible system performance resource requirements because PiT copies are created locally

Figure 10-92 shows a simple GM relationship without change volumes.

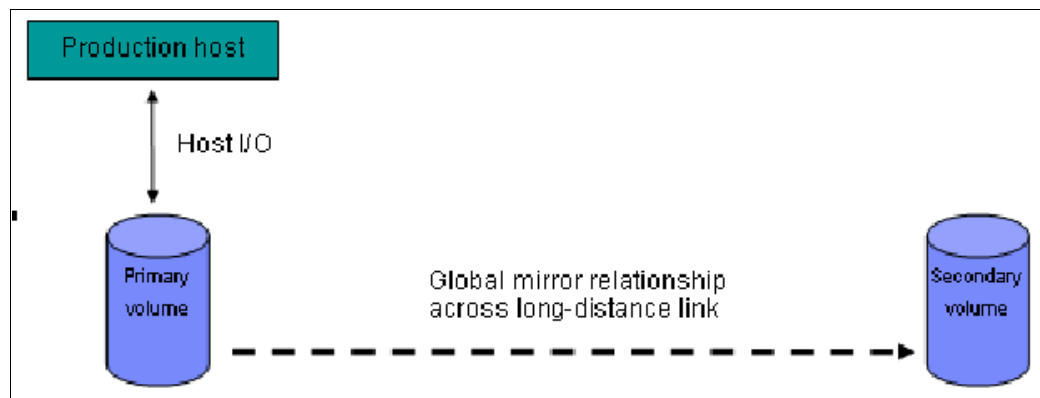


Figure 10-92 Global Mirror without Change Volumes

With change volumes, this environment looks as it is shown in Figure 10-93 on page 653.

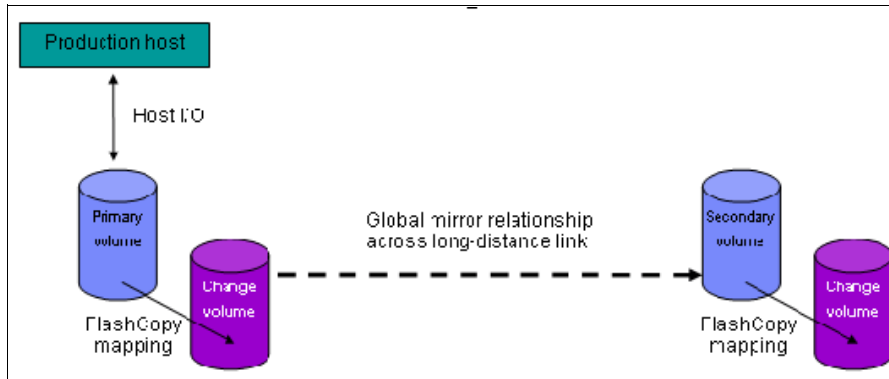


Figure 10-93 Global Mirror with Change Volumes

With change volumes, a FlashCopy mapping exists between the primary volume and the primary change volume. The mapping is updated in the cycling period (60 seconds - 1 day). The primary change volume is then replicated to the secondary GM volume at the target site, which is then captured in another change volume on the target site. This approach provides an always consistent image at the target site and protects your data from being inconsistent during resynchronization.

For more information about IBM FlashCopy, see 10.1, “IBM FlashCopy” on page 554.

You can adjust the cycling period by running the `chrcrelationship -cycleperiodseconds <60 - 86400>` command from the CLI. The default value is 300 seconds. If a copy does not complete in the cycle period, the next cycle does not start until the prior cycle completes. For this reason, the use of change volumes gives you the following possibilities for RPO:

- ▶ If your replication completes in the cycling period, your RPO is twice the cycling period.
- ▶ If your replication does not complete within the cycling period, RPO is twice the completion time. The next cycling period starts immediately after the prior cycling period is finished.

Carefully consider your business requirements versus the performance of GMCV. GMCV increases the intercluster traffic for more frequent cycling periods. Therefore, selecting the shortest cycle periods possible is not always the answer. In most cases, the default must meet requirements and perform well.

Important: When you create your GM volumes with change volumes, ensure that you remember to select the change volume on the auxiliary (target) site. Failure to do so leaves you exposed during a resynchronization operation.

10.6.14 Distribution of work among nodes

For the best performance, MM/GM volumes must have their preferred nodes evenly distributed among the nodes of the systems. Each volume within an I/O group has a preferred node property that can be used to balance the I/O load between nodes in that group. MM/GM also uses this property to route I/O between systems.

If this best practice is not maintained, such as if source volumes are assigned to only one node in the I/O group, you can change the preferred node for each volume to distribute volumes evenly between the nodes. You can also change the preferred node for volumes that are in an RC relationship without affecting the host I/O to a particular volume.

The RC relationship type does not matter. The RC relationship type can be MM, GM, or GMCV. You can change the preferred node both to the source and target volumes that are participating in the RC relationship.

10.6.15 Background copy performance

The background copy performance is subject to sufficient RAID controller bandwidth. Performance is also subject to other potential bottlenecks, such as the intercluster fabric, and possible contention from host I/O for the IBM Spectrum Virtualize system bandwidth resources.

Background copy I/O is scheduled to avoid bursts of activity that might have an adverse effect on system behavior. An entire grain of tracks on one volume is processed at around the same time, but not as a single I/O.

Double buffering is used to try to use sequential performance within a grain. However, the next grain within the volume might not be scheduled for some time. Multiple grains might be copied simultaneously, and might be enough to satisfy the requested rate, unless the available resources cannot sustain the requested rate.

GM paces the rate at which background copy is performed by the appropriate relationships. Background copy occurs on relationships that are in the `InconsistentCopying` state with a status of `Online`.

The quota of background copy (configured on the intercluster link) is divided evenly between all nodes that are performing background copy for one of the eligible relationships. This allocation is made irrespective of the number of disks for which the node is responsible. Each node in turn divides its allocation evenly between the multiple relationships that are performing a background copy.

The default value of the background copy is 25 megabytes per second (MBps) per volume.

Important: The background copy value is a system-wide parameter that can be changed dynamically, but only on a *per-system* basis and not on a *per-relationship* basis. Therefore, the copy rate of all relationships changes when this value is increased or decreased. In systems with many RC relationships, increasing this value might affect overall system or intercluster link performance. The background copy rate can be changed to 1 - 1000 MBps.

10.6.16 Thin-provisioned background copy

MM/GM relationships preserve the space-efficiency of the master. Conceptually, the background copy process detects a deallocated region of the master and sends a special *zero buffer* to the auxiliary.

If the auxiliary volume is thin-provisioned and the region is deallocated, the special buffer prevents a write and therefore, an allocation. If the auxiliary volume is not thin-provisioned or the region in question is an allocated region of a thin-provisioned volume, a buffer of “real” zeros is synthesized on the auxiliary and written as normal.

10.6.17 Methods of synchronization

This section describes two methods that can be used to establish a synchronized relationship.

Full synchronization after creation

The full synchronization after creation method is the default method. It is the simplest method in that it requires no administrative activity apart from running the necessary commands. However, in certain environments, the available bandwidth can make this method unsuitable.

Run the following command sequence for a single relationship:

1. Run `mkrcrelationship` without specifying the `-sync` option.
2. Run `starttrcrelationship` without specifying the `-clean` option.

Synchronized before creation

In this method, the administrator must ensure that the master and auxiliary volumes contain identical data before creating the relationship by using the following technique:

- ▶ Both disks are created with the security delete feature to make all data zero.
- ▶ A complete tape image (or other method of moving data) is copied from one disk to the other disk.

With this technique, do not allow I/O on the master or auxiliary before the relationship is established. Then, the administrator must run the following commands:

1. Run `mkrcrelationship` with the `-sync` flag.
2. Run `starttrcrelationship` without the `-clean` flag.

Important: Failure to perform these steps correctly can cause MM/GM to report the relationship as consistent when it is not. This use can cause loss of a data or data integrity exposure for hosts that are accessing data on the auxiliary volume.

10.6.18 Practical use of Global Mirror

The practical use of GM is similar to MM, as described in 10.6.9, “Practical use of Metro Mirror” on page 646. The main difference between the two RC modes is that GM and GMCV are mostly used on much larger distances than MM. Weak link quality or insufficient bandwidth between the primary and secondary sites can also be a reason to prefer asynchronous GM over synchronous MM. Otherwise, the use cases for MM/GM are the same.

10.6.19 IBM Spectrum Virtualize HyperSwap topology

The IBM HyperSwap topology is based on IBM Spectrum Virtualize RC mechanisms. It is also referred to as an “active-active relationship” in this document.

You can create an HyperSwap topology system configuration where each I/O group in the system is physically on a different site. These configurations can be used to maintain access to data on the system when power failures or site-wide outages occur.

In a HyperSwap configuration, each site is defined as an independent failure domain. If one site experiences a failure, the other site can continue to operate without disruption. You must also configure a third site to host a quorum device or IP quorum application that provides an automatic tie-breaker in case of a link failure between the two main sites. The main site can be in the same room or across rooms in the data center, buildings on the same campus, or buildings in different cities. Different kinds of sites protect against different types of failures.

For more information about HyperSwap implementation and best practices, see *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597.

10.6.20 Consistency Protection for Global Mirror and Metro Mirror

MM, GM, GMCV, and HyperSwap Copy Services functions create RC or remote replication relationships between volumes or consistency groups. If the secondary volume in a Copy Services relationship becomes unavailable to the primary volume, the system maintains the relationship. However, the data might become out of sync when the secondary volume becomes available.

Since V7.8, it is possible to create a FlashCopy mapping (change volume) for an RC target volume to maintain a consistent image of the secondary volume. The system recognizes it as a *Consistency Protection* and a link failure or an offline secondary volume event is handled differently now.

When Consistency Protection is configured, the relationship between the primary and secondary volumes does not stop if the link goes down or the secondary volume is offline. The relationship does not go in to the consistent stopped status. Instead, the system uses the secondary change volume to automatically copy the previous consistent state of the secondary volume. The relationship automatically moves to the consistent copying status as the system resynchronizes and protects the consistency of the data. The relationship status changes to `consistent_synchronized` when the resynchronization process completes. The relationship automatically resumes replication after the temporary loss of connectivity.

Change volumes that are used for Consistency Protection are not visible and manageable from the GUI because they are used for Consistency Protection internal behavior only.

It is not required to configure a secondary change volume on a MM/GM (without cycling) relationship. However, if the link goes down or the secondary volume is offline, the relationship goes in to the `Consistent_stopped` status. If write operations occur on the primary or secondary volume, the data is no longer synchronized (out of sync).

Consistency protection must be enabled on all relationships in a consistency group. Every relationship in a consistency group must be configured with a secondary change volume. If a secondary change volume is not configured on one relationship, the entire consistency group stops with a 1720 error if host I/O is processed when the link is down or any secondary volume in the consistency group is offline. All relationships in the consistency group are unable to retain a consistent copy during resynchronization.

The option to add consistency protection is selected by default when MM/GM relationships are created. The option must be cleared to create MM/GM relationships without consistency protection.

10.6.21 Valid combinations of FlashCopy, Metro Mirror, and Global Mirror

Table 10-11 lists the combinations of FlashCopy and MM/GM functions that are valid for a single volume.

Table 10-11 Valid combinations for a single volume

FlashCopy	MM or GM source	MM or GM target
FlashCopy Source	Supported	Supported
FlashCopy Target	Supported	Not supported

10.6.22 Remote Copy configuration limits

Table 10-12 lists the MM/GM configuration limits.

Table 10-12 Metro Mirror configuration limits

Parameter	Value
Number of Metro Mirror or GM consistency groups per system	256
Number of Metro Mirror or GM relationships per system	10000
Number of Metro Mirror or GM relationships per consistency group	10000
Total Volume size per I/O group	A per I/O group limit of 1024 TB exists on the quantity of master and auxiliary volume address spaces that can participate in Metro Mirror and GM relationships. This maximum configuration uses all 512 MiB of bitmap space for the I/O group and allows 10 MiB of space for all remaining copy services features.

10.6.23 Remote Copy states and events

This section describes the various states of a MM/GM relationship and the conditions that cause them to change. In Figure 10-94 shows an overview of the status that can apply to a MM/GM relationship in a connected state.

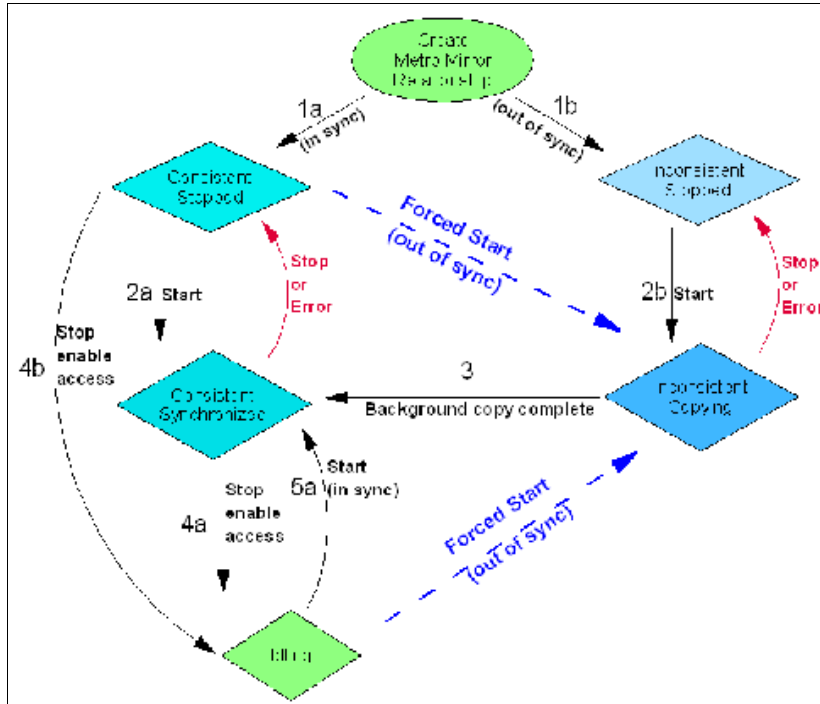


Figure 10-94 Metro Mirror or Global Mirror mapping state diagram

When the MM/GM relationship is created, you can specify whether the auxiliary volume is in sync with the master volume, and the background copy process is then skipped. This capability is useful when MM/GM relationships are established for volumes that were created with the format option.

The following step identifiers are shown in Figure 10-94:

- ▶ Step 1:
 - a. The MM/GM relationship is created with the **-sync** option, and the MM/GM relationship enters the `ConsistentStopped` state.
 - b. The MM/GM relationship is created without specifying that the master and auxiliary volumes are in sync, and the MM/GM relationship enters the `InconsistentStopped` state.
- ▶ Step 2:
 - a. When an MM/GM relationship is started in the `ConsistentStopped` state, the MM/GM relationship enters the `ConsistentSynchronized` state. Therefore, no updates (write I/O) were performed on the master volume while in the `ConsistentStopped` state. Otherwise, the **-force** option must be specified, and the MM/GM relationship then enters the `InconsistentCopying` state while the background copy is started.
 - b. When an MM/GM relationship is started in the `InconsistentStopped` state, the MM/GM relationship enters the `InconsistentCopying` state while the background copy is started.

- ▶ Step 3

When the background copy completes, the MM/GM relationship changes from the `InconsistentCopying` state to the `ConsistentSynchronized` state.
- ▶ Step 4:
 - a. When a MM/GM relationship is stopped in the `ConsistentSynchronized` state, the MM/GM relationship enters the `Idling` state when you specify the `-access` option, which enables write I/O on the auxiliary volume.
 - b. When an MM/GM relationship is stopped in the `ConsistentSynchronized` state without an `-access` parameter, the auxiliary volumes remain read-only and the state of the relationship changes to `ConsistentStopped`.
 - c. To enable write I/O on the auxiliary volume, when the MM/GM relationship is in the `ConsistentStopped` state, run the `svctask stopprcrelationship` command, which specifies the `-access` option, and the MM/GM relationship enters the `Idling` state.
- ▶ Step 5:
 - a. When an MM/GM relationship is started from the `Idling` state, you must specify the `-primary` argument to set the copy direction. If no write I/O was performed (to the master or auxiliary volume) while in the `Idling` state, the MM/GM relationship enters the `ConsistentSynchronized` state.
 - b. If write I/O was performed to the master or auxiliary volume, the `-force` option must be specified and the MM/GM relationship then enters the `InconsistentCopying` state while the background copy is started. The background process copies only the data that changed on the primary volume while the relationship was stopped.

Stop on Error

When a MM/GM relationship is stopped (intentionally, or because of an error), the state changes. For example, the MM/GM relationships in the `ConsistentSynchronized` state enter the `ConsistentStopped` state, and the MM/GM relationships in the `InconsistentCopying` state enter the `InconsistentStopped` state.

If the connection is broken between the two systems that are in a partnership, all (intercluster) MM/GM relationships enter a `Disconnected` state. For more information, see “Connected versus disconnected” on page 659.

Common states: Stand-alone relationships and consistency groups share a common configuration and state model. All MM/GM relationships in a consistency group have the same state as the consistency group.

State overview

The following sections provide an overview of the various MM/GM states.

Connected versus disconnected

Under certain error scenarios (for example, a power failure at one site that causes one complete system to disappear), communications between two systems in an MM/GM relationship can be lost. Alternatively, the fabric connection between the two systems might fail, which leaves the two systems that are running but cannot communicate with each other.

When the two systems can communicate, the systems and the relationships that relationships that span them are described as *connected*. When they cannot communicate, the systems and the relationships spanning them are described as *disconnected*.

In this state, both systems are left with fragmented relationships and are limited regarding the configuration commands that can be performed. The disconnected relationships are portrayed as having a changed state. The new states describe what is known about the relationship and the configuration commands that are permitted.

When the systems can communicate again, the relationships are reconnected. MM/GM automatically reconciles the two state fragments and considers any configuration or other event that occurred while the relationship was disconnected. As a result, the relationship can return to the state that it was in when it became disconnected, or it can enter a new state.

Relationships that are configured between volumes in the same IBM Spectrum Virtualize based system (intracluster) are never described as being in a disconnected state.

Consistent versus inconsistent

Relationships that contain volumes that are operating as secondaries can be described as being consistent or inconsistent. Consistency groups that contain relationships can also be described as being consistent or inconsistent. The consistent or inconsistent property describes the relationship of the data on the auxiliary to the data on the master volume. It can be considered a property of the auxiliary volume.

An auxiliary volume is described as *consistent* if it contains data that can be read by a host system from the master if power failed at an imaginary point while I/O was in progress, and power was later restored. This imaginary point is defined as the *recovery point*.

The requirements for consistency are expressed regarding activity at the master up to the recovery point. The auxiliary volume contains the data from all of the writes to the master for which the host received successful completion and that data was not overwritten by a subsequent write (before the recovery point).

Consider writes for which the host did not receive a successful completion (that is, it received bad completion or no completion at all). If the host then performed a read from the master of that data that returned successful completion and no later write was sent (before the recovery point), the auxiliary contains the same data as the data that was returned by the read from the master.

From the point of view of an application, consistency means that an auxiliary volume contains the same data as the master volume at the recovery point (the time at which the imaginary power failure occurred). If an application is designed to cope with an unexpected power failure, this assurance of consistency means that the application can use the auxiliary and begin operation as though it was restarted after the hypothetical power failure. Again, maintaining the application write ordering is the key property of consistency.

For more information about dependent writes, see 10.1.13, “FlashCopy and image mode volumes” on page 577.

If a relationship (or set of relationships) is inconsistent and an attempt is made to start an application by using the data in the secondaries, the following outcomes are possible:

- ▶ The application might decide that the data is corrupted and crash or exit with an event code.
- ▶ The application might fail to detect that the data is corrupted and return erroneous data.
- ▶ The application might work without a problem.

Because of the risk of data corruption, and in particular undetected data corruption, MM/GM strongly enforces the concept of consistency and prohibits access to inconsistent data.

Consistency as a concept can be applied to a single relationship or a set of relationships in a consistency group. Write ordering is a concept that an application can maintain across several disks that are accessed through multiple systems. Therefore, consistency must operate across all of those disks.

When you are deciding how to use consistency groups, the administrator must consider the scope of an application's data and consider all of the interdependent systems that communicate and exchange information.

If two programs or systems communicate and store details as a result of the information that is exchanged, either of the following actions might occur:

- ▶ All of the data that is accessed by the group of systems must be placed into a single consistency group.
- ▶ The systems must be recovered independently (each system within its own consistency group). Then, each system must perform recovery with the other applications to become consistent with them.

Consistent versus synchronized

A copy that is consistent and up-to-date is described as *synchronized*. In a synchronized relationship, the master and auxiliary volumes differ only in regions where writes are outstanding from the host.

Consistency does not mean that the data is up to date. A copy can be consistent and yet contain data that was frozen at a point in the past. Write I/O might continue to a master but not be copied to the auxiliary. This state arises when it becomes impossible to keep data up-to-date and maintain consistency. An example is a loss of communication between systems when you are writing to the auxiliary.

When communication is lost for an extended period and Consistency Protection was not enabled, MM/GM tracks the changes that occurred on the master, but not the order or the details of such changes (write data). When communication is restored, it is impossible to synchronize the auxiliary without sending write data to the auxiliary out of order. Therefore, consistency is lost.

Note: MM/GM relationships with Consistency Protection enabled use a PiT copy mechanism (FlashCopy) to keep a consistent copy of the auxiliary. The relationships stay in a consistent state, although not synchronized, even if communication is lost. For more information about Consistency Protection, see 10.6.20, "Consistency Protection for Global Mirror and Metro Mirror" on page 656.

Detailed states

The following sections describe the states that are portrayed to the user for consistency groups or relationships. Also described is the information that is available in each state. The major states are designed to provide guidance about the available configuration commands.

InconsistentStopped

InconsistentStopped is a connected state. In this state, the master is accessible for read and write I/O, but the auxiliary is not accessible for read or write I/O. A copy process must be started to make the auxiliary consistent. This state is entered when the relationship or consistency group was *InconsistentCopying* and suffered a persistent error or received a **stop** command that caused the copy process to stop.

A **start** command causes the relationship or consistency group to move to the *InconsistentCopying* state. A **stop** command is accepted, but has no effect.

If the relationship or consistency group becomes disconnected, the auxiliary side makes the transition to `InconsistentDisconnected`. The master side changes to `IdlingDisconnected`.

InconsistentCopying

`InconsistentCopying` is a connected state. In this state, the master is accessible for read and write I/O, but the auxiliary is not accessible for read or write I/O. This state is entered after a **start** command is issued to an `InconsistentStopped` relationship or a consistency group.

It is also entered when a forced start is issued to an `Idling` or `ConsistentStopped` relationship or consistency group. In this state, a background copy process runs that copies data from the master to the auxiliary volume.

In the absence of errors, an `InconsistentCopying` relationship is active, and the copy progress increases until the copy process completes. In certain error situations, the copy progress might freeze or even regress.

A persistent error or **stop** command places the relationship or consistency group into an `InconsistentStopped` state. A **start** command is accepted but has no effect.

If the background copy process completes on a stand-alone relationship or on all relationships for a consistency group, the relationship or consistency group changes to the `ConsistentSynchronized` state.

If the relationship or consistency group becomes disconnected, the auxiliary side changes to `InconsistentDisconnected`. The master side changes to `IdlingDisconnected`.

ConsistentStopped

`ConsistentStopped` is a connected state. In this state, the auxiliary contains a consistent image, but it might be out-of-date in relation to the master. This state can arise when a relationship was in a `ConsistentSynchronized` state and experienced an error that forces a Consistency Freeze. It can also arise when a relationship is created with a `CreateConsistentFlag` set to `TRUE`.

Normally, write activity that follows an I/O error causes updates to the master, and the auxiliary is no longer synchronized. In this case, consistency must be given up for a period to reestablish synchronization. You must run a **start** command with the **-force** option to acknowledge this condition, and the relationship or consistency group changes to `InconsistentCopying`. Enter this command only after all outstanding events are repaired.

In the unusual case where the master and the auxiliary are still synchronized (perhaps following a user stop, and no further write I/O was received), a **start** command takes the relationship to `ConsistentSynchronized`. No **-force** option is required. Also, in this case, you can run a **switch** command that moves the relationship or consistency group to `ConsistentSynchronized` and reverses the roles of the master and the auxiliary.

If the relationship or consistency group becomes disconnected, the auxiliary changes to `ConsistentDisconnected`. The master changes to `IdlingDisconnected`.

An informational status log is generated whenever a relationship or consistency group enters the `ConsistentStopped` state with a status of `Online`. You can configure this event to generate an SNMP trap that can be used to trigger automation or manual intervention to run a **start** command after a loss of synchronization.

ConsistentSynchronized

ConsistentSynchronized is a connected state. In this state, the master volume is accessible for read and write I/O, and the auxiliary volume is accessible for read-only I/O. Writes that are sent to the master volume are also sent to the auxiliary volume. Successful completion must be received for both writes, the write must be failed to the host, or a state must change out of the ConsistentSynchronized state before a write is completed to the host.

A **stop** command takes the relationship to the ConsistentStopped state. A **stop** command with the **-access** parameter takes the relationship to the Idling state.

A **switch** command leaves the relationship in the ConsistentSynchronized state, but it reverses the master and auxiliary roles (it switches the direction of replicating data). A **start** command is accepted, but has no effect.

If the relationship or consistency group becomes disconnected, the same changes are made as for ConsistentStopped.

Idling

Idling is a connected state. Both master and auxiliary volumes operate in the master role. Therefore, both master and auxiliary volumes are accessible for write I/O.

In this state, the relationship or consistency group accepts a **start** command. MM/GM maintains a record of regions on each disk that received write I/O while they were idling. This record is used to determine what areas must be copied following a **start** command.

The **start** command must specify the new copy direction. A **start** command can cause a loss of consistency if either volume in any relationship received write I/O, which is indicated by the Synchronized status. If the **start** command leads to loss of consistency, you must specify the **-force** parameter.

Following a **start** command, the relationship or consistency group changes to ConsistentSynchronized if there is no loss of consistency, or to InconsistentCopying if a loss of consistency occurs.

Also, the relationship or consistency group accepts a **-clean** option on the **start** command while in this state. If the relationship or consistency group becomes disconnected, both sides change their state to IdlingDisconnected.

IdlingDisconnected

IdlingDisconnected is a disconnected state. The target volumes in this half of the relationship or consistency group are all in the master role and accept read or write I/O.

The priority in this state is to recover the link to restore the relationship or consistency.

No configuration activity is possible (except for deletes or stops) until the relationship becomes connected again. At that point, the relationship changes to a connected state. The exact connected state that is entered depends on the state of the other half of the relationship or consistency group, which depends on the following factors:

- ▶ The state when it became disconnected.
- ▶ The write activity since it was disconnected.
- ▶ The configuration activity since it was disconnected.

If both halves are IdlingDisconnected, the relationship becomes Idling when it is reconnected.

While `IdlingDisconnected`, if a write I/O is received that causes the loss of synchronization (synchronized attribute transitions from `true` to `false`) and the relationship was not already stopped (through a user stop or a persistent error), an event is raised to notify you of the condition. This same event also is raised when this condition occurs for the `ConsistentSynchronized` state.

InconsistentDisconnected

`InconsistentDisconnected` is a disconnected state. The target volumes in this half of the relationship or consistency group are all in the auxiliary role, and do not accept read *or* write I/O. Except for deletes, no configuration activity is permitted until the relationship becomes connected again.

When the relationship or consistency group becomes connected again, the relationship becomes `InconsistentCopying` automatically unless either of the following conditions are true:

- ▶ The relationship was `InconsistentStopped` when it became disconnected.
- ▶ The user issued a **stop** command while disconnected.

In either case, the relationship or consistency group becomes `InconsistentStopped`.

ConsistentDisconnected

`ConsistentDisconnected` is a disconnected state. The target volumes in this half of the relationship or consistency group are all in the auxiliary role, and accept read I/O but *not* write I/O.

This state is entered from `ConsistentSynchronized` or `ConsistentStopped` when the auxiliary side of a relationship becomes disconnected.

In this state, the relationship or consistency group displays an attribute of `FreezeTime`, which is the point when consistency was frozen. When it is entered from `ConsistentStopped`, it retains the time that it had in that state. When it is entered from `ConsistentSynchronized`, the `FreezeTime` shows the last time at which the relationship or consistency group was known to be consistent. This time corresponds to the time of the last successful heartbeat to the other system.

A **stop** command with the `-access` flag set to `true` transitions the relationship or consistency group to the `IdlingDisconnected` state. This state allows write I/O to be performed to the auxiliary volume and is used as part of a DR scenario.

When the relationship or consistency group becomes connected again, the relationship or consistency group becomes `ConsistentSynchronized` only if this action does not lead to a loss of consistency. The following conditions must be true:

- ▶ The relationship was `ConsistentSynchronized` when it became disconnected.
- ▶ No writes received successful completion at the master while disconnected.

Otherwise, the relationship becomes `ConsistentStopped`. The `FreezeTime` setting is retained.

Empty

This state applies only to consistency groups. It is the state of a consistency group that has no relationships and no other state information to show.

It is entered when a consistency group is first created. It is exited when the first relationship is added to the consistency group, at which point the state of the relationship becomes the state of the consistency group.

10.7 Remote Copy commands

This section presents commands that must be issued to create and operate RC services.

10.7.1 Remote Copy process

The MM/GM process includes the following steps:

1. A system partnership is created between two IBM Spectrum Virtualize systems (for intercluster MM/GM).
2. A MM/GM relationship is created between two volumes of the same size.
3. To manage multiple MM/GM relationships as one entity, the relationships can be made part of a MM/GM consistency group to ensure data consistency across multiple MM/GM relationships, or for ease of management.
4. The MM/GM relationship is started. When the background copy completes, the relationship is consistent and synchronized. When synchronized, the auxiliary volume holds a copy of the production data at the master that can be used for DR.
5. To access the auxiliary volume, the MM/GM relationship must be stopped with the access option enabled before write I/O is submitted to the auxiliary.

Following these steps, the remote host server is mapped to the auxiliary volume and the disk is available for I/O.

The command set for MM/GM contains the following broad groups:

- ▶ Commands to create, delete, and manipulate relationships and consistency group
- ▶ Commands to cause state changes

If a configuration command affects more than one system, MM/GM coordinates configuration activity between the systems. Specific configuration commands can be run only when the systems are connected, and fail with no effect when they are disconnected.

Other configuration commands are permitted, even if the systems are disconnected. The state is reconciled automatically by MM/GM when the systems become connected again.

For any command (with one exception), a single system receives the command from the administrator. This design is significant for defining the context for a CreateRelationship **mkrcrelationship** or CreateConsistencyGroup **mkrcconsistgrp** command. In this case, the system that is receiving the command is called the *local system*.

The exception is a command that sets systems into a MM/GM partnership. The **mkfcpartnership** and **mkippartnership** commands must be issued on both the local and remote systems.

The commands in this section are described as an abstract command set, and are implemented by using one of the following methods:

- ▶ CLI can be used for scripting and automation.
- ▶ GUI can be used for one-off tasks.

10.7.2 Listing available system partners

Run the `lspartnershipcandidate` command to list the systems that are available for setting up a two-system partnership. This command is a prerequisite for creating MM/GM relationships.

Note: This command is not supported on IP partnerships. Use `mkippartnership` for IP connections.

10.7.3 Changing the system parameters

When you want to change system parameters specific to any RC or GM only, use the `chsystem` command. The `chsystem` command features the following parameters for MM/GM:

▶ `-relationshipbandwidthlimit cluster_relationship_bandwidth_limit`

This parameter controls the maximum rate at which any one RC relationship can synchronize. The default value for the relationship bandwidth limit is 25 MBps, but this value can now be specified as 1 - 100,000 MBps.

The partnership overall limit is controlled at a partnership level by the `chpartnership -linkbandwidthmbits` command, and must be set on each involved system.

Important: Do not set this value higher than the default without first establishing that the higher bandwidth can be sustained without affecting the host's performance. The limit must never be higher than the maximum that is supported by the infrastructure connecting the remote sites, regardless of the compression rates that you might achieve.

▶ `-gmlinktolerance link_tolerance`

This parameter specifies the maximum period that the system tolerates delay before stopping GM relationships. Specify values of 60 - 86,400 seconds in increments of 10 seconds. The default value is 300. Do not change this value except under the direction of IBM Support.

▶ `-gmmaxhostdelay max_host_delay`

This parameter specifies the maximum time delay, in milliseconds, at which the GM link tolerance timer starts counting down. This threshold value determines the extra effect that GM operations can add to the response times of the GM source volumes. You can use this parameter to increase the threshold from the default value of 5 milliseconds.

▶ `-maxreplicationdelay max_replication_delay`

This parameter sets a maximum replication delay in seconds. The value must be a number 0 - 360 (0 being the default value, no delay). This feature sets the maximum number of seconds to be tolerated to complete a single I/O. If I/O cannot complete within the `max_replication_delay`, the 1920 event is reported. This setting is system-wide and applies to MM/GM relationships.

Run the `chsystem` command to adjust these values, as shown in the following example:

```
chsystem -gmlinktolerance 300
```

You can view all of these parameter values by running the `lssystem <system_name>` command.

Focus on the **gm1inktolerance** parameter in particular. If poor response extends past the specified tolerance, a 1920 event is logged and one or more GM relationships automatically stop to protect the application hosts at the primary site. During normal operations, application hosts experience a minimal effect from the response times because the GM feature uses asynchronous replication.

However, if GM operations experience degraded response times from the secondary system for an extended period, I/O operations queue at the primary system. This queue results in an extended response time to application hosts. In this situation, the **gm1inktolerance** feature stops GM relationships, and the application host's response time returns to normal.

After a 1920 event occurs, the GM auxiliary volumes are no longer in the `consistent_synchronized` state. Fix the cause of the event and restart your GM relationships. For this reason, ensure that you monitor the system to track when these 1920 events occur.

You can disable the **gm1inktolerance** feature by setting the **gm1inktolerance** value to 0 (zero). However, the **gm1inktolerance** feature cannot protect applications from extended response times if it is disabled. It might be appropriate to disable the **gm1inktolerance** feature under the following circumstances:

- ▶ During SAN maintenance windows in which degraded performance is expected from SAN components, and application hosts can stand extended response times from GM volumes.
- ▶ During periods when application hosts can tolerate extended response times and it is expected that the **gm1inktolerance** feature might stop the GM relationships. For example, if you test by using an I/O generator that is configured to stress the back-end storage, the **gm1inktolerance** feature might detect the high latency and stop the GM relationships.

Disabling the **gm1inktolerance** feature prevents this result at the risk of exposing the test host to extended response times.

A 1920 event indicates that one or more of the SAN components cannot provide the performance that is required by the application hosts. This situation can be temporary (for example, a result of a maintenance activity) or permanent (for example, a result of a hardware failure or an unexpected host I/O workload).

If 1920 events are occurring, you might need to use a performance monitoring and analysis tool, such as the IBM Spectrum Control, to help identify and resolve the problem.

10.7.4 System partnership

To create a partnership, run one of the following commands, depending on the connection type:

- ▶ The **mkfcpartnership** command to establish a one-way MM/GM partnership between the local system and a remote system that are linked over an FC (or FCoE) connection.
- ▶ The **mkippartnership** command to establish a one-way MM/GM partnership between the local system and a remote system that are linked over an IP connection.

To establish a fully functional MM/GM partnership, you must run either of these commands on both systems that will be part of the partnership. This step is a prerequisite for creating MM/GM relationships between volumes on IBM Spectrum Virtualize systems.

When creating the partnership, you must specify the Link Bandwidth and can specify the Background Copy Rate:

- ▶ The Link Bandwidth, which is expressed in Mbps, is the amount of bandwidth that can be used for the FC or IP connection between the systems within the partnership.
- ▶ The Background Copy Rate is the maximum percentage of the Link Bandwidth that can be used for background copy operations. The default value is 50%.

Background copy bandwidth effect on foreground I/O latency

The combination of the Link Bandwidth value and the Background Copy Rate percentage is referred to as the *Background Copy bandwidth*. It must be at least 8 Mbps. For example, if the Link Bandwidth is set to 10000 and the Background Copy Rate is set to 20, the resulting Background Copy bandwidth that is used for background operations is 200 Mbps.

The background copy bandwidth determines the rate at which the background copy is attempted for MM/GM. The background copy bandwidth can affect foreground I/O latency in one of the following ways:

- ▶ The following results can occur if the background copy bandwidth is set too high compared to the MM/GM intercluster link capacity:
 - The background copy I/Os can back up on the MM/GM intercluster link.
 - There is a delay in the synchronous auxiliary writes of foreground I/Os.
 - The foreground I/O latency increases as perceived by applications.
- ▶ If the background copy bandwidth is set too high for the storage at the primary site, background copy read I/Os overload the primary storage and delay foreground I/Os.
- ▶ If the background copy bandwidth is set too high for the storage at the secondary site, background copy writes at the secondary site overload the auxiliary storage, and again delay the synchronous secondary writes of foreground I/Os.

To set the background copy bandwidth optimally, ensure that you consider all three resources: Primary storage, intercluster link bandwidth, and auxiliary storage. Provision the most restrictive of these three resources between the background copy bandwidth and the peak foreground I/O workload.

Perform this provisioning by calculation or by determining experimentally how much background copy can be allowed before the foreground I/O latency becomes unacceptable. Then, reduce the background copy to accommodate peaks in workload.

The `chpartnership` command

To change the bandwidth that is available for background copy in the system partnership, run the `chpartnership -backgroundcopyrate <percentage_of_link_bandwidth>` command to specify the percentage of whole link capacity to be used by the background copy process.

10.7.5 Creating a Metro Mirror/Global Mirror consistency group

Run the `mkrconsistgrp` command to create an empty MM/GM consistency group.

The MM/GM consistency group name must be unique across all consistency groups that are known to the systems owning this consistency group. If the consistency group involves two systems, the systems must be in communication throughout the creation process.

The new consistency group does not contain any relationships and is in the Empty state. You can add MM/GM relationships to the group (upon creation or afterward) by running the `chrelationship` command.

10.7.6 Creating a Metro Mirror/Global Mirror relationship

Run the `mkrcrelationship` command to create a MM/GM relationship. This relationship persists until it is deleted.

Optional parameter: If you do not use the `-global` optional parameter, an MM relationship is created rather than a GM relationship.

The auxiliary volume must be equal in size to the master volume or the command fails. If both volumes are in the same system, they must be in the same I/O group. The master and auxiliary volume cannot be in a relationship, and they cannot be the target of a FlashCopy mapping. This command returns the new relationship (`relationship_id`) when successful.

When the MM/GM relationship is created, you can add it to a consistency group, or it can be a stand-alone MM/GM relationship.

The `lsrcrelationshipcandidate` command

Run the `lsrcrelationshipcandidate` command to list the volumes that are eligible to form an MM/GM relationship.

When the command is issued, you can specify the master volume name and auxiliary system to list the candidates that comply with the prerequisites to create a MM/GM relationship. If the command is issued with no parameters, all of the volumes that are not disallowed by another configuration state, such as being a FlashCopy target, are listed.

10.7.7 Changing a Metro Mirror/Global Mirror relationship

Run the `chrcrelationship` command to modify the following properties of an MM/GM relationship:

- ▶ Change the name of an MM/GM relationship.
- ▶ Add a relationship to a group.
- ▶ Remove a relationship from a group by using the `-force` flag.

Adding an MM/GM relationship: When an MM/GM relationship is added to a consistency group that is not empty, the relationship must have the same state and copy direction as the group to be added to it.

10.7.8 Changing a Metro Mirror/Global Mirror consistency group

Run the `chrconsistgrp` command to change the name of an MM/GM consistency group.

10.7.9 Starting a Metro Mirror/Global Mirror relationship

Run the `startrcrelationship` command to start the copy process of an MM/GM relationship.

When the command is run, you can set the copy direction if it is undefined. Optionally, you can mark the auxiliary volume of the relationship as clean. The command fails if it is used as an attempt to start a relationship that is a part of a consistency group.

You can run this command only to a relationship that is connected. For a relationship that is idling, this command assigns a copy direction (master and auxiliary roles) and begins the copy process. Otherwise, this command restarts a previous copy process that was stopped by a **stop** command or by an I/O error.

If the resumption of the copy process leads to a period when the relationship is inconsistent, you must specify the **-force** parameter when the relationship is restarted. This situation can arise if, for example, the relationship was stopped and then further writes were performed on the original master of the relationship.

The use of the **-force** parameter here is a reminder that the data on the auxiliary becomes inconsistent while resynchronization (background copying) occurs. Therefore, this data is unusable for DR purposes before the background copy completes.

In the `Idling` state, you must specify the master volume to indicate the copy direction. In other connected states, you can provide the **-primary** argument, but it must match the existing setting.

10.7.10 Stopping a Metro Mirror/Global Mirror relationship

Run the **stopcrelationship** command to stop the copy process for a relationship. You can also use this command to enable write access to a consistent auxiliary volume by specifying the **-access** parameter.

This command applies to a stand-alone relationship. It is rejected if it is addressed to a relationship that is part of a consistency group. You can issue this command to stop a relationship that is copying from master to auxiliary.

If the relationship is in an inconsistent state, any copy operation stops and does not resume until you run a **startcrelationship** command. Write activity is no longer copied from the master to the auxiliary volume. For a relationship in the `ConsistentSynchronized` state, this command causes a Consistency Freeze.

When a relationship is in a consistent state (that is, in the `ConsistentStopped`, `ConsistentSynchronized`, or `ConsistentDisconnected` state), you can use the **-access** parameter with the **stopcrelationship** command to enable write access to the auxiliary volume.

10.7.11 Starting a Metro Mirror/Global Mirror consistency group

Run the **startcrconsistgrp** command to start an MM/GM consistency group. You can issue this command only to a consistency group that is connected.

For a consistency group that is idling, this command assigns a copy direction (master and auxiliary roles) and begins the copy process. Otherwise, this command restarts a previous copy process that was stopped by a **stop** command or by an I/O error.

10.7.12 Stopping a Metro Mirror/Global Mirror consistency group

Run the **stopcrconsistgrp** command to stop the copy process for an MM/GM consistency group. You can also use this command to enable write access to the auxiliary volumes in the group if the group is in a consistent state.

If the consistency group is in an inconsistent state, any copy operation stops and does not resume until you run the **startrcconsistgrp** command. Write activity is no longer copied from the master to the auxiliary volumes that belong to the relationships in the group. For a consistency group in the ConsistentSynchronized state, this command causes a Consistency Freeze.

When a consistency group is in a consistent state (for example, in the ConsistentStopped, ConsistentSynchronized, or ConsistentDisconnected state), you can use the **-access** parameter with the **stoprcconsistgrp** command to enable write access to the auxiliary volumes within that group.

10.7.13 Deleting a Metro Mirror/Global Mirror relationship

Run the **rmmrcrelationship** command to delete the relationship that is specified. Deleting a relationship deletes only the logical relationship between the two volumes. It does not affect the volumes.

If the relationship is disconnected at the time that the command is issued, the relationship is deleted on only the system on which the command is being run. When the systems reconnect, the relationship is automatically deleted on the other system.

Alternatively, if the systems are disconnected and you still want to remove the relationship on both systems, you can run the **rmmrcrelationship** command independently on both of the systems.

A relationship cannot be deleted if it is part of a consistency group. You must first remove the relationship from the consistency group.

If you delete an inconsistent relationship, the auxiliary volume becomes accessible, even though it is still inconsistent. This situation is the one case in which MM/GM does not inhibit access to inconsistent data.

10.7.14 Deleting a Metro Mirror/Global Mirror consistency group

Run the **rmmrcconsistgrp** command to delete an MM/GM consistency group. This command deletes the specified consistency group.

If the consistency group is disconnected at the time that the command is issued, the consistency group is deleted on only the system on which the command is being run. When the systems reconnect, the consistency group is automatically deleted on the other system.

Alternatively, if the systems are disconnected and you still want to remove the consistency group on both systems, you can run the **rmmrcconsistgrp** command separately on both of the systems.

If the consistency group is not empty, the relationships within it are removed from the consistency group before the group is deleted. These relationships then become stand-alone relationships. The state of these relationships is not changed by the action of removing them from the consistency group.

10.7.15 Reversing a Metro Mirror/Global Mirror relationship

Run the **switchrcrelationship** command to reverse the roles of the master volume and the auxiliary volume when a stand-alone relationship is in a consistent state. When the command is issued, the wanted master must be specified.

10.7.16 Reversing a Metro Mirror/Global Mirror consistency group

Run the `switchrconsistgrp` command to reverse the roles of the master volume and the auxiliary volume when a consistency group is in a consistent state. This change is applied to all of the relationships in the consistency group. When the command is issued, the wanted master must be specified.

Important: By reversing the roles, your current source volumes become targets, and target volumes become source. Therefore, you lose write access to your current primary volumes.

10.8 Native IP replication

IBM Spectrum Virtualize can implement RC services by using FC connections or IP connections. This section describes the IBM Spectrum Virtualize IP replication technology and implementation.

Demonstration: The IBM Client Demonstration Center shows how data is replicated by using GMCV (cycling mode set to `multiple`). This configuration perfectly fits the new IP replication function because it is well-dwell designed for links with high latency, low bandwidth, or both.

For more information, see this [web page](#) (log in required).

10.8.1 Native IP replication technology

Remote Mirroring over IP communication is supported on the SVC and IBM FlashSystem family systems by using Ethernet communication links. IBM Spectrum Virtualize software IP replication uses innovative Bridgeworks SANSlide technology to optimize network bandwidth and utilization. This function enables the use of a lower-speed and lower-cost networking infrastructure for data replication.

Bridgeworks SANSlide technology, which is integrated into the IBM Spectrum Virtualize Software, uses artificial intelligence (AI) to help optimize network bandwidth use and adapt to changing workload and network conditions.

This technology can improve remote mirroring network bandwidth usage up to three times. Improved bandwidth usage can enable clients to deploy a less costly network infrastructure, or speed up remote replication cycles to enhance DR effectiveness.

With an Ethernet network data flow, the data transfer can slow down over time. This condition occurs because of the latency that is caused by waiting for the acknowledgment of each set of packets that is sent. The next packet set cannot be sent until the previous packet is acknowledged, as shown in Figure 10-95 on page 673.

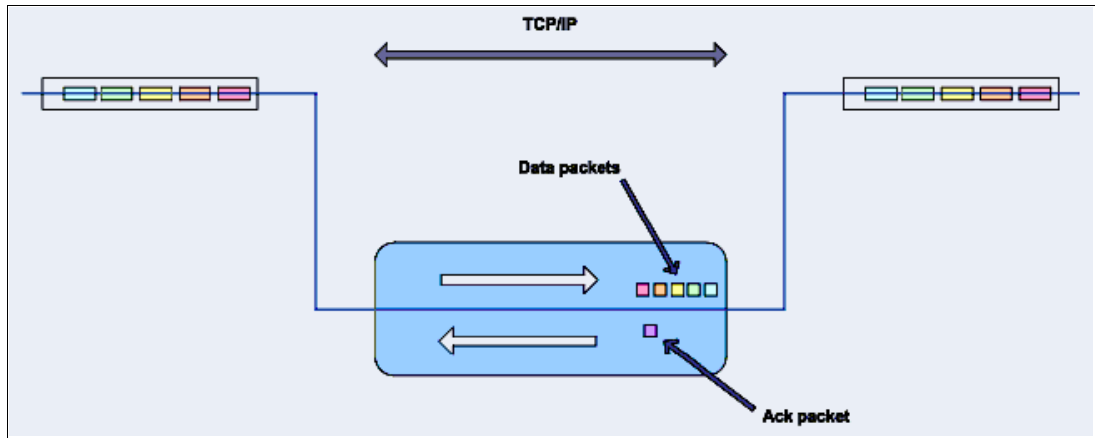


Figure 10-95 Typical Ethernet network data flow

However, by using the embedded IP replication, this behavior can be eliminated with the enhanced parallelism of the data flow by using multiple virtual connections (VCs) that share IP links and addresses. The AI engine can dynamically adjust the number of VCs, receive window size, and packet size to maintain optimum performance. While the engine is waiting for one VCs ACK, it sends more packets across other VCs. If packets are lost from any VC, data is automatically retransmitted, as shown in Figure 10-96.

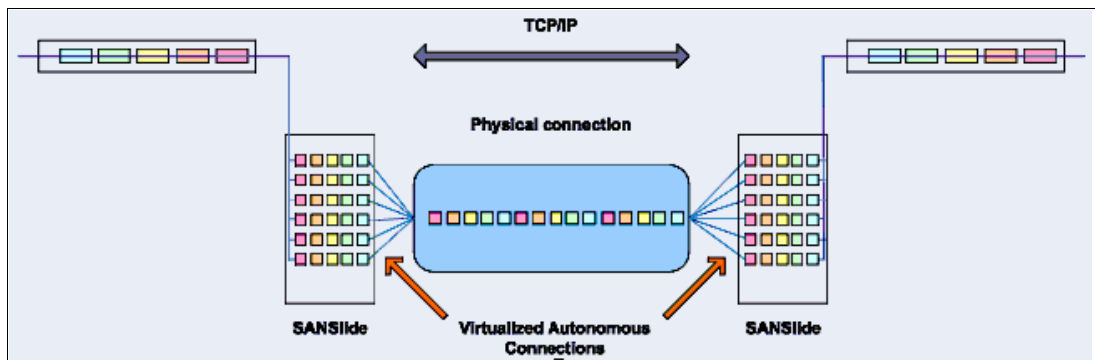


Figure 10-96 Optimized network data flow by using Bridgeworks SANSlide technology

For more information about this technology, see *IBM Storwize V7000 and SANSlide Implementation*, REDP-5023.

With native IP partnership, the following Copy Services features are supported:

- ▶ MM

Referred to as *synchronous replication*, MM provides a consistent copy of a source volume on a target volume. Data is written to the target volume synchronously after it is written to the source volume so that the copy is continuously updated.

- ▶ GM and GMCV

Referred to as *asynchronous replication*, GM provides a consistent copy of a source volume on a target volume. Data is written to the target volume asynchronously so that the copy is continuously updated. However, the copy might not contain the last few updates if a DR operation is performed. An added extension to GM is GMCV. GMCV is the preferred method for use with native IP replication.

Note: For IP partnerships, generally use the GMCV method of copying (asynchronous copy of changed grains only). This method can include performance benefits. Also, GM and MM might be more susceptible to the loss of synchronization.

10.8.2 IP partnership limitations

The following prerequisites and assumptions must be considered before IP partnership between two IBM Spectrum Virtualize systems can be established:

- ▶ The IBM Spectrum Virtualize systems are successfully installed with V7.2 or later code levels.
- ▶ The systems must have the necessary licenses that enable RC partnerships to be configured between two systems. No separate license is required to enable IP partnership.
- ▶ The storage SANs are configured correctly and the correct infrastructure to support the IBM Spectrum Virtualize systems in RC partnerships over IP links is in place.
- ▶ The two systems must be able to ping each other and perform the discovery.
- ▶ TCP ports 3260 and 3265 are used by systems for IP partnership communications. Therefore, these ports must be open.
- ▶ The maximum number of partnerships between the local and remote systems, including both IP and FC partnerships, is limited to the current maximum that is supported, which is three partnerships (four systems total).
- ▶ Only a single partnership over IP is supported.
- ▶ A system can have simultaneous partnerships over FC and IP, but with separate systems. The FC zones between two systems must be removed before an IP partnership is configured.
- ▶ IP partnerships are supported on both 10 gigabits per second (Gbps) links and 1 Gbps links. However, the intermix of both on a single link is not supported.
- ▶ The maximum supported round-trip time (RTT) is 80 ms for 1 Gbps links.
- ▶ The maximum supported RTT is 10 ms for 10 Gbps links.
- ▶ The inter-cluster heartbeat traffic uses 1 Mbps per link.
- ▶ Only nodes from two I/O groups can have ports that are configured for an IP partnership.
- ▶ Migrations of RC relationships directly from FC-based partnerships to IP partnerships are not supported.
- ▶ IP partnerships between the two systems can be over IPv4 or IPv6 only, but not both.
- ▶ Virtual local area network (VLAN) tagging of the IP addresses that are configured for RC is supported starting with V7.4.
- ▶ Management IP and internet Small Computer Systems Interface (iSCSI) IP on the same port can be in a different network starting with Version 7.4.
- ▶ An added layer of security is provided by using Challenge Handshake Authentication Protocol (CHAP) authentication.
- ▶ TCP ports 3260 and 3265 are used for IP partnership communications. Therefore, these ports must be open in firewalls between the systems.
- ▶ Only a single RC data session per physical link can be established. It is intended that only one connection (for sending/receiving RC data) is made for each independent physical link between the systems.

Note: A physical link is the physical IP link between the two sites: A (local) and B (remote). Multiple IP addresses on local system A might be connected (by Ethernet switches) to this physical link. Similarly, multiple IP addresses on remote system B might be connected (by Ethernet switches) to the same physical link. At any time, only a single IP address on cluster A can form an RC data session with an IP address on cluster B.

- ▶ The maximum throughput is restricted based on the use of 1 Gbps, 10 Gbps, or 25 Gbps Ethernet ports. It varies based on distance (for example, round-trip latency) and quality of communication link (for example, packet loss):
 - One 1 Gbps port can transfer up to 110 MBps unidirectional, 190 MBps bidirectional.
 - Two 1 Gbps ports can transfer up to 220 MBps unidirectional, 325 MBps bidirectional.
 - One 10 Gbps port can transfer up to 240 MBps unidirectional, 350 MBps bidirectional.
 - Two 10 Gbps port can transfer up to 440 MBps unidirectional, 600 MBps bidirectional.

Note: IP Replication is supported by 25 Gbps Mellanox and Chelsio adapters, but be aware there is no performance benefit or advantage for IP Replication with these adapters. However, for the purpose for consolidation where these cards are used for other purposes, such as iSCSI Extensions for Remote Direct Memory Access (RDMA) (iSER) Host Attach or iSCSI Host Attach/Backend Virtualization, they can be used for IP replication.

The minimum supported link bandwidth is 10 Mbps. However, this requirement scales up with the amount of host I/O that you choose to do. Figure 10-97 shows scaling host I/O.

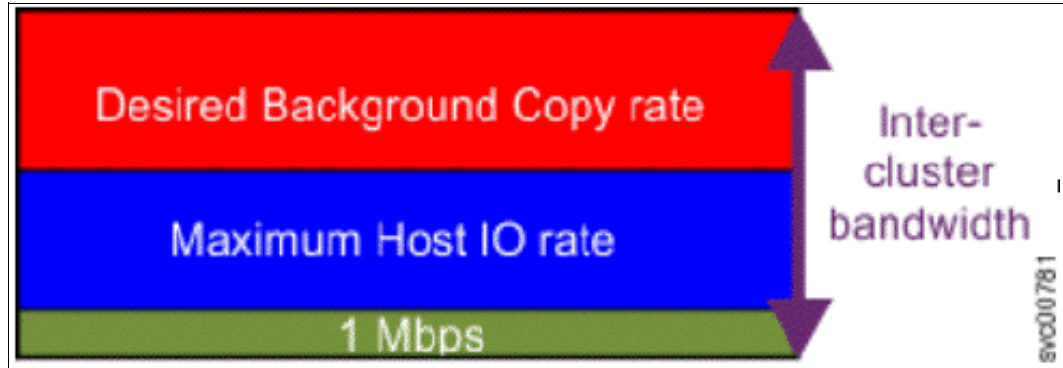


Figure 10-97 Scaling of host I/O

The following equation describes the approximate minimum bandwidth that is required between two systems with < 5 ms RTT and errorless link:

$$\text{Minimum intersite link bandwidth in Mbps} > \text{Required Background Copy in Mbps} + \text{Maximum Host I/O in Mbps} + 1 \text{ Mbps heartbeat traffic}$$

Increasing latency and errors results in a higher requirement for minimum bandwidth.

Note: The Bandwidth setting definition when the IP partnerships are created changed in V7.7. Previously, the bandwidth setting defaulted to 50 MiB, and was the maximum transfer rate from the primary site to the secondary site for initial sync/resync of volumes.

The Link Bandwidth setting is now configured by using megabits (Mb) not MB. You set the Link Bandwidth setting to a value that the communication link can sustain, or to what is allocated for replication. The Background Copy Rate setting is now a percentage of the Link Bandwidth. The Background Copy Rate setting determines the available bandwidth for the initial sync and resyncs or for GMCV.

10.8.3 IP Partnership and data compression

When creating an IP partnership between two systems, you can specify whether you want to use the data compression feature. When enabled, IP partnership compression compresses the data that is sent from a local system to the remote system and potentially uses less bandwidth than with uncompressed data. It is also used to decompress data that is received by a local system from a remote system.

Data compression is supported for IPv4 or IPv6 partnerships. To enable data compression, both systems in an IP partnership must be running a software level that supports IP partnership compression (Version 7.7 or later) and both must have the compression feature enabled.

A compression license is not needed on any local or remote system.

Volumes that are replicated by using IP partnership compression can be either compressed or uncompressed on the system. Volume compression and IP replication compression are not linked features. As an example, the following steps replicate a compressed volume over an IP partnership with the compression feature enabled:

1. Read operations in the local system decompress the data when reading from the source volume.
2. Decompressed data is transferred to the RC code.
3. Data is compressed before being sent over the IP partnership link.
4. The remote system RC code decompresses the received data.
5. Write operations in the remote system compress the data when writing to the target volume.

10.8.4 VLAN support

Starting with V7.4, VLAN tagging is supported for iSCSI host attachment and IP replication. Hosts and RC operations can connect to the system through Ethernet ports. Each traffic type has different bandwidth requirements, which can interfere with each other if they share IP connections. VLAN tagging creates two separate connections on the same IP network for different types of traffic. The system supports VLAN configuration on both IPv4 and IPv6 connections.

When the VLAN ID is configured for IP addresses that is used for iSCSI host attach or IP replication, the VLAN settings on the Ethernet network and servers must be configured correctly to avoid connectivity issues. After the VLANs are configured, changes to the VLAN settings disrupt iSCSI and IP replication traffic to and from the partnerships.

During the VLAN configuration for each IP address, the VLAN settings for the local and failover ports on two nodes of an I/O group can differ. To avoid any service disruption, switches must be configured so that the failover VLANs are configured on the local switch ports and the failover of IP addresses from a failing node to a surviving node succeeds. If failover VLANs are not configured on the local switch ports, no paths are available to the IBM Spectrum Virtualize system nodes during a node failure and the replication fails.

Consider the following requirements and procedures when implementing VLAN tagging:

- ▶ VLAN tagging is supported for IP partnership traffic between two systems.
- ▶ VLAN provides network traffic separation at the layer 2 level for Ethernet transport.
- ▶ VLAN tagging by default is disabled for any IP address of a node port (N_Port). You can use the CLI or GUI to optionally set the VLAN ID for port IP addresses on both systems in the IP partnership.
- ▶ When a VLAN ID is configured for the port IP addresses that are used in RC port groups, appropriate VLAN settings on the Ethernet network must also be configured to prevent connectivity issues.

Setting VLAN tags for a port is disruptive. Therefore, VLAN tagging requires that you stop the partnership first before you configure VLAN tags. Restart the partnership after the configuration is complete.

10.8.5 IP partnership and terminology

The IP partnership terminology and abbreviations that are used are listed in Table 10-13.

Table 10-13 Terminology for IP partnership

IP partnership terminology	Description
RC group or RC port group	<p>The following numbers group a set of IP addresses that are connected to the same physical link. Therefore, only IP addresses that are part of the same RC group can form RC connections with the partner system:</p> <ul style="list-style-type: none"> ▶ 0: Ports that are not configured for RC ▶ 1: Ports that belong to RC port group 1 ▶ 2: Ports that belong to RC port group 2 <p>Each IP address can be shared for iSCSI host attach and RC functions. Therefore, appropriate settings must be applied to each IP address.</p>
IP partnership	Two systems that are partnered to perform RC over native IP links.
FC partnership	Two systems that are partnered to perform RC over native FC links.
Failover	Failure of a node within an I/O group causes the volume access to go through the surviving node. The IP addresses fail over to the surviving node in the I/O group. When the configuration node of the system fails, management IP addresses also fail over to an alternative node.
Failback	When the failed node rejoins the system, all failed over IP addresses are failed back from the surviving node to the rejoined node, and volume access is restored through this node.

IP partnership terminology	Description
linkbandwidthmbits	Aggregate bandwidth of all physical links between two sites in Mbps.
IP partnership or partnership over native IP links	These terms are used to describe the IP partnership feature.
Discovery	<p>Process by which two IBM Spectrum Virtualize systems exchange information about their IP address configuration. For IP-based partnerships, only IP addresses configured for RC are discovered.</p> <p>For example, the first Discovery takes place when the user is running the mkippartnership CLI command. Subsequent Discoveries can take place as a result of user activities (configuration changes) or as a result of hardware failures (for example, node failure, ports failure, and so on).</p>

10.8.6 States of IP partnership

The different partnership states in IP partnership are listed in Table 10-14.

Table 10-14 States of IP partnership

State	Systems connected	Support for active RC I/O	Comments
Partially_Configured_Local	No	No	This state indicates that the initial discovery is complete.
Fully_Configured	Yes	Yes	Discovery successfully completed between two systems, and the two systems can establish RC relationships.
Fully_Configured_Stopped	Yes	Yes	The partnership is stopped on the system.
Fully_Configured_Remote_Stopped	Yes	No	The partnership is stopped on the remote system.
Not_Present	Yes	No	The two systems cannot communicate with each other. This state is also seen when data paths between the two systems are not established.
Fully_Configured_Exceeded	Yes	No	There are too many systems in the network, and the partnership from the local system to remote system is disabled.
Fully_Configured_Excluded	No	No	The connection is excluded because of too many problems, or either system cannot support the I/O work load for the MM and GM relationships.

The process to establish two systems in the IP partnerships includes the following steps:

1. The administrator configures the CHAP secret on both the systems. This step is not mandatory, and users can choose to not configure the CHAP secret.
2. The administrator configures the system IP addresses on both local and remote systems so that they can discover each other over the network.

3. If you want to use VLANs, configure your local area network (LAN) switches and Ethernet ports to use VLAN tagging.
4. The administrator configures the systems ports on each node in both of the systems by using the GUI (or the `cfgport ip` CLI command), and completes the following steps:
 - a. Configure the IP addresses for RC data.
 - b. Add the IP addresses in the respective RC port group.
 - c. Define whether the host access on these ports over iSCSI is allowed.
5. The administrator establishes the partnership with the remote system from the local system where the partnership state then changes to `Partially_Configured_Local`.
6. The administrator establishes the partnership from the remote system with the local system. If this process is successful, the partnership state then changes to the `Fully_Configured`, which implies that the partnerships over the IP network were successfully established. The partnership state momentarily remains `Not_Present` before moving to the `Fully_Configured` state.
7. The administrator creates MM, GM, and GMCV relationships.

Partnership consideration: When the partnership is created, no master or auxiliary status is defined or implied. The partnership is equal. The concepts of *master or auxiliary* and *primary or secondary* apply to volume relationships only, not to system partnerships.

10.8.7 Remote Copy groups

This section describes Remote Copy (RC) groups (or RC port groups) and different ways to configure the links between the two remote systems. The two IBM Spectrum Virtualize systems can be connected to each other over one link or at most two links. To address the requirement to enable the systems to know about the physical links between the two sites, the concept of RC port groups was introduced.

RC port group ID is a numerical tag that is associated with an IP port of an IBM Spectrum Virtualize system to indicate to which physical IP link it is connected. Multiple nodes might be connected to the same physical long-distance link, and must therefore share RC port group ID.

In scenarios with two physical links between the local and remote clusters, two RC port group IDs must be used to designate which IP addresses are connected to which physical link. This configuration must be done by the system administrator by using the GUI or running the `cfgport ip` CLI command.

Remember: IP ports on both partners must be configured with identical RC port group IDs for the partnership to be established correctly.

The IBM Spectrum Virtualize system IP addresses that are connected to the same physical link are designated with identical RC port groups. The system supports three RC groups: 0, 1, and 2.

The systems' IP addresses are, by default, in RC port group 0. Ports in port group 0 are not considered for creating RC data paths between two systems. For partnerships to be established over IP links directly, IP ports must be configured in RC group 1 if a single inter-site link exists, or in RC groups 1 and 2 if two inter-site links exist.

You can assign one IPv4 address and one IPv6 address to each Ethernet port on the system platforms. Each of these IP addresses can be shared between iSCSI host attach and the IP partnership. The user must configure the required IP address (IPv4 or IPv6) on an Ethernet port with a an RC port group.

The administrator might want to use IPv6 addresses for RC operations and use IPv4 addresses on that same port for iSCSI host attach. This configuration also implies that for two systems to establish an IP partnership, both systems must have IPv6 addresses that are configured.

Administrators can choose to dedicate an Ethernet port for IP partnership only. In that case, host access must be disabled for that IP address and any other IP address that is configured on that Ethernet port.

Note: To establish an IP partnership, each IBM Spectrum Virtualize controller node must have only a single RC port group that is configured, either 1 or 2. The remaining IP addresses must be in RC port group 0.

10.8.8 Supported configurations

Note: For explanation purposes, this section shows a node with two ports available: 1 and 2. This number generally increments when the latest models of IBM Spectrum Virtualize systems are used.

The following supported configurations for IP partnership that were in the first release are described in this section:

- ▶ Two 2-node systems in IP partnership over a single inter-site link, as shown in Figure 10-98 (configuration 1).

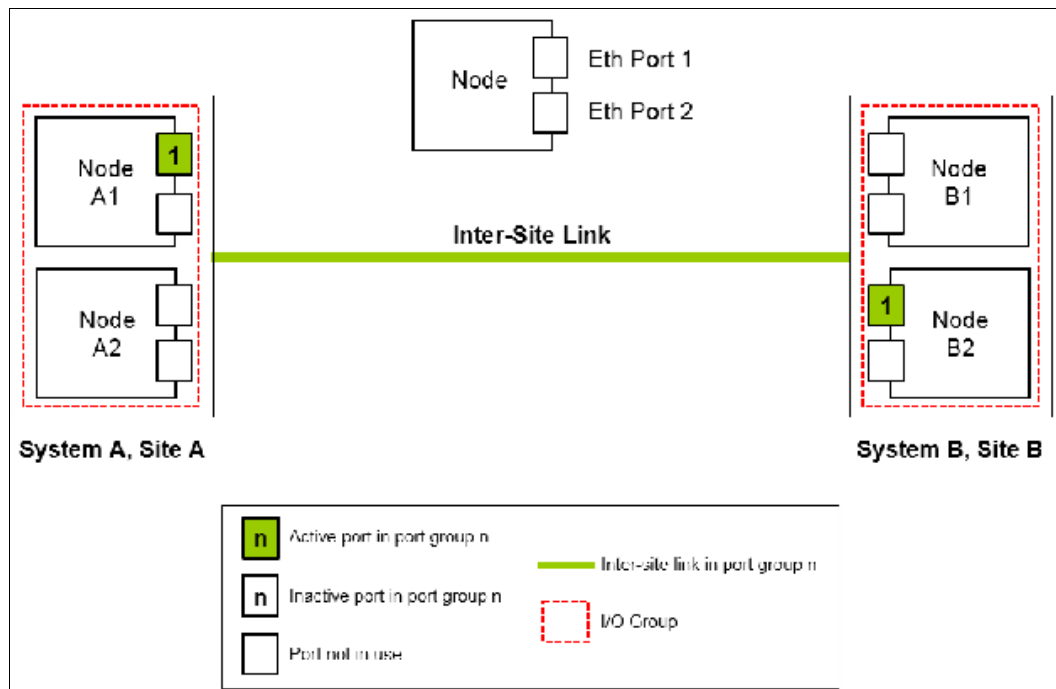


Figure 10-98 Single link with only one Remote Copy port group configured in each system

As shown in Figure 10-98 on page 680, two systems are available:

- System A.
- System B.

A single RC port group 1 is created on Node A1 on System A and on Node B2 on System B because only a single inter-site link is used to facilitate the IP partnership traffic. An administrator might choose to configure the RC port group on Node B1 on System B rather than Node B2.

At any time, only the IP addresses that are configured in RC port group 1 on the nodes in System A and System B participate in establishing data paths between the two systems after the IP partnerships are created. In this configuration, no failover ports are configured on the partner node in the same I/O group.

This configuration has the following characteristics:

- Only one node in each system has an RC port group that is configured, and no failover ports are configured.
 - If the Node A1 in System A or the Node B2 in System B encounter a failure, the IP partnership stops and enters the Not_Present state until the failed nodes recover.
 - After the nodes recover, the IP ports fail back, the IP partnership recovers, and the partnership state goes to the Fully_Configured state.
 - If the inter-site system link fails, the IP partnerships change to the Not_Present state.
 - This configuration is not recommended because it is not resilient to node failures.
- Two 2-node systems in IP partnership over a single inter-site link (with failover ports configured), as shown in Figure 10-99 on page 682 (configuration 2).

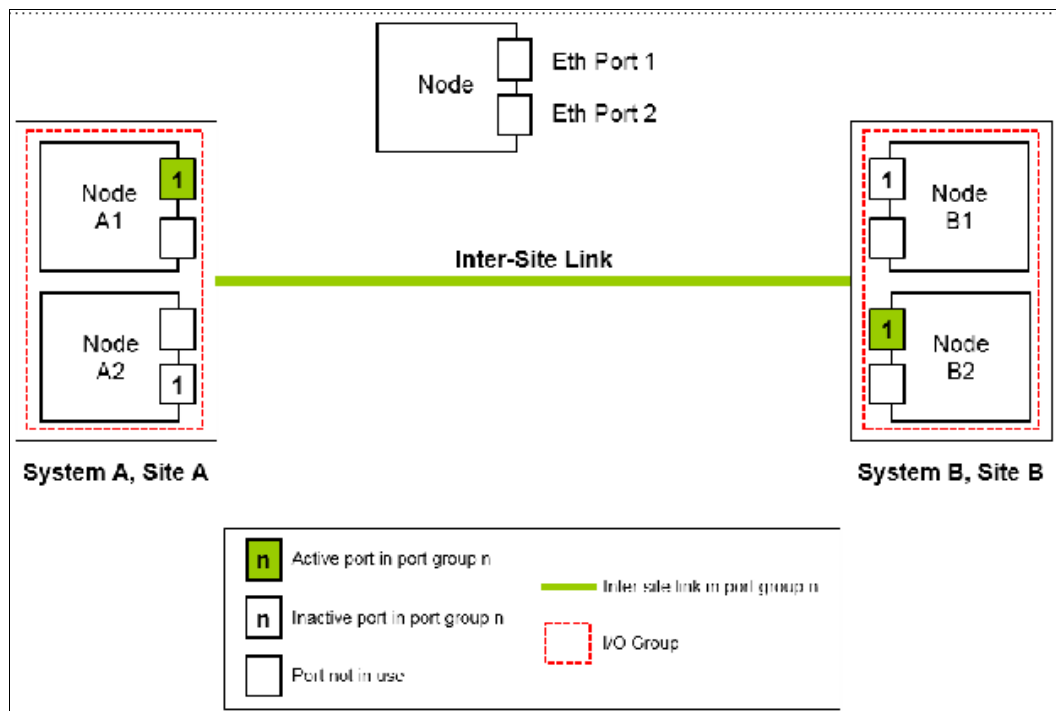


Figure 10-99 One Remote Copy group on each system and nodes with failover ports configured

As shown in Figure 10-99, two systems are available:

- System A.
- System B.

A single RC port group 1 is configured on two Ethernet ports, one each on Node A1 and Node A2 on System A. Similarly, a single RC port group is configured on two Ethernet ports on Node B1 and Node B2 on System B.

Although two ports on each system are configured for RC port group 1, only one Ethernet port in each system actively participates in the IP partnership process. This selection is determined by a path configuration algorithm that is designed to choose data paths between the two systems to optimize performance.

The other port on the partner node in the I/O group behaves as a standby port that is used if a node fails. If Node A1 fails in System A, IP partnership continues servicing replication I/O from Ethernet Port 2 because a failover port is configured on Node A2 on Ethernet Port 2.

However, it might take some time for discovery and path configuration logic to reestablish paths post failover. This delay can cause partnerships to change to Not_Present for that time. The details of the particular IP port that is actively participating in IP partnership is provided in the **l sport ip** output (reported as used).

This configuration has the following characteristics:

- Each node in the I/O group has the same RC port group that is configured. However, only one port in that RC port group is active at any time at each system.
 - If the Node A1 in System A or the Node B2 in System B fails in the respective systems, IP partnerships rediscovery is triggered and continues servicing the I/O from the failover port.
 - The discovery mechanism that is triggered because of failover might introduce a delay where the partnerships momentarily change to the Not_Present state and recover.
- Two 4-node systems in IP partnership over a single inter-site link (with failover ports configured), as shown in Figure 10-100 on page 683 (configuration 3).

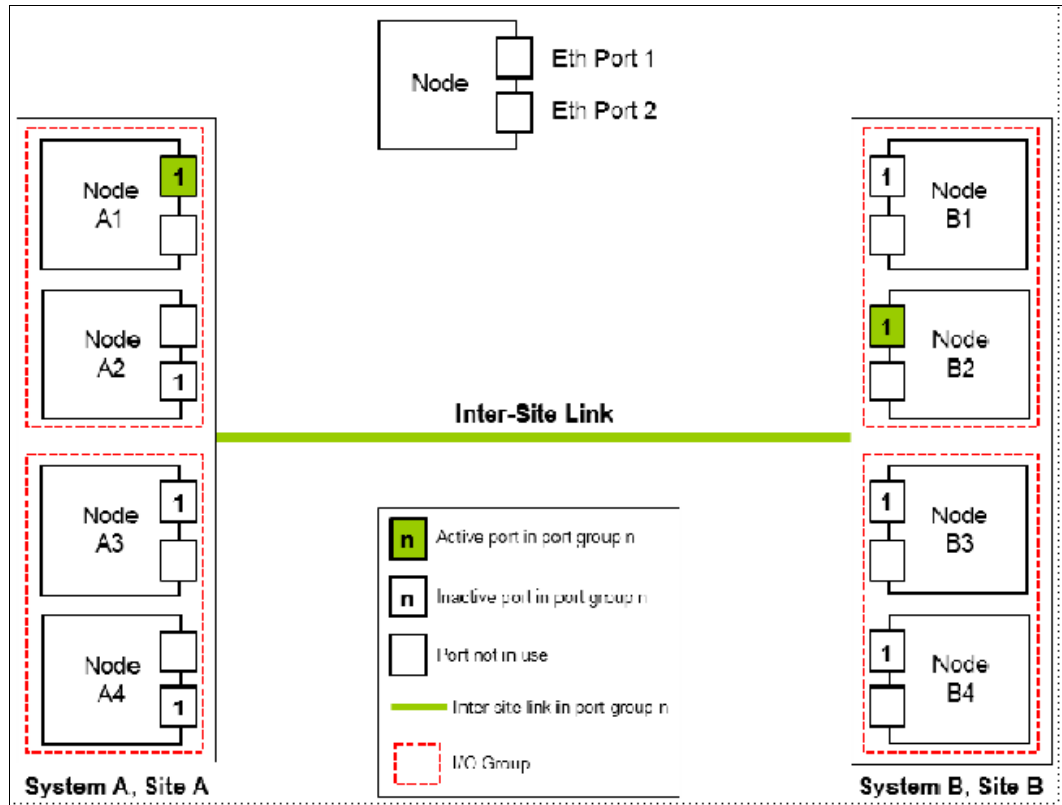


Figure 10-100 Multinode systems single inter-site link with only one RC port group

As shown in Figure 10-100, two 4-node systems are available:

- System A.
- System B.

A single RC port group 1 is configured on nodes A1, A2, A3, and A4 on System A, Site A; and on nodes B1, B2, B3, and B4 on System B, Site B. Although four ports are configured for RC group 1, only one Ethernet port in each RC port group on each system actively participates in the IP partnership process.

Port selection is determined by a path configuration algorithm. The other ports play the role of standby ports.

If Node A1 fails in System A, the IP partnership selects one of the remaining ports that is configured with RC port group 1 from any of the nodes from either of the two I/O groups in System A. However, it might take some time (generally seconds) for discovery and path configuration logic to reestablish paths post failover. This process can cause partnerships to change to the Not_Present state.

This result causes RC relationships to stop. The administrator might need to manually verify the issues in the event log and start the relationships or RC consistency groups, if they do not autorecover. The details of the particular IP port actively participating in the IP partnership process is provided in the `lsportip` view (reported as used).

This configuration has the following characteristics:

- Each node has the RC port group that is configured in both I/O groups. However, only one port in that RC port group remains active and participates in IP partnership on each system.
 - If the Node A1 in System A or the Node B2 in System B were to encounter some failure in the system, IP partnerships discovery is triggered and it continues servicing the I/O from the failover port.
 - The discovery mechanism that is triggered because of failover might introduce a delay wherein the partnerships momentarily change to the Not_Present state and then recover.
 - The bandwidth of the single link is used completely.
- Eight-node system in IP partnership with four-node system over single inter-site link, as shown in Figure 10-101 (configuration 4).

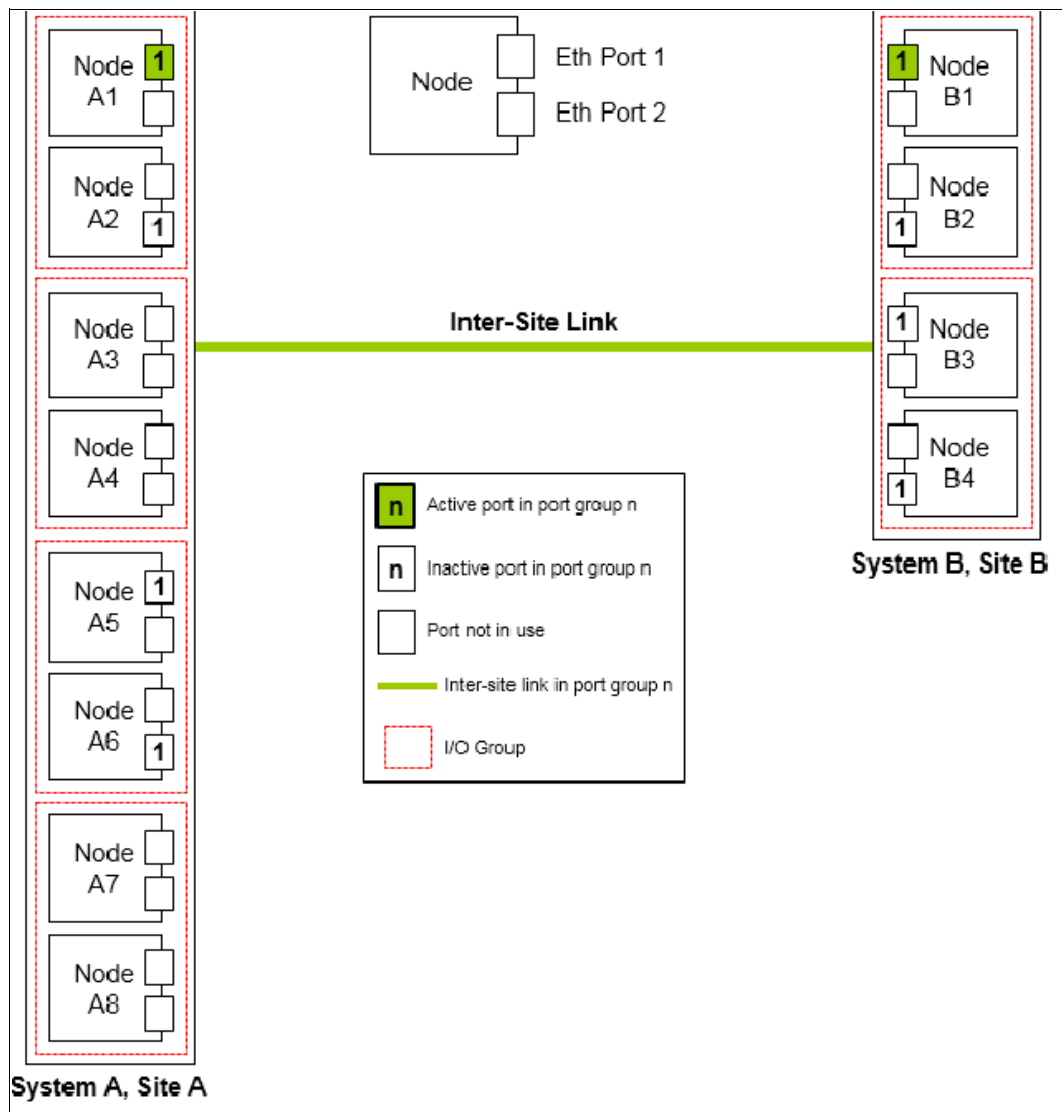


Figure 10-101 Multinode systems single inter-site link with only one Remote Copy port group

As shown in Figure 10-101 on page 684, an eight-node system (System A in Site A) and a four-node system (System B in Site B) are used. A single RC port group 1 is configured on nodes A1, A2, A5, and A6 on System A at Site A. Similarly, a single RC port group 1 is configured on nodes B1, B2, B3, and B4 on System B.

Although four I/O groups (eight nodes) are in System A, any two I/O groups at maximum are supported to be configured for IP partnerships. If Node A1 fails in System A, IP partnership continues by using one of the ports that is configured in RC port group from any of the nodes from either of the two I/O groups in System A.

However, it might take some time for discovery and path configuration logic to reestablish paths post-failover. This delay might cause partnerships to change to the `Not_Present` state.

This process can lead to RC relationships stopping, and the administrator must manually start them if the relationships do not auto-recover. The details of which particular IP port is actively participating in IP partnership process are provided in `lspport ip` output (reported as used).

This configuration features the following characteristics:

- ▶ Each node has the RC port group that is configured in both the I/O groups that are identified for participating in IP Replication. However, only one port in that RC port group remains active on each system and participates in IP Replication.
- ▶ If the Node A1 in System A or the Node B2 in System B fails in the system, the IP partnerships trigger discovery and continue servicing the I/O from the failover ports.
- ▶ The discovery mechanism that is triggered because of failover might introduce a delay wherein the partnerships momentarily change to the `Not_Present` state and then recover.
- ▶ The bandwidth of the single link is used completely.

- ▶ Two 2-node systems with two inter-site links, as shown in Figure 10-102 (configuration 5).

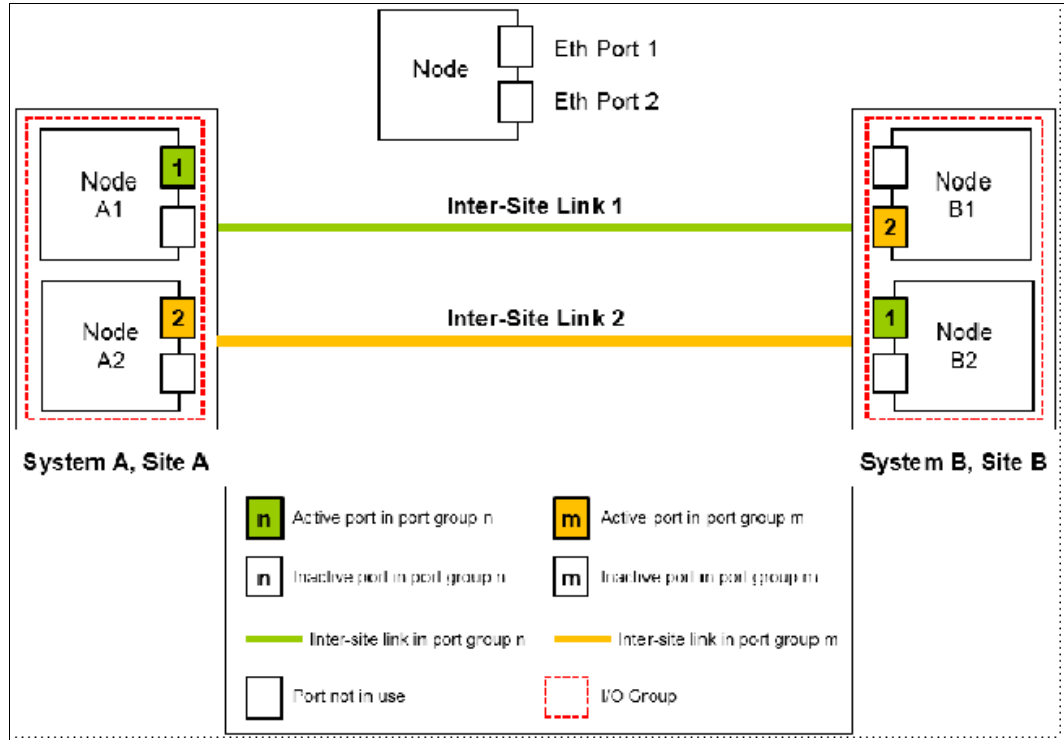


Figure 10-102 Dual links with two Remote Copy groups on each system configured

As shown in Figure 10-102, RC port groups 1 and 2 are configured on the nodes in System A and System B because two inter-site links are available. In this configuration, the failover ports are not configured on partner nodes in the I/O group. Instead, the ports are maintained in different RC port groups on both of the nodes. They remain active and participate in IP partnership by using both of the links.

However, if either of the nodes in the I/O group fail (that is, if Node A1 on System A fails), the IP partnership continues only from the available IP port that is configured in RC port group 2. Therefore, the effective bandwidth of the two links is reduced to 50% because only the bandwidth of a single link is available until the failure is resolved.

This configuration has the following characteristics:

- Two inter-site links and two RC port groups are configured.
- Each node has only one IP port in RC port group 1 or 2.
- Both the IP ports in the two RC port groups participate simultaneously in IP partnerships. Therefore, both of the links are used.
- During node failure or link failure, the IP partnership traffic continues from the other available link and the port group. Therefore, if two links of 10 Mbps each are available and you have 20 Mbps of effective link bandwidth, bandwidth is reduced to 10 Mbps only during a failure.
- After the node failure or link failure is resolved and failback occurs, the entire bandwidth of both of the links is available as before.

- ▶ Two 4-node systems in IP partnership with dual inter-site links, as shown in Figure 10-103 (configuration 6).

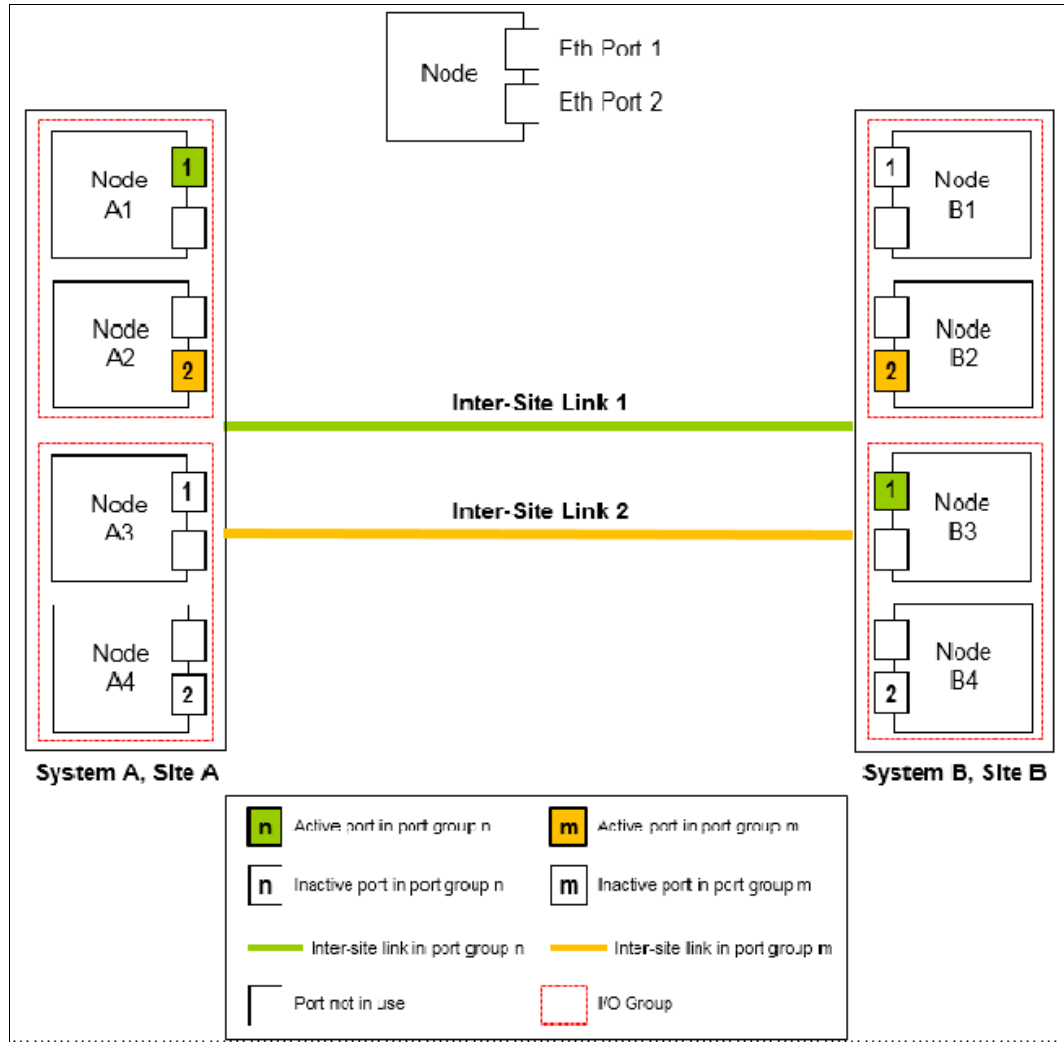


Figure 10-103 Multinode systems with dual inter-site links between the two systems

As shown in Figure 10-103, two 4-node systems are used:

- System A.
- System B.

This configuration is an extension of Configuration 5 to a multinode multi-I/O group environment. This configuration has two I/O groups, and each node in the I/O group has a single port that is configured in RC port groups 1 or 2.

Although two ports are configured in RC port groups 1 and 2 on each system, only one IP port in each RC port group on each system actively participates in IP partnership. The other ports that are configured in the same RC port group act as standby ports in the event of failure. Which port in a configured RC port group participates in IP partnership at any moment is determined by a path configuration algorithm.

In this configuration, if Node A1 fails in System A, IP partnership traffic continues from Node A2 (that is, RC port group 2) and at the same time the failover also causes discovery in RC port group 1.

Therefore, the IP partnership traffic continues from Node A3 on which RC port group 1 is configured. The details of the particular IP port that is actively participating in IP partnership process is provided in the `lsport ip` output (reported as used).

This configuration has the following characteristics:

- Each node has the RC port group that is configured in the I/O groups 1 or 2. However, only one port per system in both RC port groups remains active and participates in IP partnership.
 - Only a single port per system from each configured RC port group participates simultaneously in IP partnership. Therefore, both of the links are used.
 - During node failure or port failure of a node that is actively participating in IP partnership, IP partnership continues from the alternative port because another port is in the system in the same RC port group but in a different I/O group.
 - The pathing algorithm can start discovery of available ports in the affected RC port group in the second I/O group and pathing is reestablished, which restores the total bandwidth, so both of the links are available to support IP partnership.
- Eight-node system in IP partnership with a four-node system over dual inter-site links, as shown in Figure 10-104 on page 689 (configuration 7).

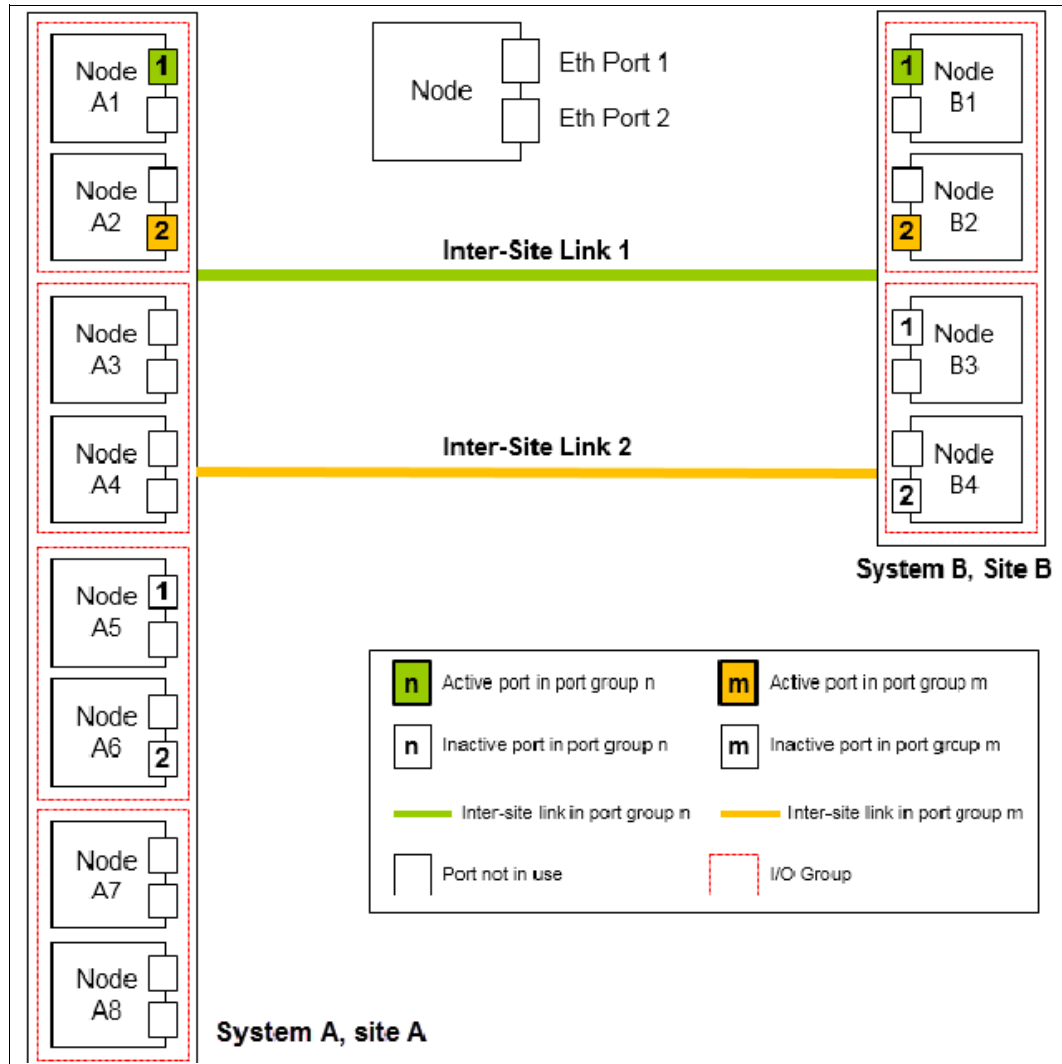


Figure 10-104 Multinode systems (two I/O groups on each system) with dual inter-site links between the two systems

As shown in Figure 10-104, an eight-node System A in Site A and a four-node System B in Site B is used. Because a maximum of two I/O groups in IP partnership is supported in a system, although four I/O groups (eight nodes) exist, nodes from only two I/O groups are configured with RC port groups in System A. The remaining or all of the I/O groups can be configured to be RC partnerships over FC.

In this configuration, two links and two I/O groups are configured with RC port groups 1 and 2, but path selection logic is managed by an internal algorithm. Therefore, this configuration depends on the pathing algorithm to decide which of the nodes actively participates in IP partnership. Even if Node A5 and Node A6 are configured with RC port groups properly, active IP partnership traffic on both of the links might be driven from Node A1 and Node A2 only.

If Node A1 fails in System A, IP partnership traffic continues from Node A2 (that is, RC port group 2). The failover also causes IP partnership traffic to continue from Node A5 on which RC port group 1 is configured. The details of the particular IP port actively participating in IP partnership process is provided in the `1 sport ip` output (reported as used).

This configuration has the following characteristics:

- Two I/O groups with nodes in those I/O groups are configured in two RC port groups because two inter-site links are used for participating in IP partnership. However, only one port per system in a particular RC port group remains active and participates in IP partnership.
 - One port per system from each RC port group participates in IP partnership simultaneously. Therefore, both of the links are used.
 - If a node or port on the node that is actively participating in IP partnership fails, the RC data path is established from that port because another port is available on an alternative node in the system with the same RC port group.
 - The path selection algorithm starts discovery of available ports in the affected RC port group in the alternative I/O groups and paths are reestablished, which restores the total bandwidth across both links.
 - The remaining or all of the I/O groups can be in RC partnerships with other systems.
- An example of an *unsupported* configuration for a single inter-site link is shown in Figure 10-105 (configuration 8).

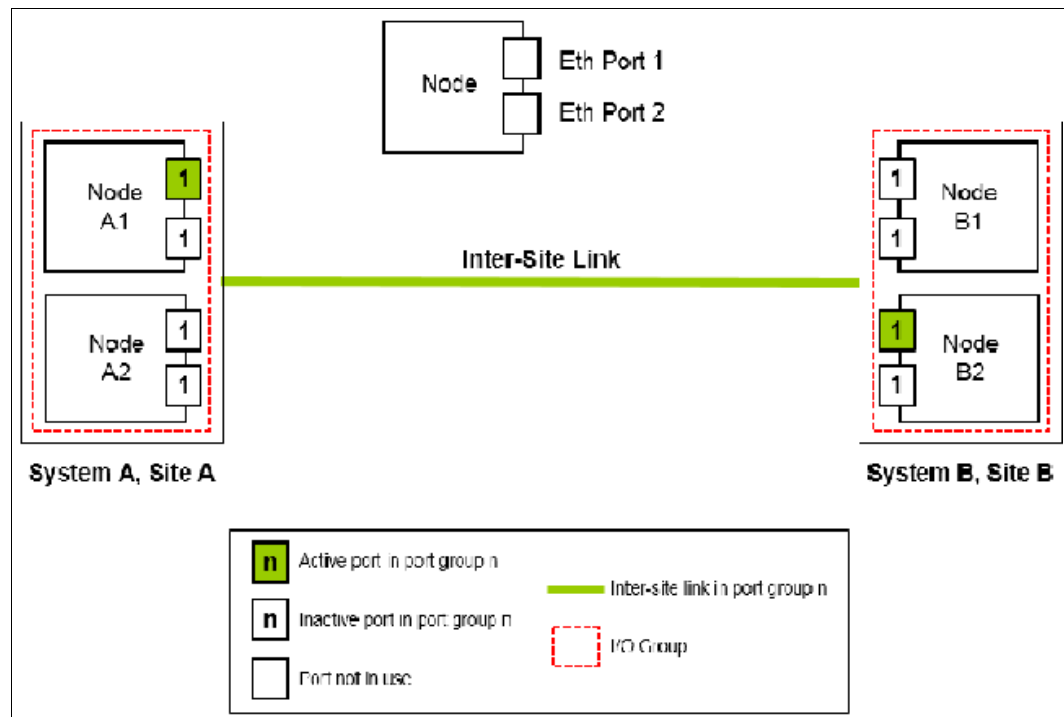


Figure 10-105 Two node systems with single inter-site link and Remote Copy port groups configured

As shown in Figure 10-105, this configuration is similar to Configuration 2, but differs because each node now has the same RC port group that is configured on more than one IP port.

On any node, only one port at any time can participate in IP partnership. Configuring multiple ports in the same RC group on the same node is *not supported*.

- An example of an *unsupported* configuration for a dual inter-site link is shown in Figure 10-106 (configuration 9).

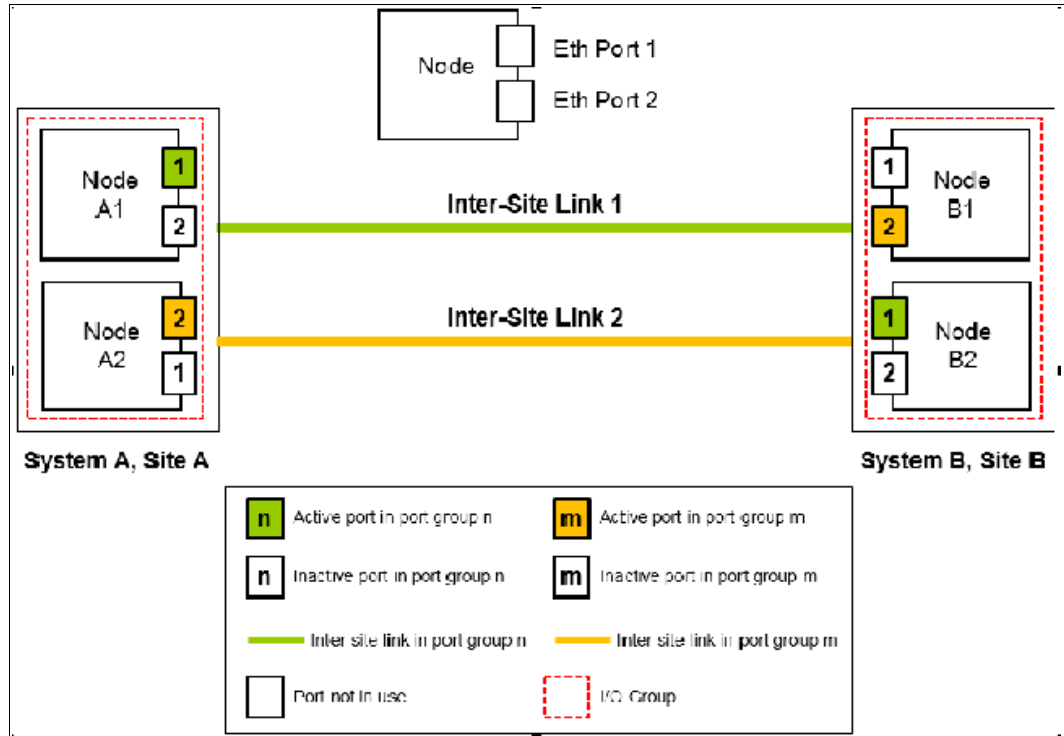


Figure 10-106 Dual links with two Remote Copy Port Groups with failover Port Groups configured

As shown in Figure 10-106, this configuration is similar to Configuration 5, but differs because each node now also has two ports that are configured with RC port groups. In this configuration, the path selection algorithm can select a path that might cause partnerships to change to the Not_Present state and then recover, which results in a configuration restriction. The use of this configuration is not recommended until the configuration restriction is lifted in future releases.

- An example deployment for configuration 2 with a dedicated inter-site link is shown in Figure 10-107 (configuration 10).

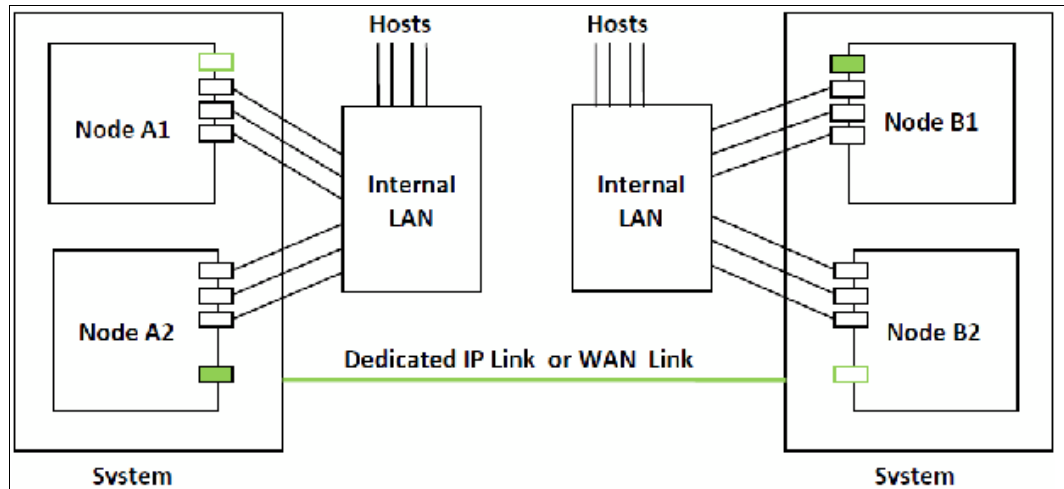


Figure 10-107 Deployment example

In this configuration, one port on each node in System A and System B is configured in RC group 1 to establish IP partnership and support RC relationships. A dedicated inter-site link is used for IP partnership traffic, and iSCSI host attach is disabled on those ports.

The following configuration steps are used:

- a. Configure system IP addresses properly. As such, they can be reached over the inter-site link.
 - b. Qualify if the partnerships must be created over IPv4 or IPv6, and then assign IP addresses and open firewall ports 3260 and 3265.
 - c. Configure IP ports for RC on both the systems by using the following settings:
 - RC group: 1
 - Host: No
 - Assign IP address
 - d. Check that the maximum transmission unit (MTU) levels across the network meet the requirements as set (default MTU is 1500).
 - e. Establish IP partnerships from both of the systems.
 - f. After the partnerships are in the Fully_Configured state, you can create the RC relationships.
- Figure 10-107 on page 691 is an example deployment for the configuration that is shown in Figure 10-101 on page 684. Ports that are shared with host access are shown in Figure 10-108 (configuration 11).

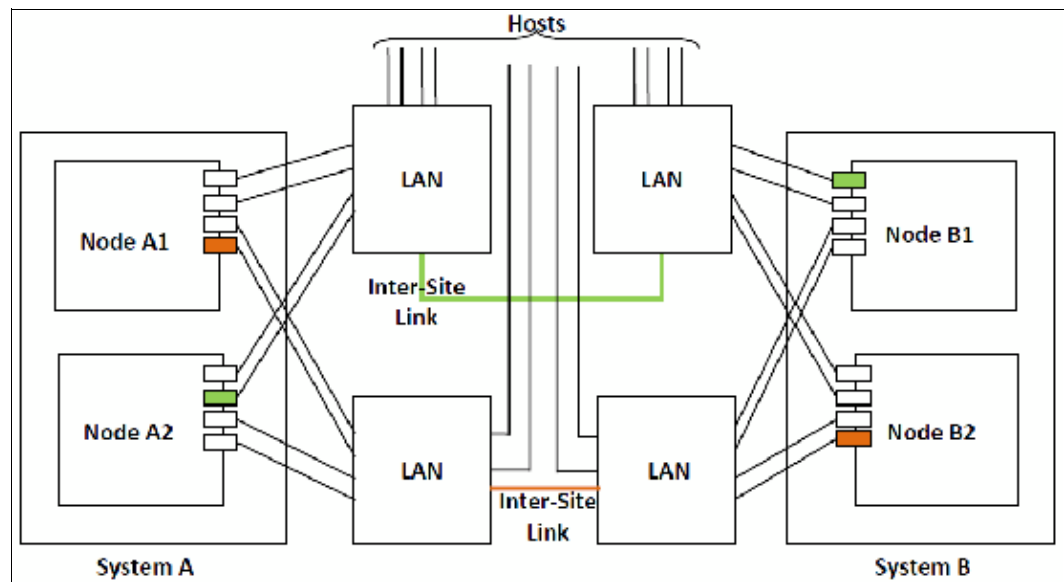


Figure 10-108 Deployment example

In this configuration, IP ports are to be shared by both iSCSI hosts and for IP partnership.

The following configuration steps are used:

- a. Configure System IP addresses properly so that they can be reached over the inter-site link.
- b. Qualify if the partnerships must be created over IPv4 or IPv6, and then assign IP addresses and open firewall ports 3260 and 3265.

- c. Configure IP ports for RC on System A1 by using the following settings:
 - Node 1:
 - Port 1, RC port group 1
 - Host: Yes
 - Assign IP address
 - Node 2:
 - Port 4, RC port group 2
 - Host: Yes
 - Assign IP address
- d. Configure IP ports for RC on System B1 by using the following settings:
 - Node 1:
 - Port 1, RC port group 1
 - Host: Yes
 - Assign IP address
 - Node 2:
 - Port 4, RC port group 2
 - Host: Yes
 - Assign IP address
- e. Check the MTU levels across the network (the default MTU is 1500 on SVC and IBM Spectrum Virtualize systems).
- f. Establish IP partnerships from both systems.
- g. After the partnerships are in the Fully_Configured state, you can create the RC relationships.

10.9 Managing Remote Copy by using the GUI

It is often easier to control MM/GM with the GUI if you have few mappings. When many mappings are used, run your commands by using the CLI. This section describes the tasks that you can perform at an RC level.

Note: The **Copy Services** → **Consistency Groups** menu relates to FlashCopy consistency groups only, not RC groups.

The following windows are used to visualize and manage your remote copies:

► Remote Copy window

To open the Remote Copy window, click **Copy Services** → **Remote Copy** in the main menu, as shown in Figure 10-109.

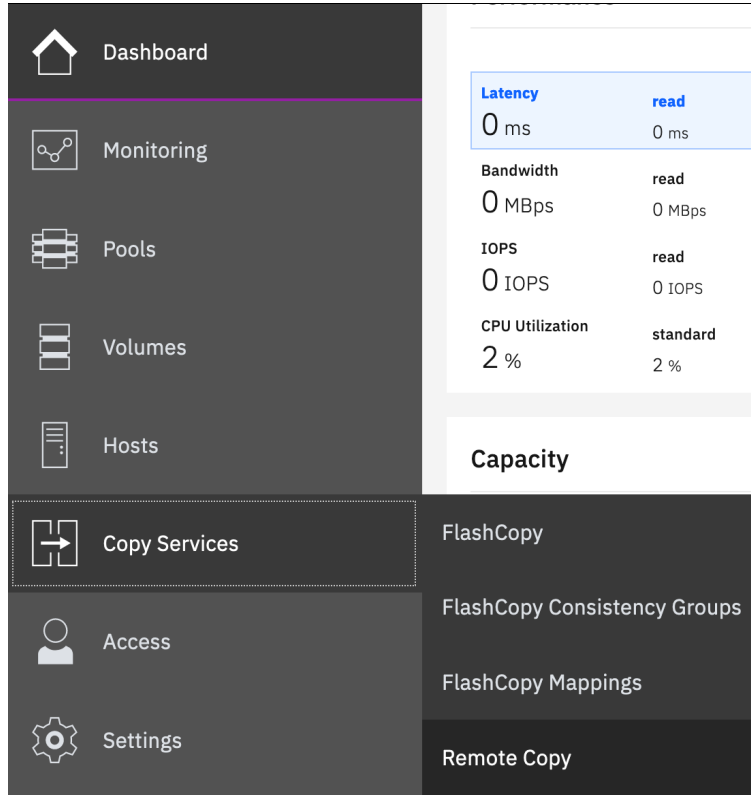


Figure 10-109 Remote Copy menu

The Remote Copy window opens, as shown in Figure 10-110.

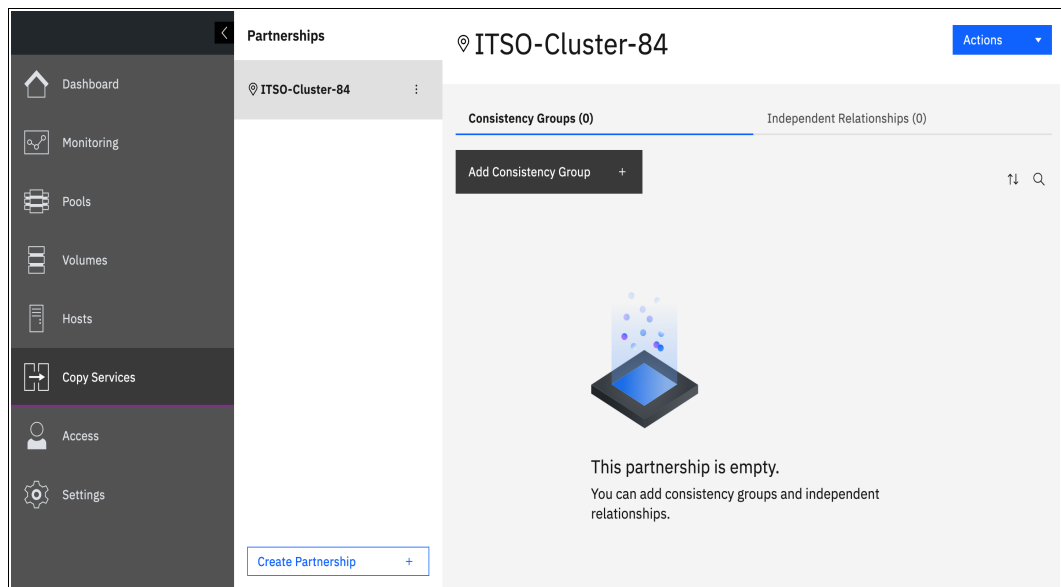


Figure 10-110 Remote Copy window

10.9.1 Creating a Fibre Channel partnership

Intra-cluster MM: If you are creating intra-cluster MM, do not perform this next step to create the MM partnership. Instead, see 10.9.2, “Creating Remote Copy relationships” on page 697.

To create an FC partnership between IBM Spectrum Virtualize systems by using the GUI, open the Remote Copy window that is shown in the Figure 10-110 on page 694 and click **Create Partnership** to create a partnership.

In the Create Partnership window, enter the following information:

1. Select the Replication topology between two or three sites, as shown in the Figure 10-111. In this example, we use a two-site partnership.

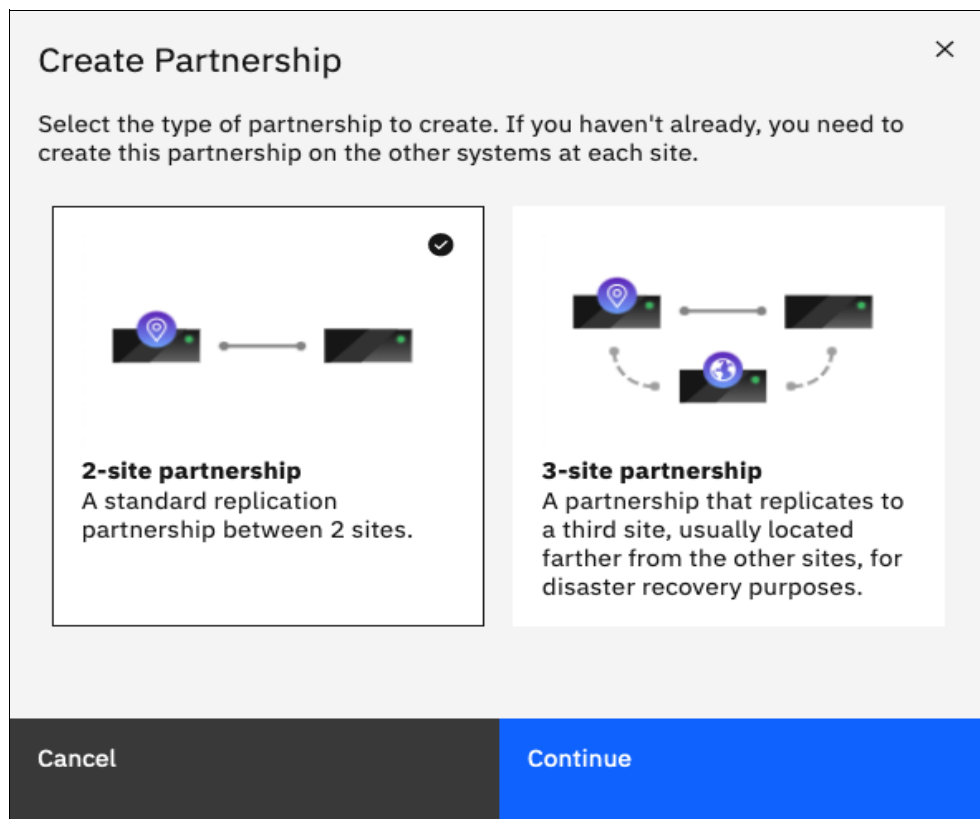


Figure 10-111 Selecting a Remote Copy topology

2. Select the partnership type (**Fibre Channel** or **IP**). If you choose an IP partnership, you must provide the IP address of the partner system and the partner system’s CHAP key.
3. If your partnership is based on Fibre Channel Protocol (FCP), select an available partner system from the menu. To be able to select a partner system, the two clusters must be properly zoned between each other. If no other candidate cluster is available, the This system does not have any candidates error message is displayed.
4. Enter a link bandwidth in Mbps that is used by the background copy process between the systems in the partnership.

5. Enter the background copy rate.
6. Click **OK** to confirm the partnership relationship (see Figure 10-112).

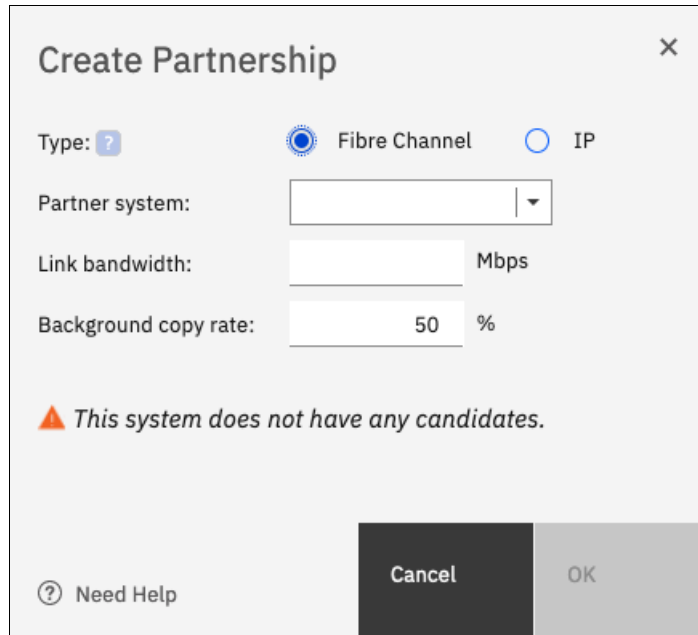


Figure 10-112 Creating a partnership details

To fully configure the partnership between both systems, perform the same steps on the other system in the partnership. If not configured on the partner system, the partnership is displayed as **Partial Local**.

When both sides of the system partnership are defined, the partnership shows a **Configured** green status, as shown in Figure 10-113.

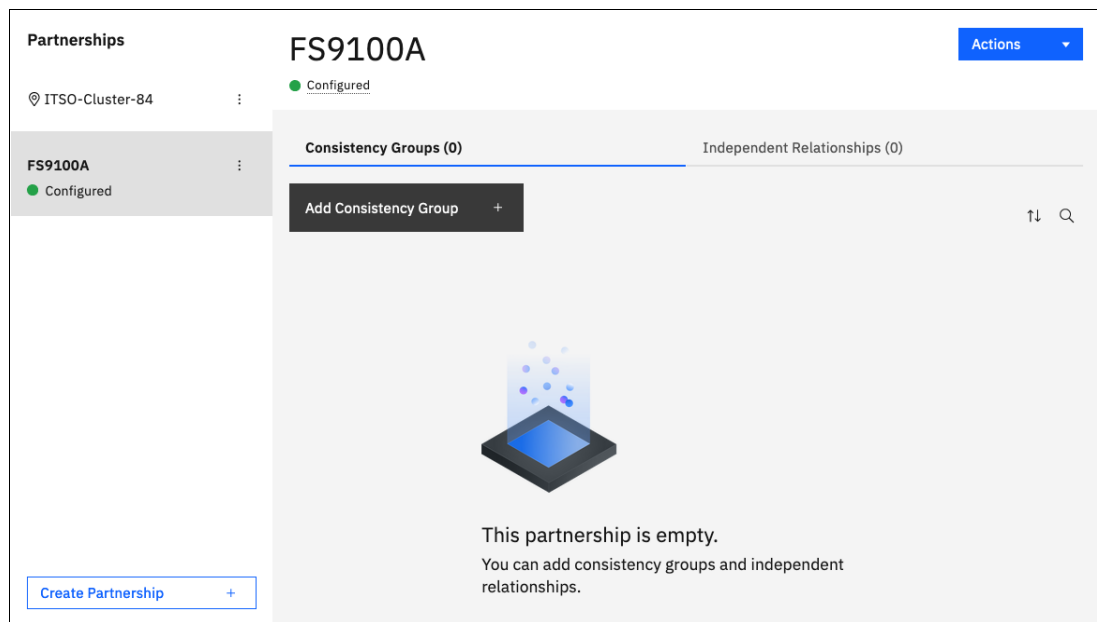


Figure 10-113 Fully configured FC partnership

10.9.2 Creating Remote Copy relationships

This section shows how to create RC relationships for volumes with their respective remote targets. Before creating a relationship between a volume on the local system and a volume on a remote system, both volumes must exist and have the same virtual size.

To create an RC relationship, complete the following steps:

1. Select **Copy Services** → **Remote Copy**.
2. Select the target system with which you will create an RC relationship.
3. If you want to add the relationship to an existing consistency group, select the consistency group for which you want to create the relationship and click **Create Relationship**, as shown in Figure 10-114.

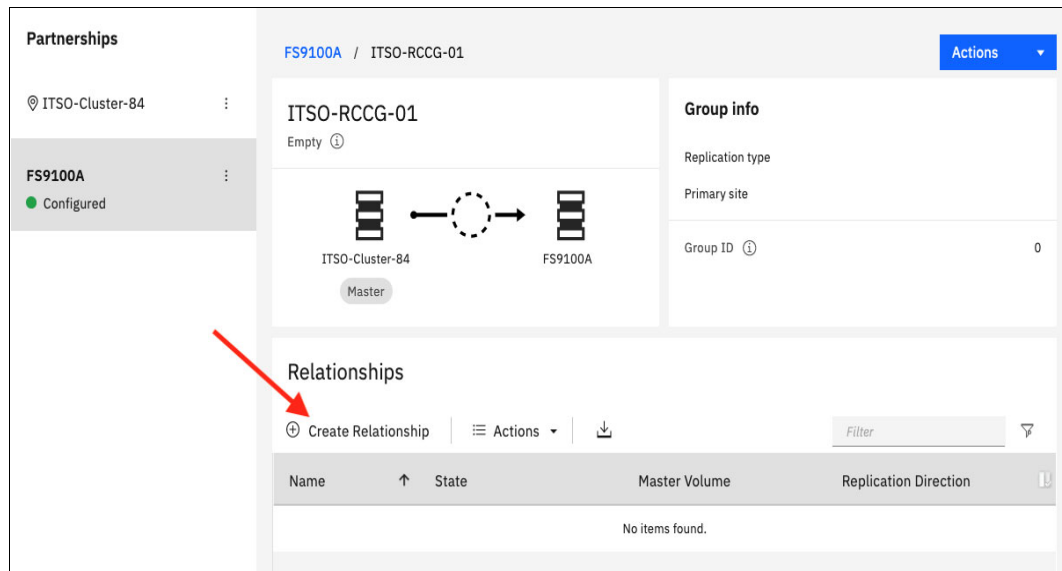


Figure 10-114 Creating a Remote Copy relationship in an existing consistency group

4. If you want to add a stand-alone relationship, select the **Independent Relationships** tab and click **Create Relationship**, as shown in the Figure 10-115.

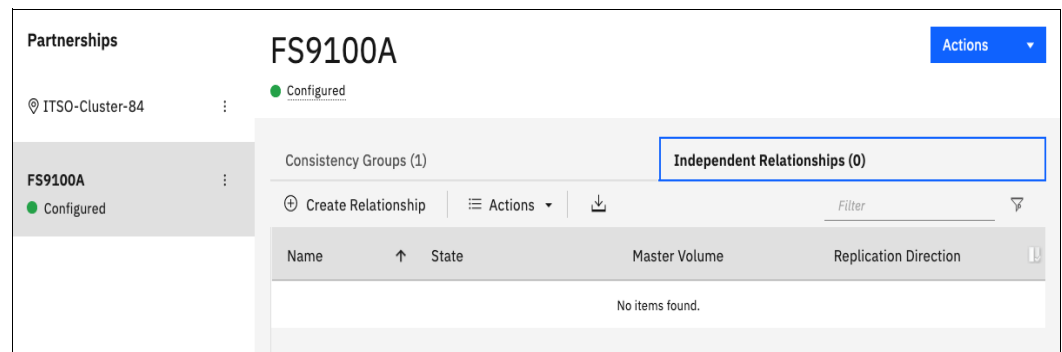


Figure 10-115 Creating a stand-alone Remote Copy relationship

5. In the Create Relationship window, select one of the following types of relationships that you want to create, as shown in Figure 10-116:

- **Metro Mirror**
- **Global Mirror** (with or without Consistency Protection)
- **Global Mirror with Change Volumes**

Click **Next**.

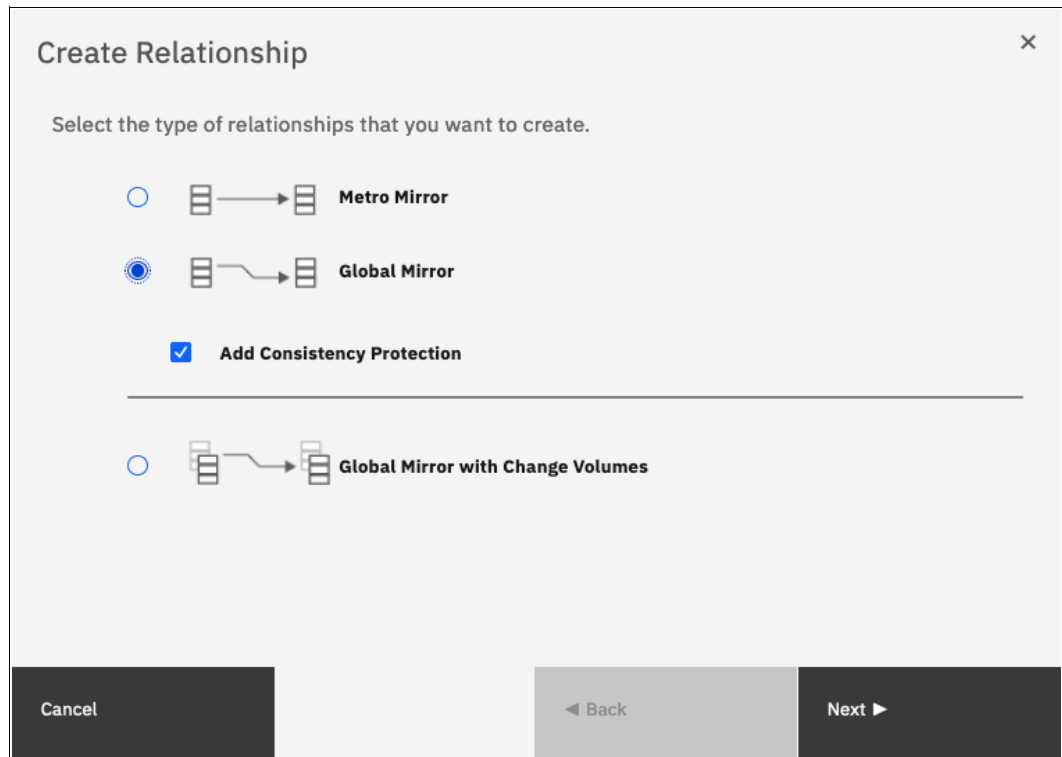


Figure 10-116 Creating a Remote Copy relationship

6. In the next window, select the target system for this RC relationship and click **Next**, as shown in the Figure 10-117, “Selecting the target system for the RC relationship” on page 699.

Create Relationship ×

Where are the auxiliary volumes located?

On this system

On another system

FS9100A ▼

Cancel ◀ Back Next ▶

Figure 10-117 Selecting the target system for the RC relationship

7. Select the master and auxiliary volumes, as shown in Figure 10-118. Click **ADD**.

Figure 10-118 Selecting the master and auxiliary volumes

Important: The master and auxiliary volumes must be of equal size. Therefore, only the targets with the correct size are shown in the list for a specific source volume.

8. In the next window, you can add change volumes if needed, as shown in Figure 10-119. Click **Finish**.

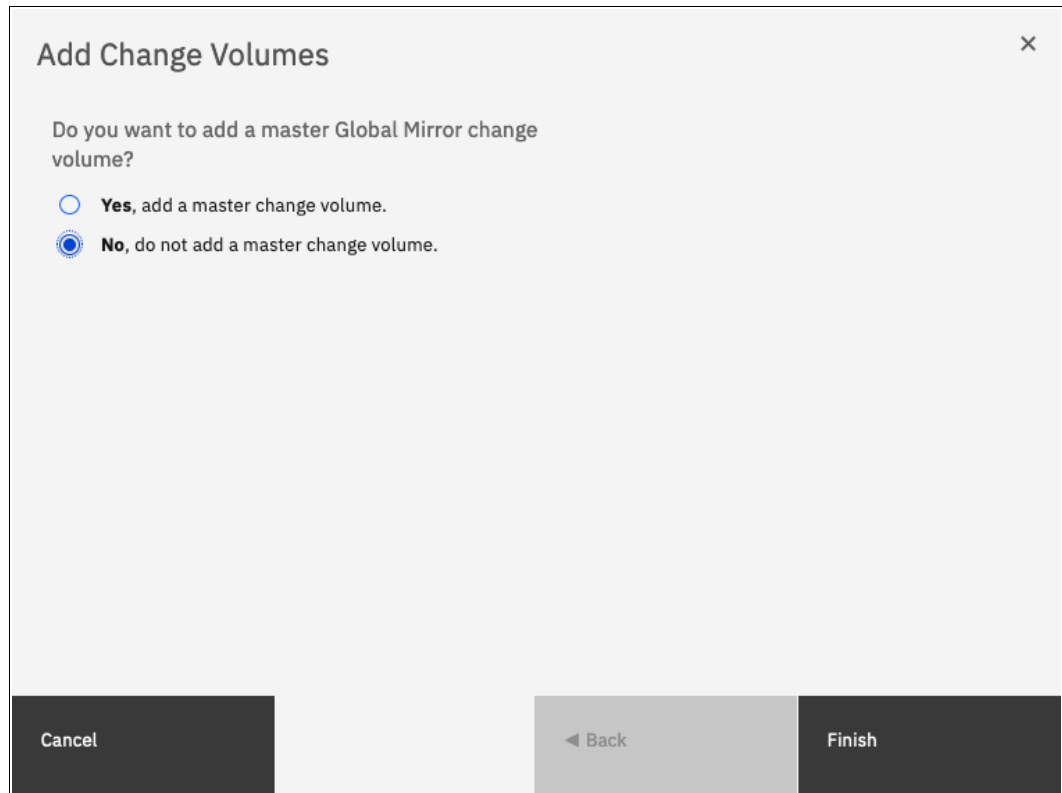


Figure 10-119 Add Change Volumes window

9. If you want to add more relationships, repeat steps 7 on page 700 and 8 on page 701, as shown in the Figure 10-120. When you are finished creating more relationships, click **Next**.

Create Relationship ×

Select the master and auxiliary volumes to use in the relationship.

Master **Auxiliary**

ITSO-FC-VOL-01 ⇌ ITSO-RC-TGT-VOL1 ✗

Figure 10-120 Checking and adding the relationship

10. In the next window, select whether the volumes are synchronized so that the relationship is created, as shown in Figure 10-121. Click **Next**.

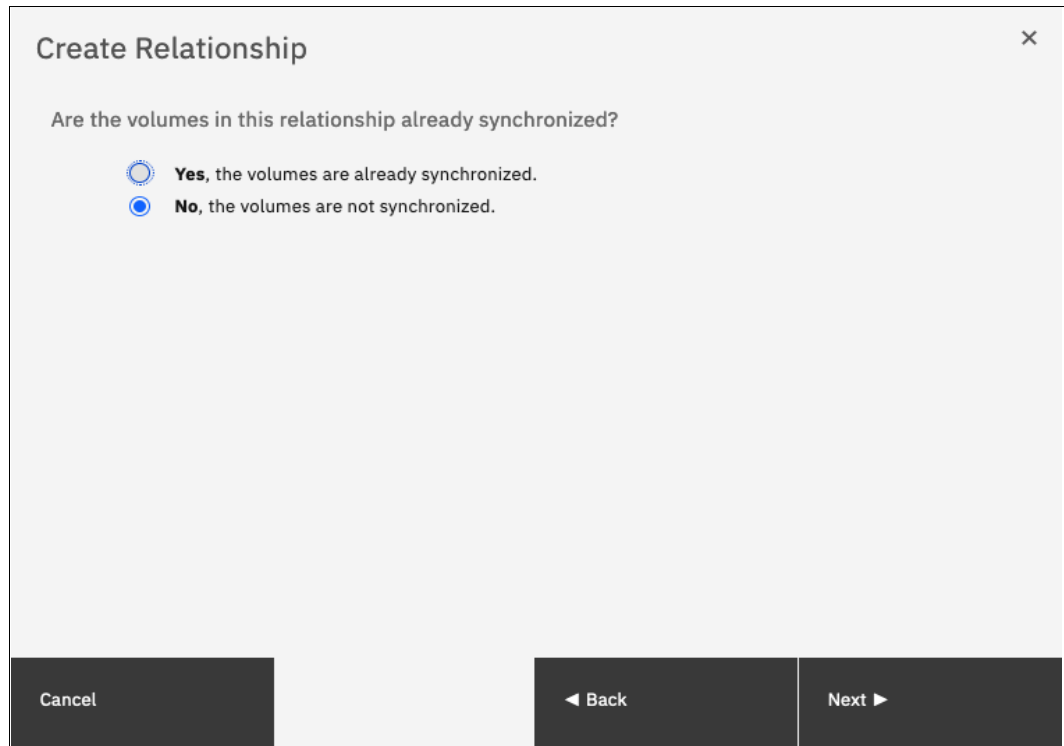


Figure 10-121 Selecting whether the volumes are synchronized

11. Select whether you want to start synchronizing the Master and Auxiliary volumes at the time of creation of the relationship or start the copy later, as shown in Figure 10-122. Click **Finish**.

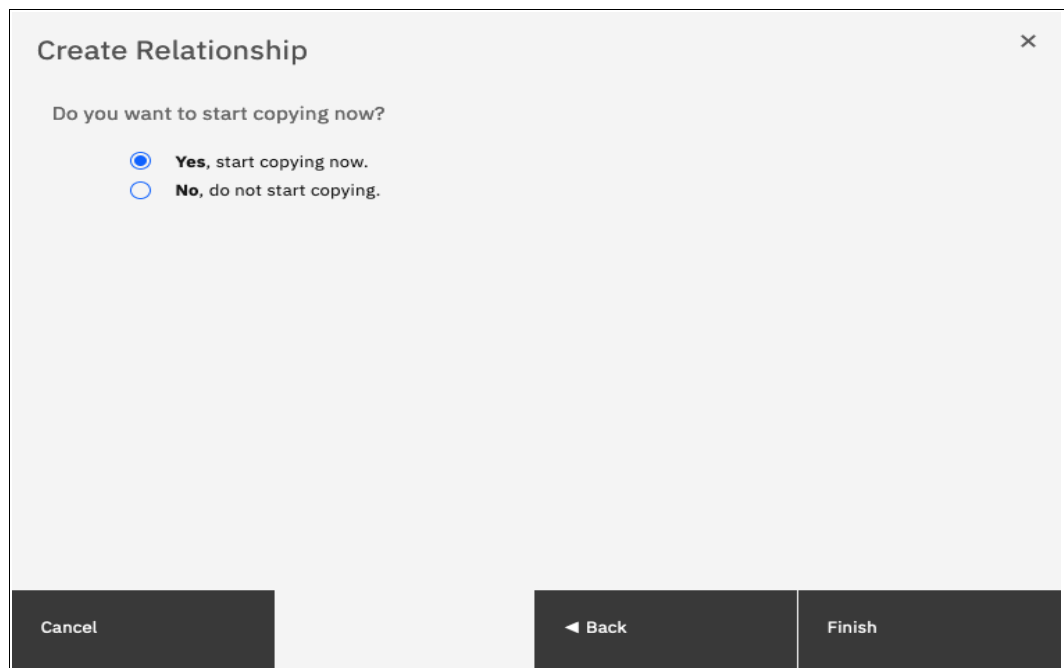


Figure 10-122 Start copying the Remote Copy relationship

Note: If the volumes are not synchronized, the initial copy copies the entire source volume to the remote target volume. If you suspect that the volumes are different or if you have any other doubts, synchronize them to ensure consistency on both sides of the relationship.

10.9.3 Creating a consistency group

To create a consistency group, complete the following steps:

1. Select **Copy Services** → **Remote Copy**, and select the target system of the RC. Then, click **Add Consistency Group**, as shown in Figure 10-123.

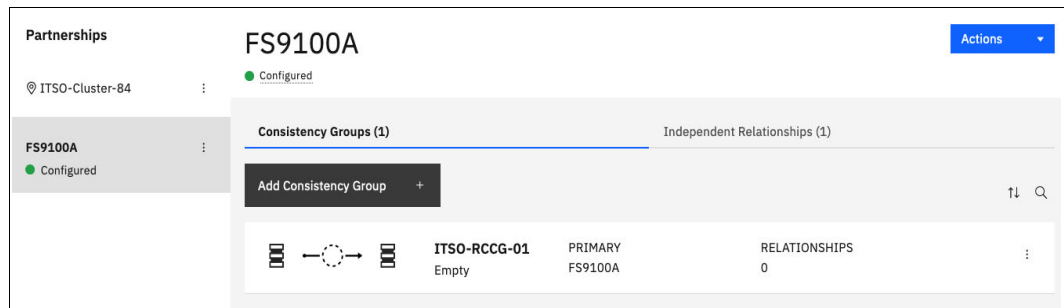


Figure 10-123 Creating a Remote Copy consistency group

2. Enter a name for the consistency group, select the target system, and click **Add**, as shown in Figure 10-124. The consistency group is added to the configuration with no relationships.

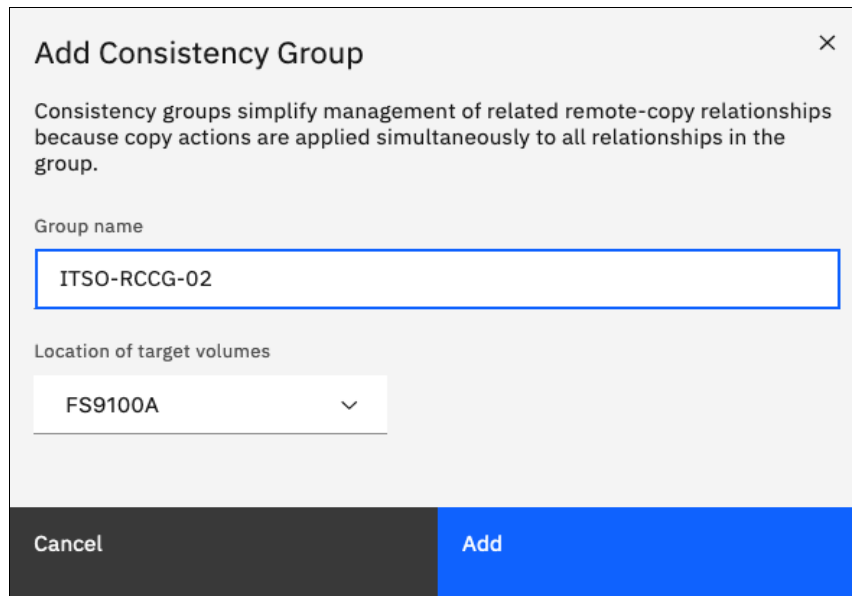


Figure 10-124 Entering a name for the new consistency group

- Then, you can either add existing stand-alone relationships to the recently added consistency group, by selecting the **Independent Relationships** tab, right-clicking the relationship and clicking **Add to Consistency Group**, or you can create new relationships directly to this consistency group, by selecting it in the **Consistency Group** tab, as shown in the Figure 10-125, and then clicking **Create Relationship**.

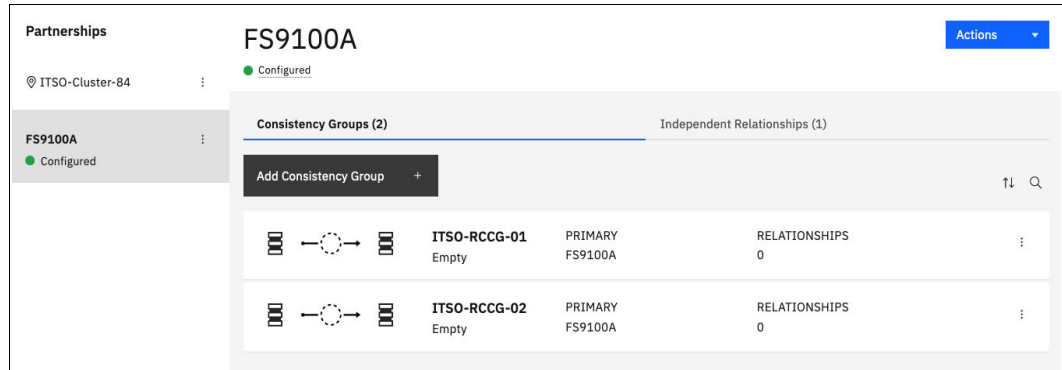


Figure 10-125 Remote Copy Consistency Groups tab

To create an RC relationship, see 10.9.2, “Creating Remote Copy relationships” on page 697.

10.9.4 Renaming Remote Copy relationships

To rename one or multiple RC relationships, complete the following steps:

- Select **Copy Services** → **Remote Copy**.
- Select the appropriate tab for the relationship that you want to rename, whether it is part of a consistency group or an independent relationship. If it is part of a consistency group, hover your cursor over the consistency group’s name to view its relationships.
- Right-click the relationships to be renamed and select **Rename**, as shown in Figure 10-126.

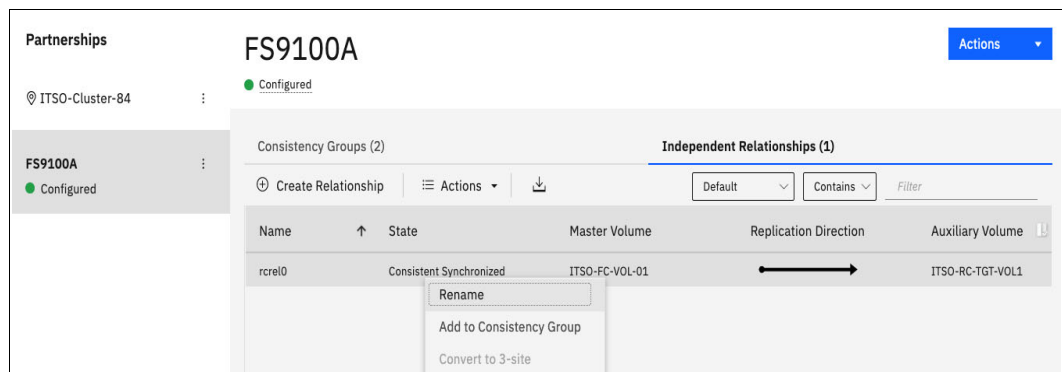


Figure 10-126 Renaming Remote Copy relationships

4. Enter the new name that you want to assign to the relationships and click **Rename**, as shown in Figure 10-127.

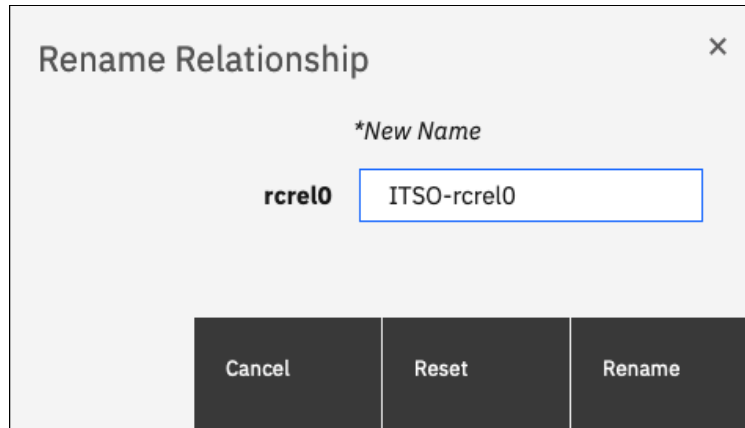


Figure 10-127 Renaming Remote Copy relationships

RC relationship name: You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The RC name can be 1 - 15 characters. Blanks cannot be used.

10.9.5 Renaming a Remote Copy consistency group

To rename an RC consistency group, complete the following steps:

1. Select **Copy Services** → **Remote Copy**.
2. Select the target system for the RC consistency group that you want to rename, click the three dots for the consistency group to be renamed, and select **Rename Group**, as shown in Figure 10-128.

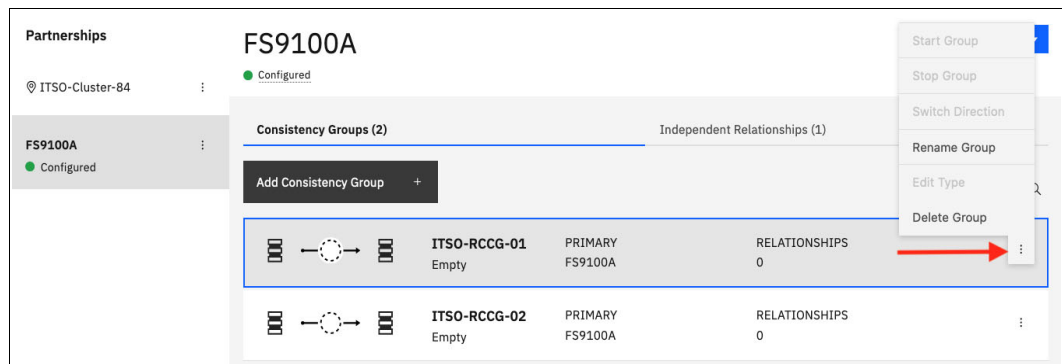


Figure 10-128 Renaming a Remote Copy consistency group

3. Enter the new name that you want to assign to the consistency group and click **Rename**, as shown in Figure 10-129.

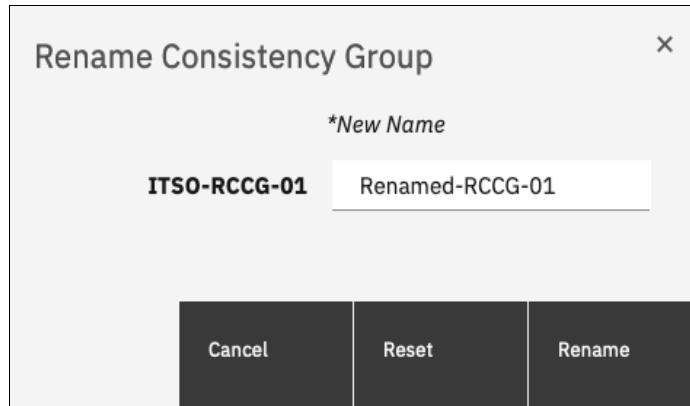


Figure 10-129 Entering a new name for a consistency group

RC consistency group name: You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (`_`) character. The RC name can be 1 - 15 characters. Blanks cannot be used.

10.9.6 Moving stand-alone Remote Copy relationships to a consistency group

To add one or multiple stand-alone relationships to an RC consistency group, complete the following steps:

1. Select **Copy Services** → **Remote Copy**.
2. Select the target system for the relationship to be moved, and go to the **Independent Relationships** tab. Right-click the relationship to be moved and select **Add to Consistency Group**, as shown in Figure 10-130.

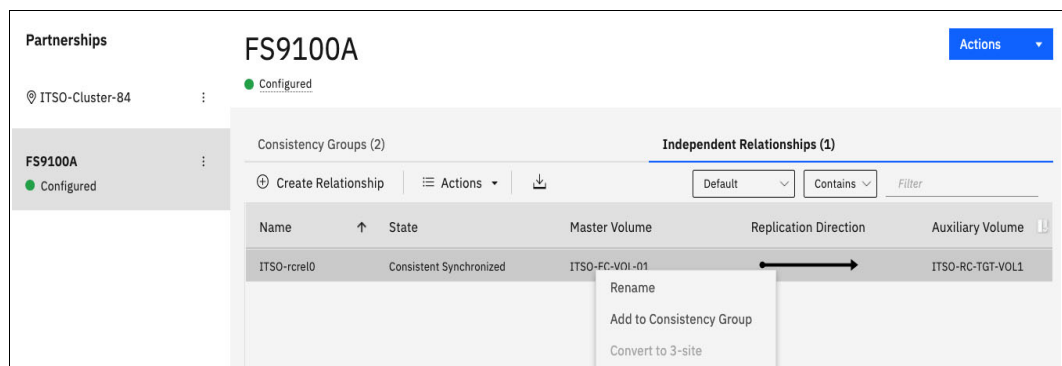


Figure 10-130 Moving relationships to a consistency group

3. Select the consistency group for this RC relationship by using the menu, as shown in Figure 10-131. Click **Add to Consistency Group** to confirm your changes.

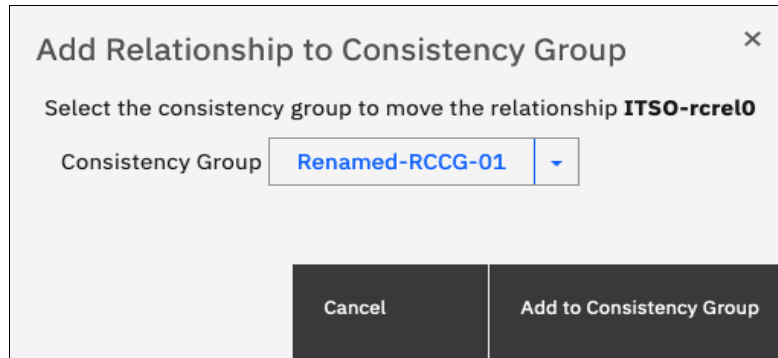


Figure 10-131 Selecting the consistency group to add the relationships to

10.9.7 Removing Remote Copy relationships from a consistency group

To remove one or multiple relationships from an RC consistency group, complete the following steps:

1. Select **Copy Services** → **Remote Copy**.
2. Select the target RC system and go to the **Consistency Groups** tab. Next, hover your cursor over the consistency group to view its relationships.
3. Right-click the relationships to be removed and select **Remove from Consistency Group**, as shown in Figure 10-132.

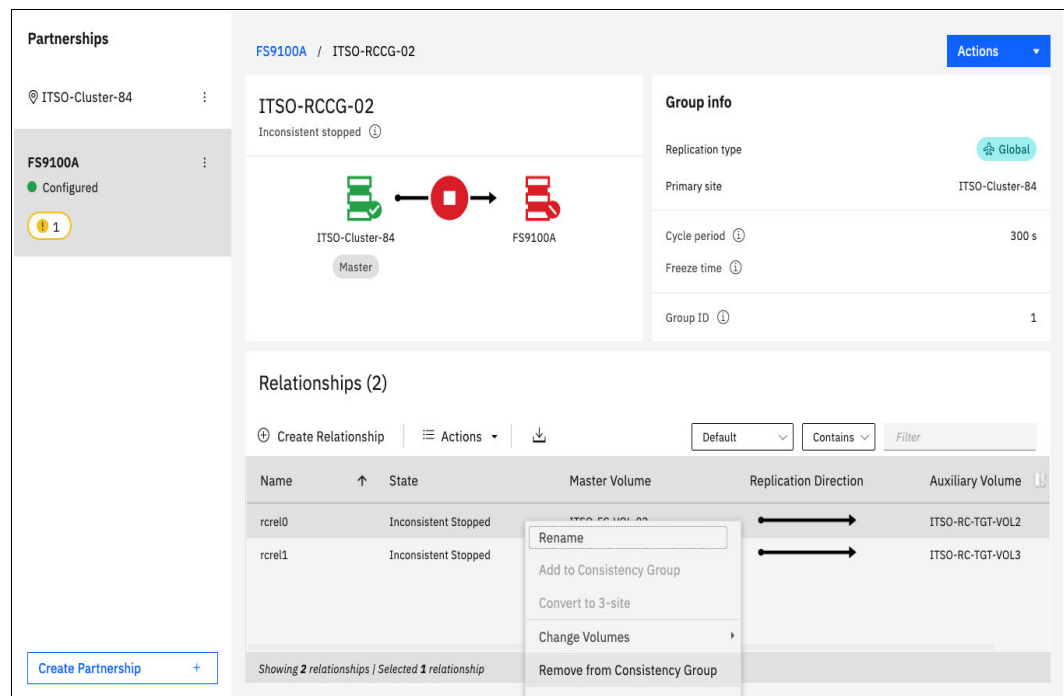


Figure 10-132 Removing relationships from a consistency group

- Confirm your selection and click **Remove**, as shown in Figure 10-133.

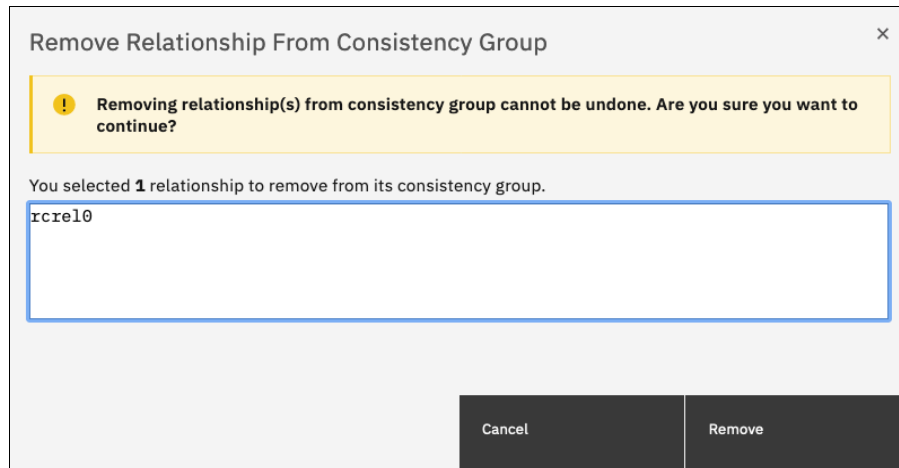


Figure 10-133 Confirming the removal of relationships from a consistency group

10.9.8 Starting Remote Copy relationships

When an RC relationship is created, the RC process can be started. Only relationships that are not members of a consistency group, or the entire consistency group, can be started.

To start one or multiple stand-alone relationships, complete the following steps:

- Select **Copy Services** → **Remote Copy**.
- Select the target RC system and go to the **Independent Relationships** tab. Next, right-click the relationships to be started and select **Start**, as shown in Figure 10-134.

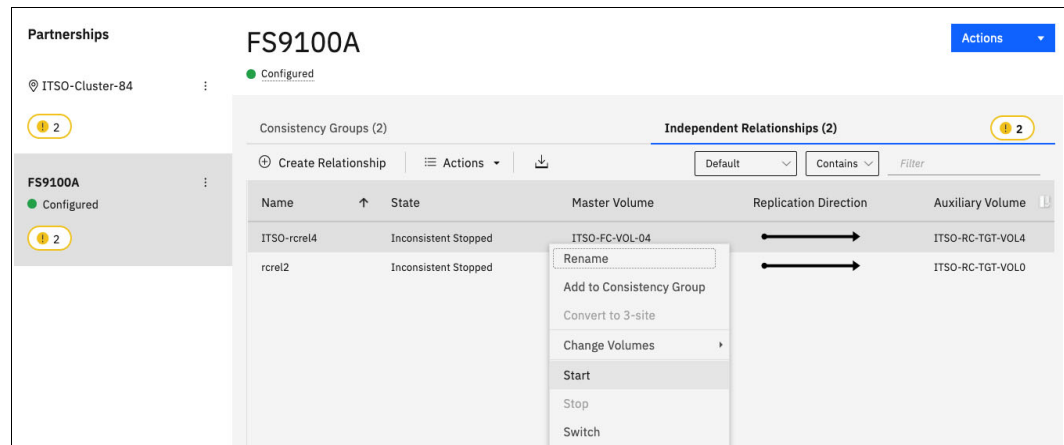


Figure 10-134 Starting Remote Copy relationships

10.9.9 Starting a Remote Copy consistency group

When an RC consistency group is created, the RC process can be started for all the relationships that are part of the consistency groups.

To start a consistency group, select **Copy Services** → **Remote Copy**, select the target RC system, and go to the **Consistency Groups** tab. Click the three dots for the consistency group to be started, and select **Start Group**, as shown in Figure 10-135.

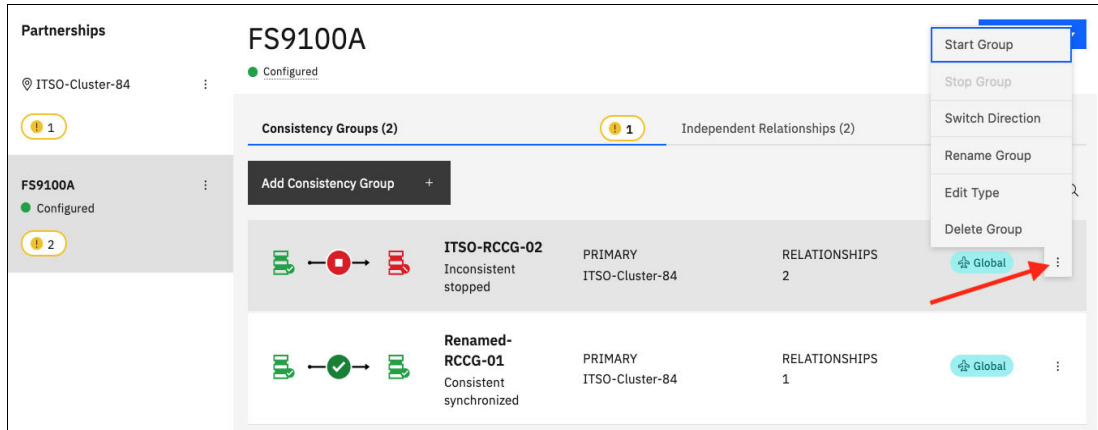


Figure 10-135 Starting a Remote Copy consistency group

10.9.10 Switching a relationship copy direction

When an RC relationship is in the Consistent synchronized state, the copy direction for the relationship can be changed. Only relationships that are not member of a consistency group, or the entire consistency group, can be switched.

Important: When the copy direction is switched, it is crucial that no outstanding I/O exists to the volume that changes from primary to secondary because all of the I/O is disallowed to that volume when it becomes the secondary. Therefore, careful planning is required before you switch the copy direction for a relationship.

To switch the direction of a stand-alone RC relationship, complete the following steps:

1. Select **Copy Services** → **Remote Copy**.
2. Select the target RC system for the relationship to be switched, and go to the **Independent Relationships** tab. Right-click the relationship to be switched and select **Switch**, as shown in Figure 10-136.

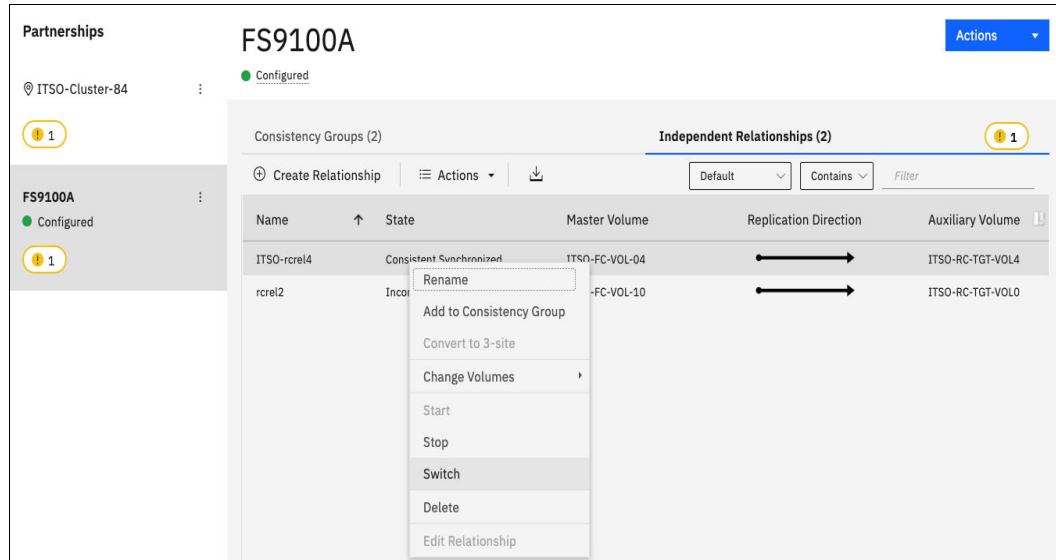


Figure 10-136 Switching the Remote Copy relationship direction

3. Because the master-auxiliary relationship direction is reversed, write access is disabled on the new auxiliary volume (former master volume), and it is enabled on the new master volume (former auxiliary volume). A warning message is displayed, as shown in Figure 10-137. Click **Yes**.

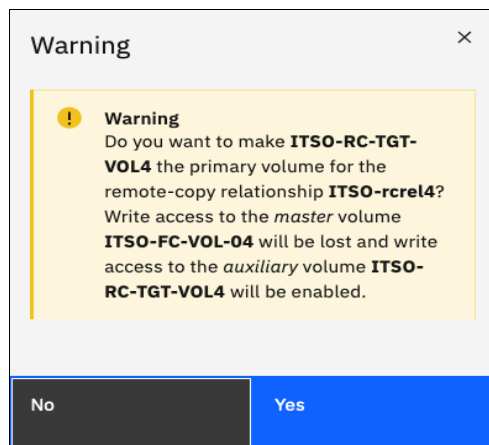


Figure 10-137 Switching the master-auxiliary direction of a relationships changes the write access

When an RC relationship is switched, an icon is displayed in the Remote Copy window and the Replication Direction is also updated, as shown in Figure 10-138.

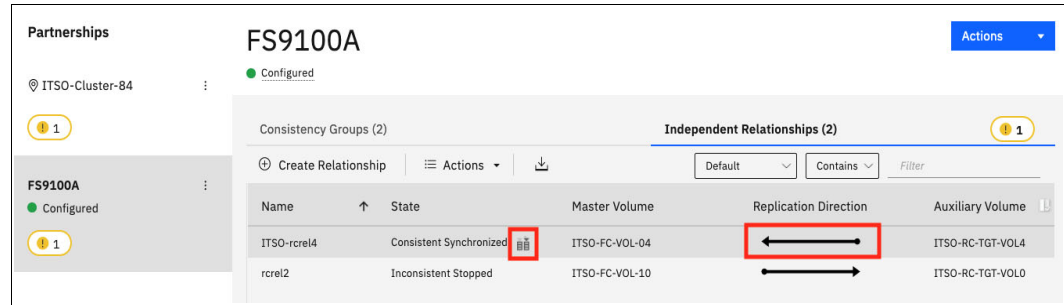


Figure 10-138 Switched Remote Copy relationship

10.9.11 Switching a consistency group direction

When an RC consistency group is in the consistent synchronized state, the copy direction for the consistency group can be changed.

Important: When the copy direction is switched, it is crucial that no outstanding I/O exists to the volume that changes from primary to secondary because all the I/O is disallowed to that volume when it becomes the secondary. Therefore, careful planning is required before you switch the copy direction for a relationship.

To switch the direction of an RC consistency group, complete the following steps:

1. Select **Copy Services** → **Remote Copy**.
2. Select the target RC system and go to the **Consistency Groups** tab. Next, click the three dots for the consistency group to be switched and select **Switch Direction**, as shown in Figure 10-139.

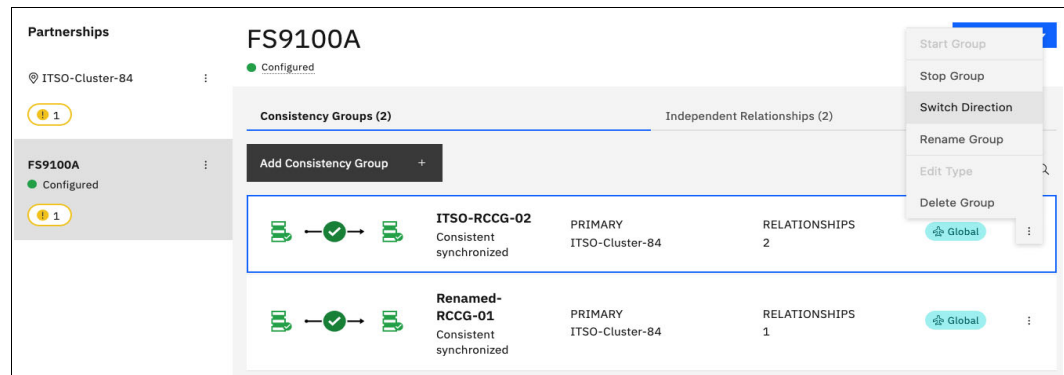


Figure 10-139 Switching a consistency group direction

3. Because the master-auxiliary relationship direction is reversed, write access is disabled on the new auxiliary volume (former master volume), while it is enabled on the new master volume (former auxiliary volume). A warning message is displayed, as shown in Figure 10-140 on page 713. Click **Yes**.

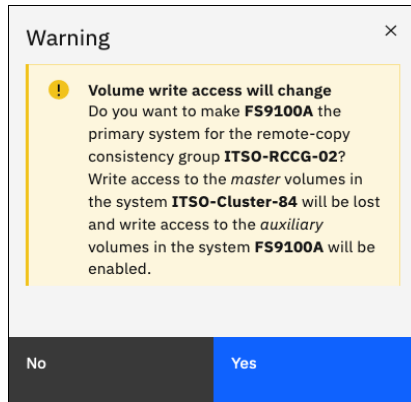


Figure 10-140 Switching the direction of a consistency group changes the write access

10.9.12 Stopping Remote Copy relationships

When an RC relationship is created and started, the RC process can be stopped. Only relationships that are not members of a consistency group, or the entire consistency group, can be stopped.

To stop one or multiple relationships, complete the following steps:

1. Select **Copy Services** → **Remote Copy**.
2. Select the target RC system and go to the **Independent Relationships** tab. Right-click the relationships to be stopped and select **Stop**, as shown in Figure 10-141.

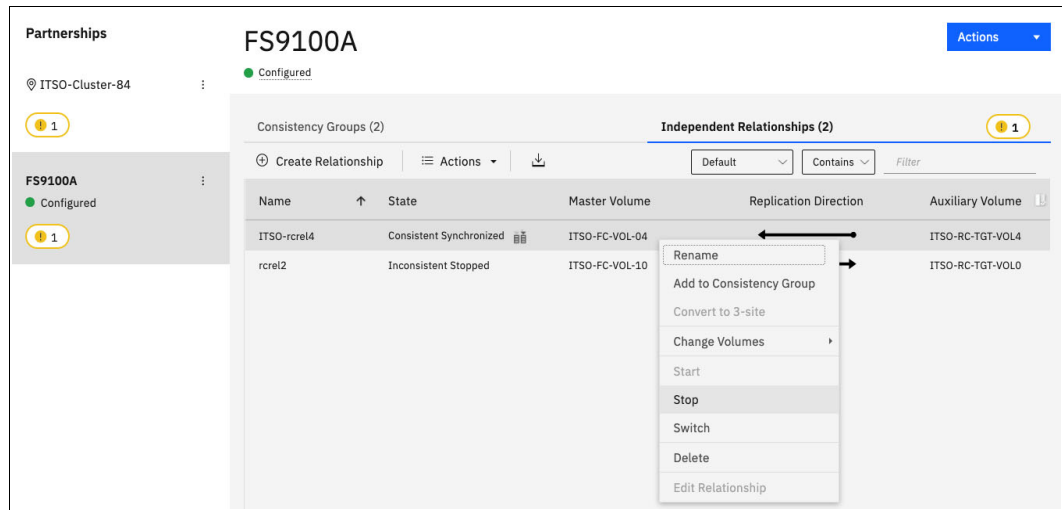


Figure 10-141 Stopping a Remote Copy relationship

- When an RC relationship is stopped, access to the auxiliary volume can be changed so that it can be read and written by a host. A confirmation message appears, as shown in Figure 10-142.

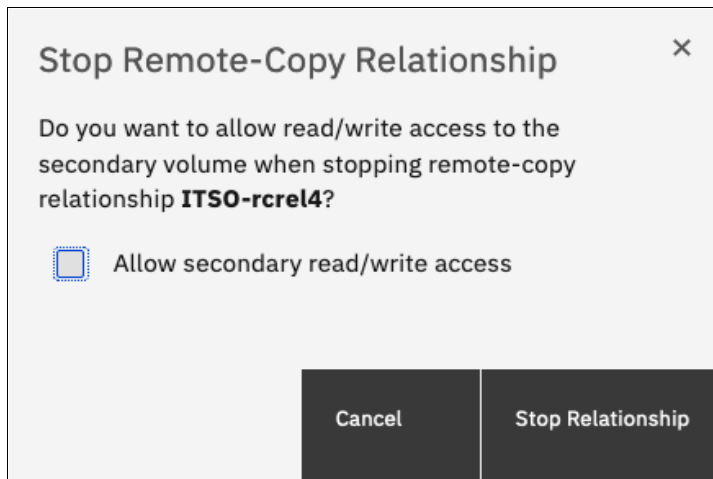


Figure 10-142 Granting read/write access to the auxiliary volume

10.9.13 Stopping a consistency group

When an RC consistency group is created and started, the RC process can be stopped.

To stop a consistency group, complete the following steps:

- Select **Copy Services** → **Remote Copy**.
- Select the target RC system and go to the **Consistency Groups** tab. Next, click the three dots for the consistency group to be stopped and select **Stop Group**, as shown in Figure 10-143.

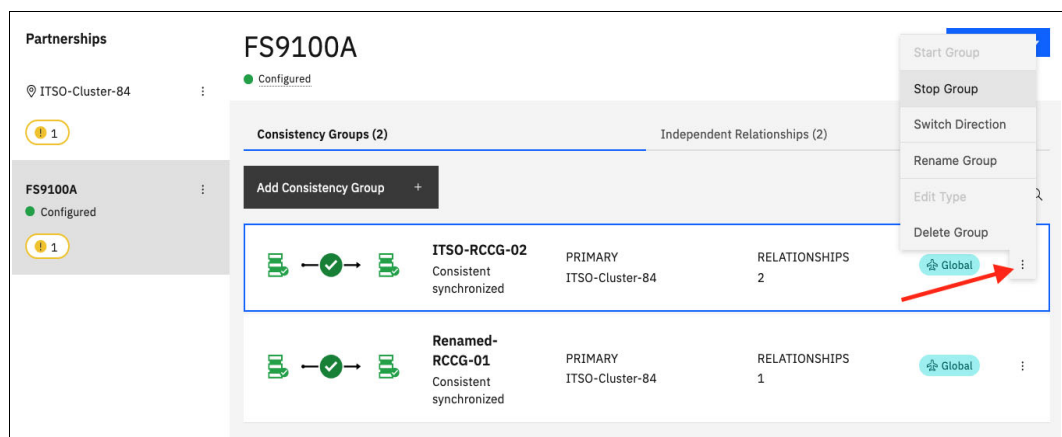


Figure 10-143 Stopping a consistency group

- When an RC consistency group is stopped, access to the auxiliary volumes can be changed so it can be read and written by a host. A confirmation message opens, as shown in Figure 10-144 on page 715.

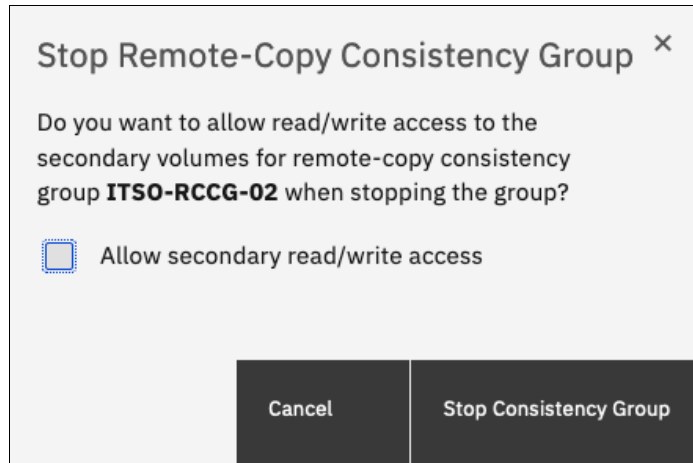


Figure 10-144 Granting read/write access write to the auxiliary volumes

10.9.14 Deleting Remote Copy relationships

To delete RC relationships, complete the following steps:

1. Select **Copy Services** → **Remote Copy**.
2. Select the target RC system and go to the **Independent Relationships** tab. Next, right-click the relationships that you want to delete and select **Delete**, as shown in Figure 10-145.

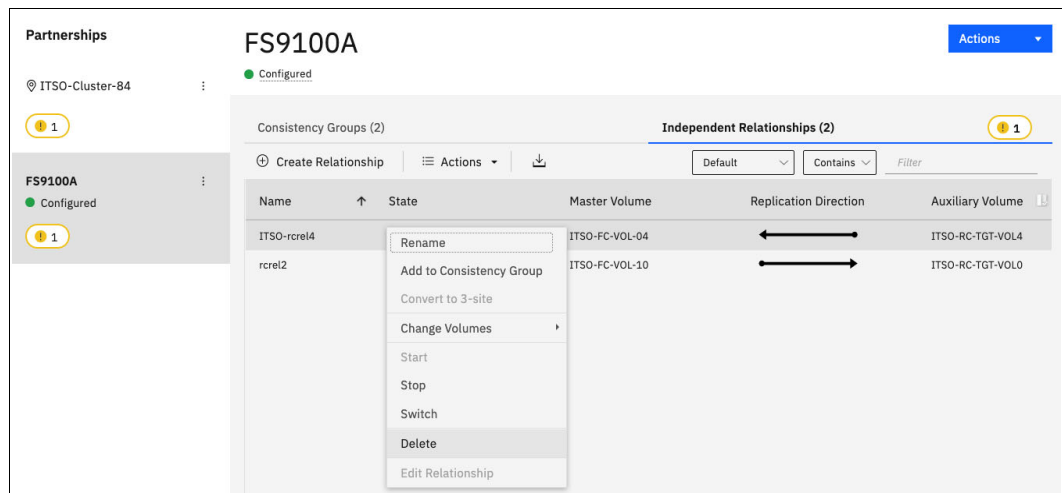


Figure 10-145 Deleting Remote Copy relationships

3. A confirmation message appears that requests that the user enter the number of relationships to be deleted, as shown in Figure 10-146.

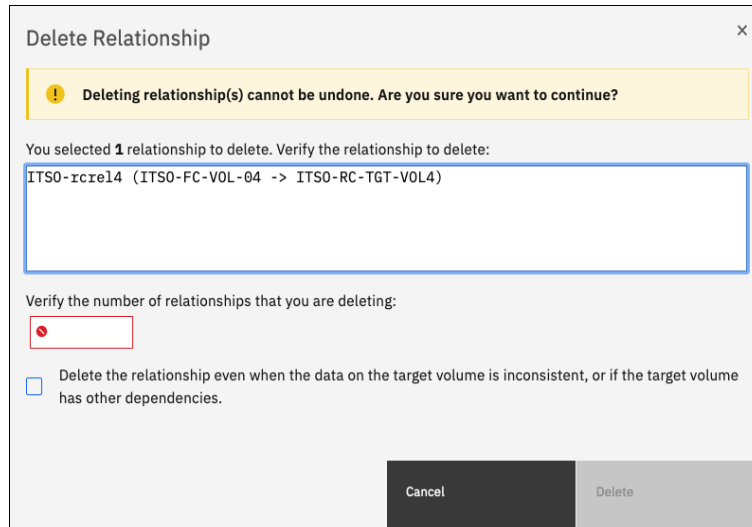


Figure 10-146 Confirming a relationship deletion

10.9.15 Deleting a consistency group

To delete an RC consistency group, complete the following steps:

1. Select **Copy Services** → **Remote Copy**.
2. Select the target RC system and go to the **Consistency Groups** tab. Next, click the three dots for the consistency group that you want to delete and select **Delete Group**, as shown in Figure 10-147.

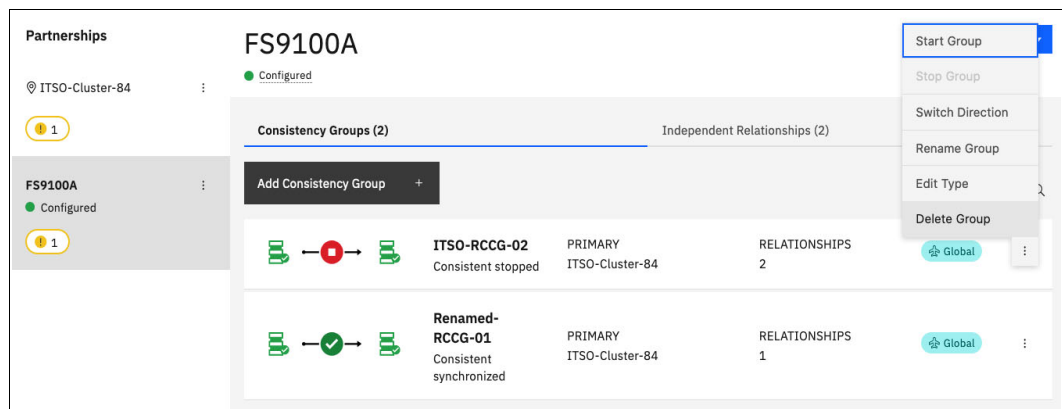


Figure 10-147 Deleting a consistency group

3. A confirmation message opens, as shown in Figure 10-148. Click **Yes**.

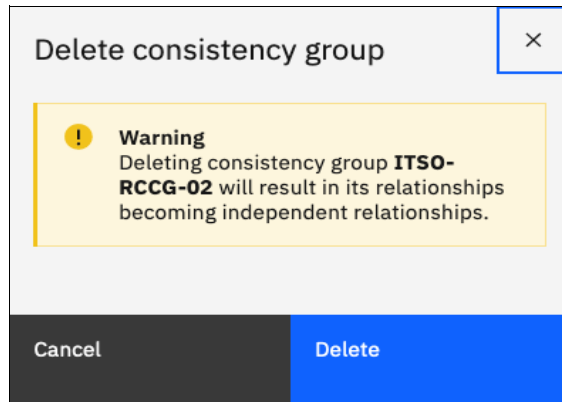


Figure 10-148 Confirming a consistency group deletion

Important: Deleting a consistency group does *not* delete its RC mappings.

10.10 Remote Copy memory allocation

Copy Services features require that small amounts of volume cache be converted from cache memory into bitmap memory to allow the functions to operate at an I/O group level. If you do not have enough bitmap space that is allocated when you try to use one of the functions, the configuration cannot be completed.

The total memory that can be dedicated to these functions is not defined by the physical memory in the system. The memory is constrained by the software functions that use the memory.

For every RC relationship that is created on an IBM Spectrum Virtualize system, a bitmap table is created to track the copied grains. By default, the system allocates 20 MiB of memory for a minimum of 2 TiB of remote copied source volume capacity. Every 1 MiB of memory provides the following volume capacity for the specified I/O group: for 256 KiB grains size, 2 TiB of total MM, GM, or active-active volume capacity.

To help calculate the memory requirements and confirm that your system can accommodate the total installation size, see the values in Table 10-15.

Table 10-15 Memory allocation for Remote Copy services

Minimum allocated bitmap space	Default allocated bitmap space	Maximum allocated bitmap space	Minimum functionality when using the default values ^a
0	20 MiB	512 MiB	40 TiB of remote mirroring volume capacity

a. RC includes MM, GM, and active-active relationships.

When you configure GMCV, two internal FlashCopy mappings are created for each change volume.

Two bitmaps exist for MM, GM, and HyperSwap active-active relationships. For MM/GM relationships, one is used for the master clustered system and one is used for the auxiliary system because the direction of the relationship can be reversed. For active-active relationships, which are configured automatically when HyperSwap volumes are created, one bitmap is used for the volume copy on each site because the direction of these relationships can be reversed.

MM/GM relationships do not automatically increase the available bitmap space. You might need to run the `chiogrp` command to manually increase the space in one or both of the master and auxiliary systems.

You can modify the resource allocation for each I/O group of an SVC system by selecting **Settings** → **System** and clicking the **Resources** menu, as shown in Figure 10-149. At the time of writing, this GUI option is not available for other IBM Spectrum Virtualize based systems, so the resource allocation can be adjusted by running the `chiogrp` command. For more information about this command, see [IBM Documentation](#).

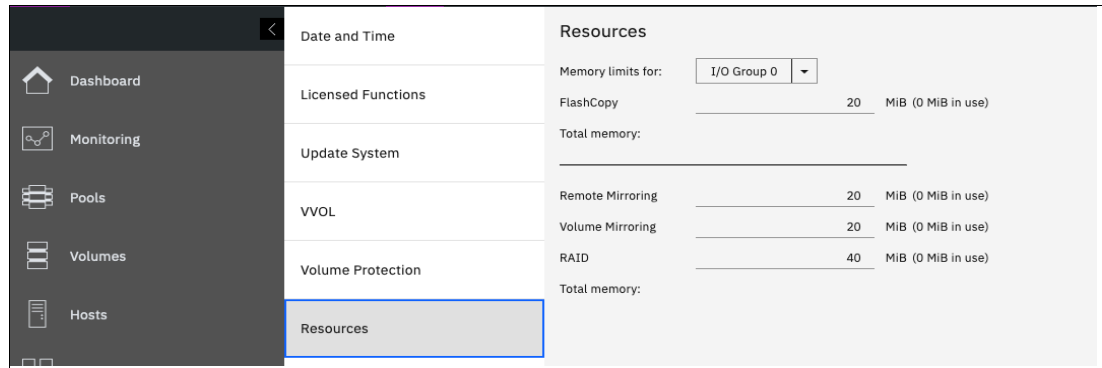


Figure 10-149 Modifying resources allocation

10.11 Troubleshooting Remote Copy

RC (MM and GM) features two primary error codes that are displayed:

- ▶ A 1920 error can be considered as a voluntary stop of a relationship by the system when it evaluates the replication causes errors on the hosts. A 1920 is a congestion error. This error means that the source, the link between the source and target, or the target cannot keep up with the requested copy rate. The system then triggers a 1920 error to prevent replication from having undesired effects on hosts.
- ▶ A 1720 error is a heartbeat or system partnership communication error. This error often is more serious because failing communication between your system partners involves extended diagnostic time.

10.11.1 1920 error

A 1920 error is deliberately generated by the system and is considered as a control mechanism. It occurs after 985003 (“Unable to find path to disk in the remote cluster (system) within the timeout period”) or 985004 (“Maximum replication delay has been exceeded”) events.

It can have several triggers, including the following probable causes:

- ▶ Primary system or SAN fabric problem (10%)
- ▶ Primary system or SAN fabric configuration (10%)
- ▶ Secondary system or SAN fabric problem (15%)
- ▶ Secondary system or SAN fabric configuration (25%)
- ▶ Intercluster link problem (15%)
- ▶ Intercluster link configuration (25%)

In practice, the most often overlooked cause is latency. GM has an RTT tolerance limit of 80 or 250 milliseconds, depending on the firmware version and the hardware model. A message that is sent from the source IBM Spectrum Virtualize system to the target system and the accompanying acknowledgment must have a total time of 80- or 250-millisecond round trip. That is, it must have up to 40- or 125-millisecond latency each way.

The primary component of your RTT is the physical distance between sites. For every 1000 kilometers (621.4 miles), you observe a 5-millisecond delay each way. This delay does not include the time that is added by equipment in the path. Every device adds a varying amount of time, depending on the device, but a good rule is 25 microseconds for pure hardware devices.

For software-based functions (such as compression that is implemented in applications), the added delay tends to be much higher (usually in the millisecond plus range.) The following is an example of a physical delay.

Company A has a production site that is 1900 kilometers (1180.6 miles) away from its recovery site. The network service provider uses a total of five devices to connect the two sites. In addition to those devices, Company A uses a SAN FC router at each site to provide FCIP to encapsulate the FC traffic between sites.

Now, there are seven devices and 1900 kilometers (1180.6 miles) of distance delay. All the devices are adding 200 microseconds of delay each way. The distance adds 9.5 milliseconds each way, for a total of 19 milliseconds. Combined with the device latency, the delay is 19.4 milliseconds of physical latency minimum, which is under the 80-millisecond limit of GM until you realize that this number is the best case number.

The link quality and bandwidth play a large role. Your network provider likely ensures a latency maximum on your network link. Therefore, be sure to stay as far beneath the GM RTT limit as possible. You can easily double or triple the expected physical latency with a lower quality or lower bandwidth network link. Then, you are within the range of exceeding the limit if high I/O occurs that exceeds the bandwidth capacity.

When you get a 1920 event, always check the latency first. The FCIP routing layer can introduce latency if it is not properly configured. If your network provider reports a much lower latency, you might have a problem at your FCIP routing layer. Most FCIP routing devices have built-in tools to enable you to check the RTT. When you are checking latency, remember that TCP/IP routing devices (including FCIP routers) report RTT by using standard 64-byte ping packets.

Effective transit time must be measured only by using packets that are large enough to hold an FC frame, or 2148 bytes (2112 bytes of payload and 36 bytes of header). Allow estimated resource requirements to be a safe amount because various switch vendors have optional features that might increase this size. After you verify your latency by using the proper packet size, proceed with normal hardware troubleshooting.

Before proceeding, look at the second largest component of your RTT, which is *serialization delay*. Serialization delay is the amount of time that is required to move a packet of data of a specific size across a network link of a certain bandwidth. The required time to move a specific amount of data decreases as the data transmission rate increases.

The amount of time in microseconds that is required to transmit a packet across network links of varying bandwidth capacity is compared. The following packet sizes are used:

- ▶ 64 bytes: The size of the common ping packet
- ▶ 1500 bytes: The size of the standard TCP/IP packet
- ▶ 2148 bytes: The size of an FC frame

Finally, your path (MTU) affects the delay that is incurred to get a packet from one location to another location. An MTU might cause fragmentation or be too large and cause too many retransmits when a packet is lost.

Note: Unlike 1720 errors, 1920 errors are deliberately generated by the system because it evaluated that a relationship can affect the host's response time. The system has no indication about if or when the relationship can be restarted. Therefore, the relationship cannot be restarted automatically and it must be done manually.

10.11.2 1720 error

The 1720 error (event ID 050020) is the other problem RC might encounter. The amount of bandwidth that is needed for system-to-system communications varies based on the number of nodes. It is important that it is not zero. When a partner on either side stops communication, a 1720 is displayed in your error log. According to the product documentation, there are no likely field-replaceable unit (FRU) breakages or other causes.

The source of this error is most often a fabric problem or a problem in the network path between your partners. When you receive this error, check your fabric configuration for zoning of more than one host bus adapter (HBA) port for each node per I/O group if your fabric has more than 64 HBA ports zoned. The suggested zoning configuration for fabrics is one port for each node per I/O group per fabric that is associated with the host.

For those fabrics with 64 or more host ports, this suggestion becomes a rule. Therefore, you see four paths to each volume discovered on the host because each host must have at least two FC ports from separate HBA cards, each in a separate fabric. On each fabric, each host FC port is zoned to two IBM Spectrum Virtualize N_Ports, where each N_Port comes from a different IBM Spectrum Virtualize node. This configuration provides four paths per volume. More than four paths per volume are supported but not recommended.

Improper zoning can lead to SAN congestion, which can inhibit remote link communication intermittently. Checking the zero buffer credit timer and port send delay percentage by using IBM Spectrum Control and comparing them against your sample interval reveals potential SAN congestion. If a zero buffer credit or port send delay percentage is more than 2% of the total time of the sample interval, it might cause problems.

Always ask your network provider to check the status of the link. If the link is acceptable, watch for repeats of this error. It is possible in a normal and functional network setup to have occasional 1720 errors, but multiple occurrences might indicate a larger problem.

If you receive multiple 1720 errors, recheck your network connection and then check the system partnership information to verify its status and settings. Then, perform diagnostics for every piece of equipment in the path between your two IBM Spectrum Virtualize systems. It

often helps to have a diagram that shows the path of your replication from both logical and physical configuration viewpoints.

Note: With Consistency Protection enabled on GM relationships, the system tries to resume the replication when possible. Therefore, it is not necessary to manually restart the failed relationship after a 1720 error is triggered.

If your investigations fail to resolve your RC problems, contact your IBM Support representative for a more complete analysis.



Ownership groups

The ownership groups feature, or object-based access control (OBAC), provides a method of implementing a multi-tenant solution on IBM FlashSystem systems. The ownership group principles of operations and implementation steps are provided in this chapter.

This chapter includes the following sections:

- ▶ 11.1, “Ownership groups principles of operations” on page 724
- ▶ 11.2, “Implementing ownership groups on a new system” on page 726
- ▶ 11.3, “Migrating existing objects to ownership groups” on page 731

11.1 Ownership groups principles of operations

Ownership groups enable the allocation of storage resources to several independent tenants with the assurance that one tenant cannot access resources that are associated with another tenant.

Ownership groups restrict access for users in the ownership group to only those objects that are defined within that ownership group. An owned object can belong to one ownership group. Users in an ownership group are restricted to viewing and managing objects within their ownership group. Users that are not in an ownership group can continue to view or manage all the objects on the system based on their defined user role, including objects within ownership groups.

Only users with Security Administrator roles (for example, *superuser*) can configure and manage ownership groups.

The system supports several resources that you assign to ownership groups:

- ▶ Child pools
- ▶ Volumes
- ▶ Volume groups
- ▶ Hosts
- ▶ Host clusters
- ▶ Host mappings
- ▶ IBM FlashCopy mappings
- ▶ FlashCopy consistency groups

An owned object can belong to only one ownership group. An owner is a user with an ownership group that can view and manipulate objects within that group.

Before you create ownership groups and assign resources and users, review the following guidelines:

- ▶ Users can be in only one ownership group at a time (applies to both local and remotely authenticated users).
- ▶ Objects can be within at most one ownership group.
- ▶ Global resources, such as drives, enclosures, and arrays, cannot be assigned to ownership groups.
- ▶ Global users that do not belong to an ownership group can view and manage (depending on their user role) all resources on the system, including the ones that belong to an ownership group, and users within an ownership group.
- ▶ Users within an ownership group cannot have the Security Administrator role. All Security Administrator role users are global users.
- ▶ Users within an ownership group can view or change resources within the ownership group in which they belong.

- ▶ Users within an ownership group cannot change any objects outside of their ownership group. This restriction includes global resources that are related to resources within the ownership group. For example, a user can change a volume in the ownership group, but not the drive that provides the storage for that volume.
- ▶ Users within an ownership group cannot view or change resources if those resources are assigned to another ownership group or are not assigned to any ownership group. However, users within ownership groups can view and display global resources. For example, users can display information on drives on the system because drives are a global resource that cannot be assigned to any ownership group.

When a user group is assigned to an ownership group, the users in that user group retain their role but are restricted to only those resources that belong to the same ownership group. The role that is associated with a user group can define the permitted operations on the system, and the ownership group can further limit access to individual resources. For example, you can configure a user group with the Copy Operator role, which limits user access to FlashCopy operations. Access to individual resources, such as a specific FlashCopy consistency group, can be further restricted by assigning it to an ownership group.

A child pool is a key requirement for the ownership groups feature. By defining a child pool and assigning it to an ownership group, the system administrator provides capacity for volumes that ownership group users can create or manage.

Depending on the type of resource, the owning group for the resource can be defined explicitly or inherited from explicitly defined objects. For example, a child pool needs an ownership group parameter to be set by a system administrator, but volumes that are created in that child pool automatically inherit the ownership group from a child pool. For more information about ownership inheritance, see [IBM FlashSystem 9200 documentation](#) and expand **Product overview** → **Technical overview** → **Ownership groups**.

When the user logs on to the management GUI or command-line interface (CLI), only resources that they have access to through the ownership group are available. Additionally, only events and commands that are related to the ownership group in which a user belongs are viewable by those users.

11.2 Implementing ownership groups on a new system

This section describes ownership group implementation process for a new system that has no volumes and users that must be migrated to ownership groups.

11.2.1 Creating an ownership group

To create the first ownership group, select **Access** → **Ownership Groups**, as shown in Figure 11-1. Enter a name for the ownership group, and click **Create Ownership Group**.

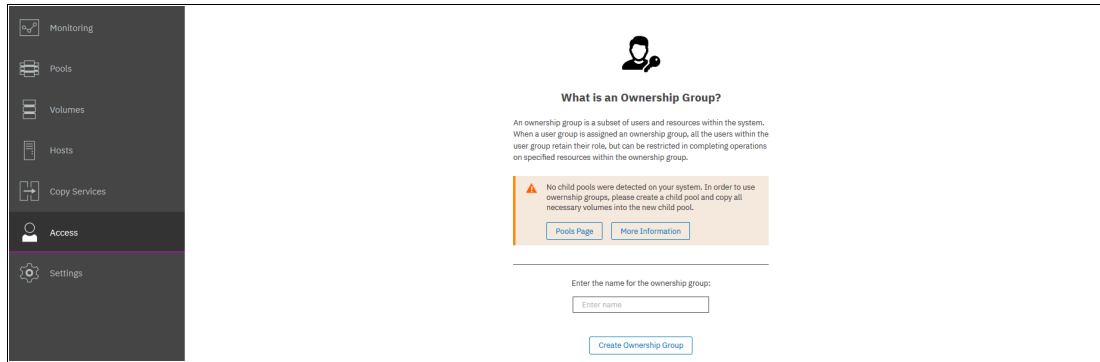


Figure 11-1 Creating the first ownership group

After the first group is created, the window changes to ownership group mode, as shown in Figure 11-2. The new ownership group has no user groups and no resources that are assigned to it.

To create more ownership groups, click **Create Ownership Group**.

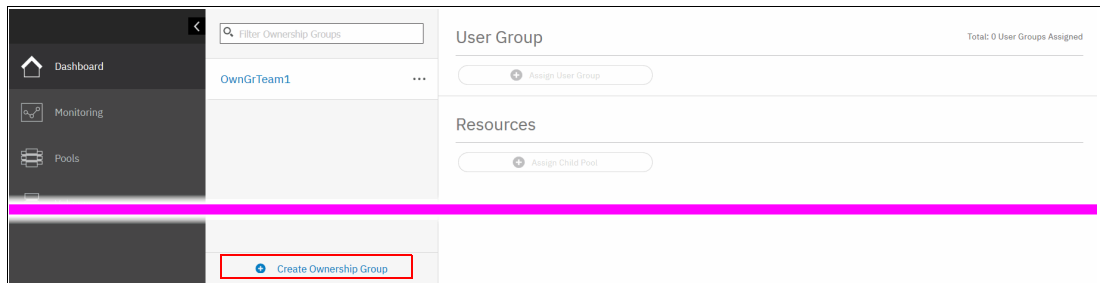


Figure 11-2 Ownership groups management window

11.2.2 Assigning users to an ownership group

To create accounts for users that use ownership group resources, the user group must be created and assigned to the ownership group. To create a user group, select **Access** → **Users by Group**, and click **Create User Group**. The Create User Group window opens, as shown in Figure 11-3 on page 727. Specify the User Group name, select an ownership group to tie this user group to, and select a role for the users in this group.

For a description of user roles, see [IBM FlashSystem 9200 documentation](#) and expand **Product overview** → **Technical overview** → **User roles**.

To create volume, host, and other objects in an ownership group, users must have an *Administrator* or *Restricted Administrator* role. Users with the *Security Administrator* role cannot be assigned to an ownership group.

You may also set up a user group to use remote authentication, if it is enabled. To do so, select the **Lightweight Directory Access Protocol (LDAP)** checkbox.

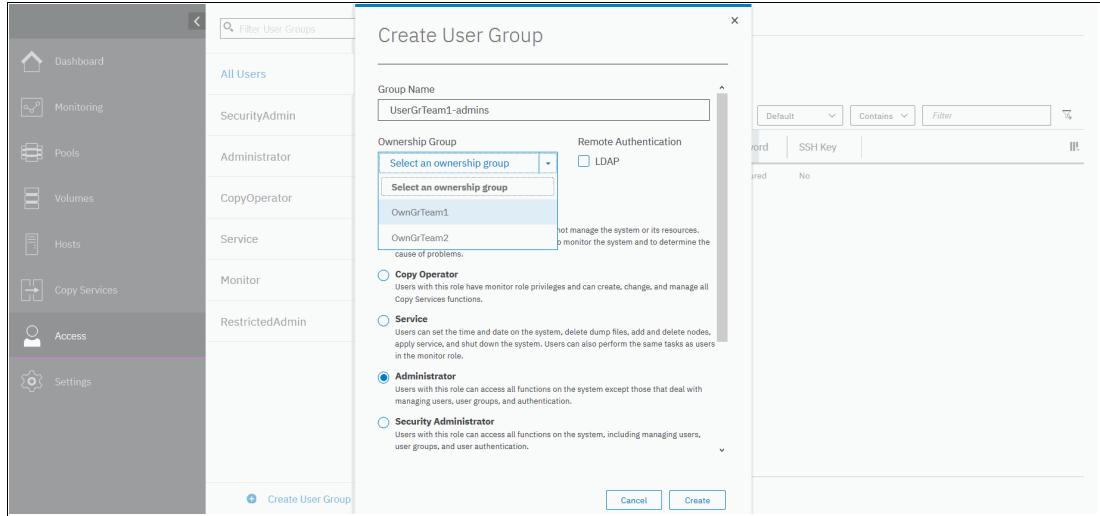


Figure 11-3 Creating and assigning a user group

Note: Users that use LDAP can belong to multiple user groups, but belong to only one ownership group that is associated with one of the user groups.

If remote authentication is not configured, you must create a user (or users) and assign it to a created user group, as shown in Figure 11-4.

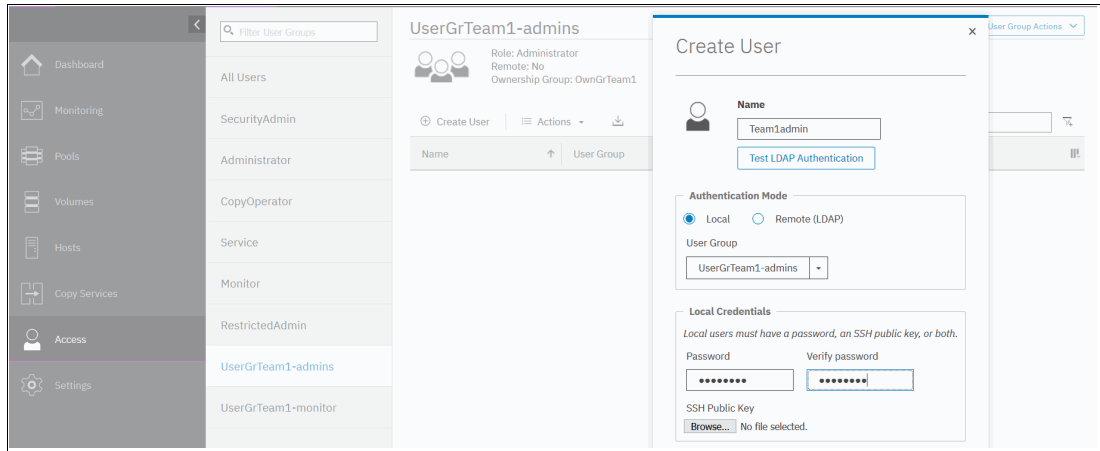


Figure 11-4 Creating a user

You can manage user groups that are assigned to an ownership group by selecting **Access** → **Ownership Groups**, as shown in Figure 11-5. To assign a user group that exists but is not assigned to any ownership group, click **Assign User Group**. To unassign user group, click the ... icon next to the assigned user group name.

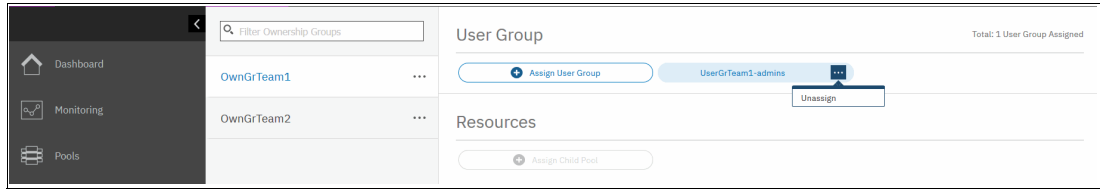


Figure 11-5 Unassigning user groups

Multiple user groups with different user roles may be assigned to one ownership group. For example, you may create and assign a user group with the Monitor role in addition to a group with the Administrator role to have two sets of users with different privilege levels accessing an ownership group’s resources.

11.2.3 Creating ownership group resources

To be able to create ownership group volumes and other resources, a child pool must be created and assigned to the ownership group. To do this task, select **Pools** → **Pools**, right-click a parent pool that is designated as a container for child pools, and click **Create Child Pool**, as shown in Figure 11-6.

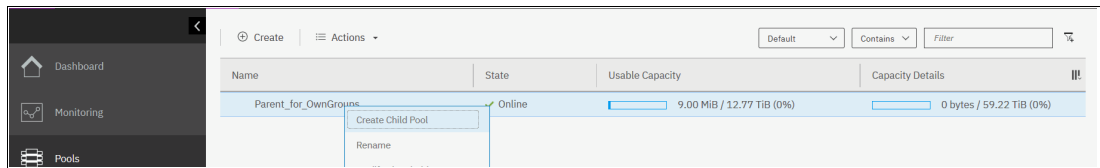


Figure 11-6 Creating a child pool

When creating a child pool, specify an ownership group for it and assign a part of the parent’s pool capacity, as shown in Figure 11-7. Ownership group objects can use only capacity that is provisioned for them with the child pool.

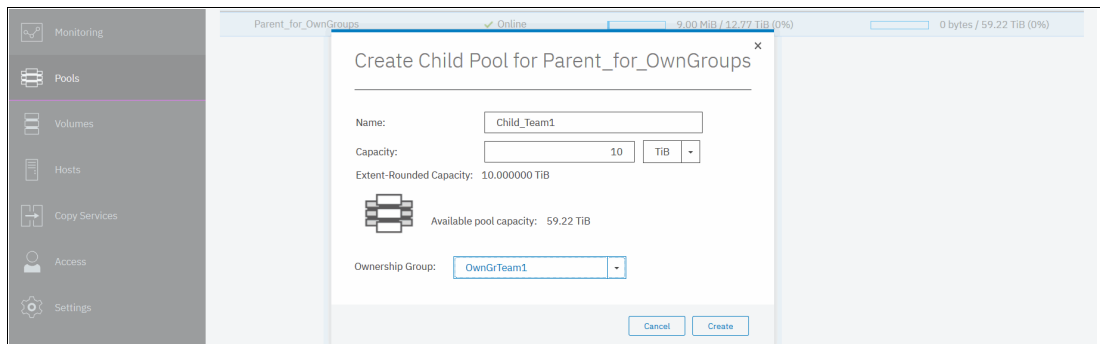


Figure 11-7 Creating a child pool and assigning it to an ownership group

Multiple child pools that are created from the same or different parent pools can be assigned to a single ownership group.

After a child pool is created and assigned, the ownership group management window, which you open by selecting **Access** → **Ownership Groups**, changes to show the assigned and available resources, as shown in Figure 11-8.

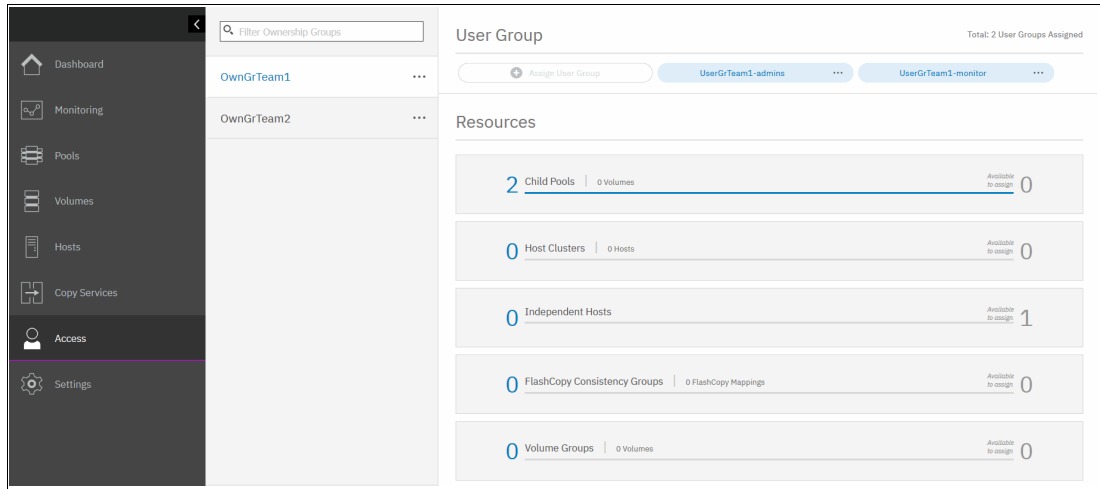


Figure 11-8 Ownership group management window

Any volumes that are created on a child pool that is assigned to an ownership group inherits ownership from the child pool.

After a child pool and user group are assigned to an ownership group, ownership group administrators can log in with their credentials and start creating volumes, host and host clusters, or FlashCopy mappings. For more information about creating those objects, see Chapter 6, “Volumes” on page 299, Chapter 7, “Hosts” on page 405, and Chapter 10, “Advanced Copy Services” on page 553.

Although an ownership group administrator can create objects only within the resources that are assigned to them, the system administrator can create, monitor, and assign objects for any ownership group.

11.2.4 Listing ownership group resources

By default, the **Ownership Group** attribute is not enabled in the GUI windows that list volumes and other objects that can be owned. For convenience, the system administrator can enable this attribute, as shown in Figure 11-9.

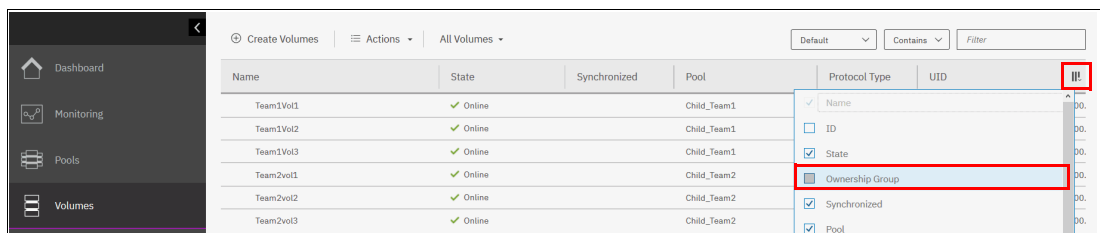


Figure 11-9 Enabling the ownership group attribute display

For example, the volume listing for a system administrator looks like Figure 11-10.

Name	State	Ownership Group	Synchronized	Pool	Protocol Type	UID
Team1Vol1	Online	OwnGrTeam1		Child_Team1		600507681
Team1Vol2	Online	OwnGrTeam1		Child_Team1		600507681
Team1Vol3	Online	OwnGrTeam1		Child_Team1		600507681
Team2vol1	Online	OwnGrTeam2		Child_Team2		600507681
Team2vol2	Online	OwnGrTeam2		Child_Team2		600507681
Team2vol3	Online	OwnGrTeam2		Child_Team2		600507681
NonOwnedVol1	Online			Parent_for_OwnGroups		600507681
NonOwnedVol2	Online			Parent_for_OwnGroups		600507681

Figure 11-10 Listing volumes for all ownership groups

The global system administrator can see and manage the resources of all ownership groups and resources that are not assigned to any groups.

When the ownership group user logs in, they can see and manage only resources that are assigned to their group. Figure 11-11 shows the initial login window for an ownership group user with the Administrator role.

This user does not see a dashboard with global system performance and capacity parameters, but instead can see only tiles for their existing ownership group resources. Out of eight volumes that are configured on a system and shown in Figure 11-10, they can see and manage only three volumes that belong to the group.

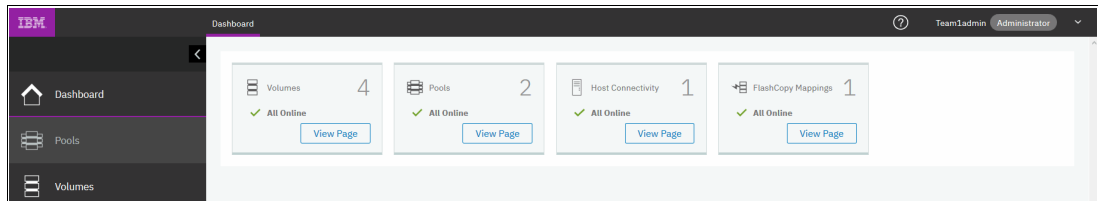


Figure 11-11 Ownership group administrator view

The ownership group user can use the GUI to browse, create, and delete (depending on their user role) resources that are assigned to their group. To see information about the global resources (for example, list managed disks (MDisks) or arrays on the pool), they must use the CLI. Ownership group users cannot manage global resources, but can only view them.

11.2.5 Actions on ownership groups

The global system administrator can rename or remove an ownership group. To do so, select **Access** → **Ownership Groups**, click the “...” icon next to the group name, and select the required task, as shown in Figure 11-12.

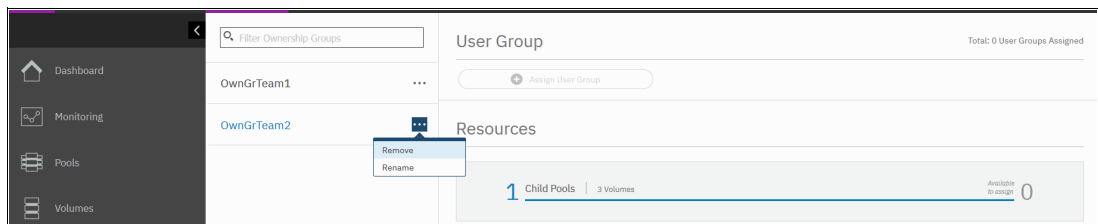


Figure 11-12 Renaming or removing an ownership group

When an ownership group is removed by using the GUI, all ownership assignment information for all the objects of the ownership group is removed, but the objects remain configured. Only the system administrator can manage those resources afterward.

11.3 Migrating existing objects to ownership groups

If you want to use ownership groups for objects that exist on your system, you must reconfigure certain resources if you want to configure ownership groups.

If child pools are on the system, you can define an ownership group to the child pool or child pools. Before you define an ownership group to existing child pools, determine other related objects that you want to migrate. Any volumes that are currently in the child pool inherit the ownership group that is defined for the child pool.

If no child pools are on the system, you must create child pools and move any volumes to those child pools before you can assign them to ownership groups. If volumes currently are in a parent pool, volume mirroring can be used to create copies of the volume within the child pool. Alternatively, volume migration can be used to relocate a volume from a parent pool to a child pool within that parent pool without requiring copying.

To non-disruptively migrate a volume to become an ownership group object, complete the following steps:

1. Create a child pool. Do not assign it to an ownership group yet.
2. Migrate volumes that must be assigned to an ownership group to that child pool by using the volume migration or volume mirroring function.

Figure 11-13 shows the NonOwnedVol1 volume, which is in a parent pool and does not belong to any ownership group. This volume is mapped to a Small Computer System Interface (SCSI) host.

To start migration, right-click the volume and click **Add Volume Copy**. You can also use **Migrate to Another Pool**, but this method is suitable only if you are migrating from a pool with the same extent size (for example, from a parent pool to a child pool of the same pool), and provides less flexibility.

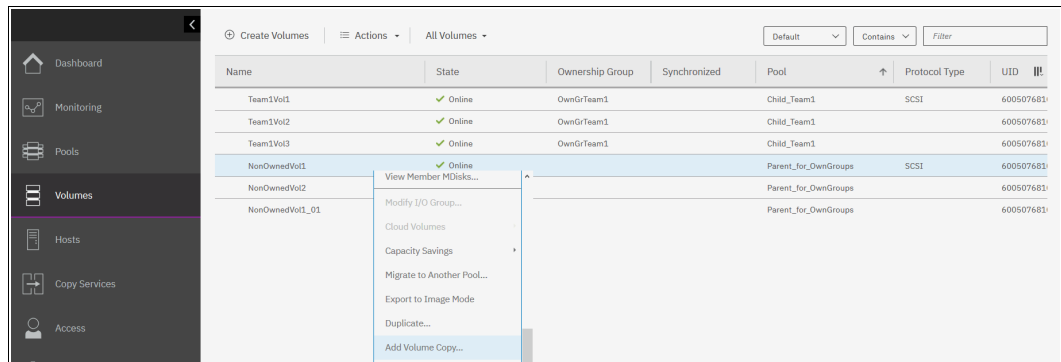


Figure 11-13 Adding a volume copy for migration

On the Volume Copy window, select the child pool that will be assigned to an ownership group, as shown in Figure 11-14.

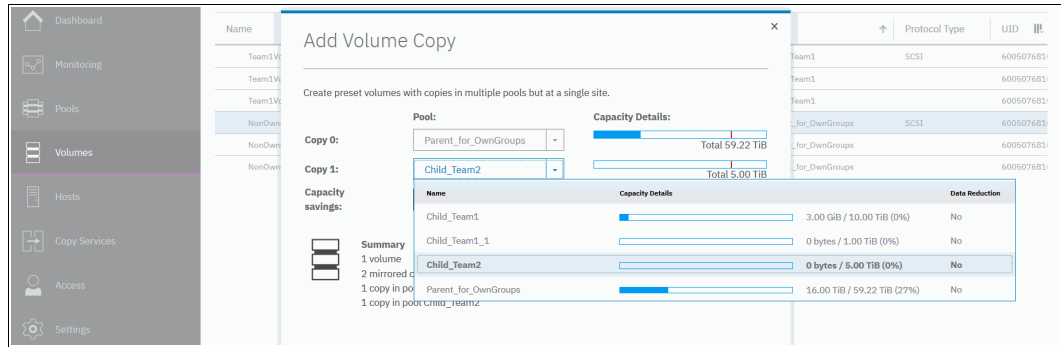


Figure 11-14 Migrating to a child pool

- Repeat step 2 on page 731 for all volumes that must belong to an ownership group, and then remove the source copies.
- Create an ownership group as described in 11.2.1, “Creating an ownership group” on page 726. Assign a user group to it, as described in 11.2.2, “Assigning users to an ownership group” on page 726.
- As shown in Figure 11-15, in **Access** → **Ownership Groups**, select the wanted ownership group and click **Assign Child Pool**.

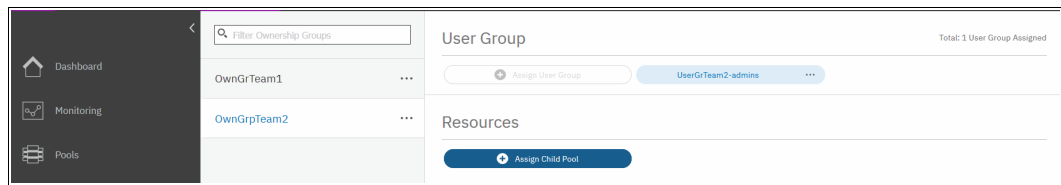


Figure 11-15 Assigning a child pool to an ownership group

Select a child pool to assign, as shown in Figure 11-16.

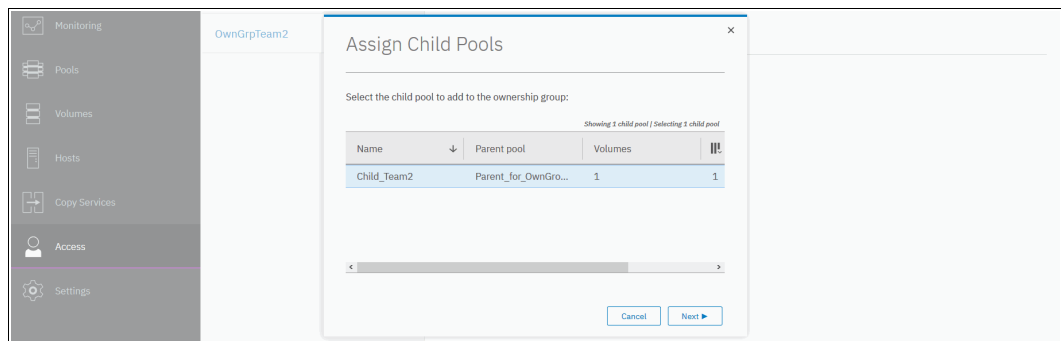


Figure 11-16 Selecting a child pool to assign

After you click **Next**, the system notifies you that there are more resources that will inherit ownership from a volume, and because the volume is mapped to a host, the host will become an ownership group object, as shown in Figure 11-17 on page 733.

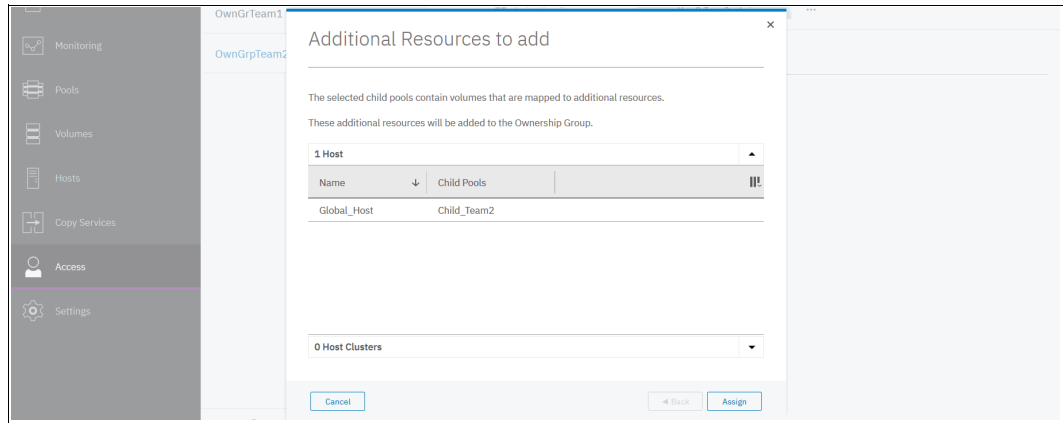


Figure 11-17 Additional Resources to add

- As shown in Figure 11-18, a volume and a host both belong to an ownership group. As a host and a volume are in a group, host mapping inherits ownership and becomes a part of an ownership group too.

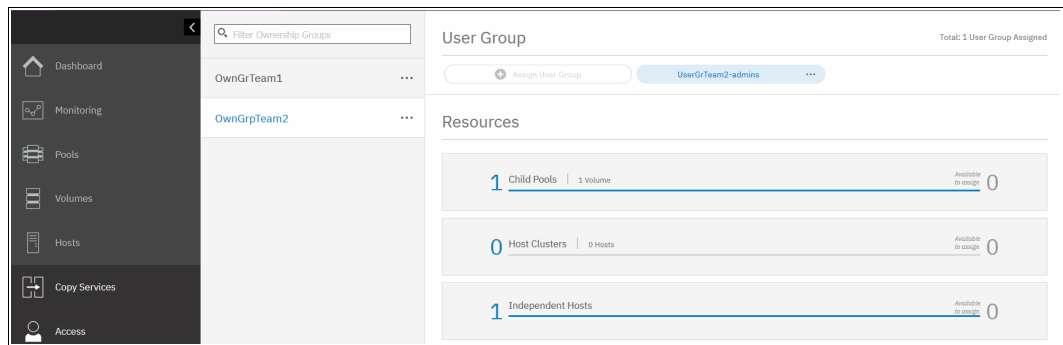


Figure 11-18 Resources of an ownership group

Now, a child pool is assigned to an ownership group. If you must migrate more volumes to the child pool later, the same approach can be used. However, during migration one volume copy is in an owned child pool, and the original copy remains in an unowned parent pool. Such a condition causes *inconsistent ownership*, as shown in Figure 11-19.

Name	State	Ownership Group	Synchronized	Pool	Protocol Type	UUID
Team1Vol1	Online	OwnGrTeam1		Child_Team1	SCSI	600507681
Team1Vol2	Online	OwnGrTeam1		Child_Team1	SCSI	600507681
Team1Vol3	Online	OwnGrTeam1		Child_Team1	SCSI	600507681
NonOwnedVol1	Online	OwnGrTeam2		Child_Team2	SCSI	600507681
NonOwnedVol2	Online	Inconsistent		Parent_for_OwnGroups		600507681
Copy 0*	Online		Yes	Parent_for_OwnGroups		600507681
Copy 1	Online		No	Child_Team2		600507681

Figure 11-19 Example of inconsistent volume ownership

Until the inconsistent volume ownership is resolved, the volume does not belong to an ownership group and cannot be seen or managed by an ownership group administrator. To resolve it, delete one of the copies after both are synchronized.



Encryption

Encryption protects against the potential exposure of sensitive user data that is stored on discarded, lost, or stolen storage devices. For storage devices that use IBM Spectrum Virtualize, it supports optional encryption of data-at-rest.

This chapter includes the following topics:

- ▶ 12.2, “Planning for encryption” on page 737
- ▶ 12.3, “Defining encryption of data-at-rest” on page 737
- ▶ 12.4, “Activating encryption” on page 742
- ▶ 12.5, “Enabling encryption” on page 752
- ▶ 12.6, “Configuring more providers” on page 774
- ▶ 12.7, “Migrating between providers” on page 777
- ▶ 12.8, “Recovering from a provider loss” on page 780
- ▶ 12.9, “Using encryption” on page 781
- ▶ 12.10, “Rekeying an encryption-enabled system” on page 789
- ▶ 12.11, “Disabling encryption” on page 792

12.1 General types of encryption across IBM Spectrum Virtualize

Within IBM Spectrum Virtualize SAN Volume Controller (SVC) and IBM FlashSystem, there are essentially three different types of encryption:

- ▶ Externally virtualized storage
- ▶ Serial-attached SCSI internal storage
- ▶ Non-Volatile Memory Express internal storage

12.1.1 Externally virtualized storage

Data is decrypted and encrypted as read/write I/Os are issued to the external storage. You can have an encryption key per storage pool or per child storage pool. Migrating a volume between pools (by using volume mirroring) can be used as a technique for encrypting and decrypting the data.

The key per pool (and allowing different keys for child pools) supports some part of the multi-tenant use case (if you delete a pool, you delete the key and cryptoerase the data), but all the keys are wrapped and protected by a single master key that is obtained either from a USB stick or an external key server.

As a special case, it is possible to turn off encryption for individual MDisks within the storage pool, which means that if an external storage controller supports encryption that you can choose to allow it to encrypt the data instead.

12.1.2 Serial-attached SCSI internal storage

Data is decrypted and encrypted by the serial-attached Small Computer System Interface (SCSI) (SAS) controller. There is an encryption key per redundant array of independent disks (RAID) array. Normally, all arrays in a storage pool are encrypted to form an encrypted storage pool. You can create child storage pools, but there is still only one key per RAID array. Multi-tenancy is possible only if you have more than one array and storage pool, which usually is not practical.

You can migrate volumes from a non-encrypted storage pool to an encrypted storage pool, or you can add an encrypted array to a storage pool and then delete the unencrypted array (which migrates all the data automatically) as a way of encrypting data.

12.1.3 Non-Volatile Memory Express internal storage

Data is decrypted and encrypted by the Non-Volatile Memory Express (NVMe) drives. Each drive has a media encryption key that is wrapped and protected by an encryption key per RAID array. So, it has the same properties as SAS internal storage.

A storage pool can include a mixture of two or all three types of storage. In this case, the SAS and NVMe internal storage use a key per RAID array for encryption, and the externally virtualized storage uses the pool level key. Because it is almost impossible to control exactly what storage is used for each volume, from a security viewpoint you effectively have a single key for the whole pool, and a cryptographic erase is possible only by deleting the entire storage pool and arrays.

12.2 Planning for encryption

Data-at-rest encryption is a powerful tool that can help organizations protect the confidentiality of sensitive information. However, encryption, like any other tool, must be used correctly to fulfill its purpose.

Multiple drivers exist for an organization to implement data-at-rest encryption. These drivers can be internal, such as protection of confidential company data and ease of storage sanitization, or external, such as compliance with legal requirements or contractual obligations.

Therefore, before configuring encryption on storage, the organization defines its needs and if it decides that data-at-rest encryption is required, it includes it in the security policy. Without defining the purpose of the particular implementation of data-at-rest encryption, it is difficult or impossible to choose the best approach to implement encryption and verify whether the implementation meets the set of goals.

The following items are worth considering during the design of a solution that includes data-at-rest encryption:

- ▶ Legal requirements
- ▶ Contractual obligations
- ▶ Organization's security policy
- ▶ Attack vectors
- ▶ Expected resources of an attacker
- ▶ Encryption key management
- ▶ Physical security

Multiple regulations mandate data-at-rest encryption, from processing of Sensitive Personal Information to the guidelines of the Payment Card Industry. If any regulatory or contractual obligations govern the data that is held on the storage system, they often provide a wide and detailed range of requirements and characteristics that must be realized by that system. Apart from mandating data-at-rest encryption, these documents might contain requirements concerning encryption key management.

Another document that should be consulted when planning data-at-rest encryption is the organization's security policy.

The outcome of a data-at-rest encryption planning session answers the following questions:

1. What are the goals that the organization wants to realize by using data-at-rest encryption?
2. How will data-at-rest encryption be implemented?
3. How can it be demonstrated that the proposed solution realizes the set of goals?

12.3 Defining encryption of data-at-rest

Encryption is the process of encoding data so that only authorized parties can read it. Secret keys are used to encode the data according to well-known algorithms.

Encryption of data-at-rest as implemented in IBM Spectrum Virtualize is defined by the following characteristics:

- ▶ *Data-at-rest* means that the data is encrypted on the end device (drives).
- ▶ The algorithm that is used is the Advanced Encryption Standard (AES) US government standard from 2001.

- ▶ Encryption of data-at-rest complies with the Federal Information Processing Standard 140-2 (FIPS-140-2) standard.
- ▶ AES 256 is used for master access keys.
- ▶ XTS-AES 256 encryption is a FIPS 140-2 compliant algorithm.
- ▶ XTS-AES-256 is used for data encryption.
- ▶ The algorithm is public. The only secrets are the keys.
- ▶ A symmetric key algorithm is used. The same key is used to encrypt and decrypt data.

The encryption of system data and metadata is not required, so they are not encrypted.

12.3.1 Encryption methods

There are two types of encryption on devices running IBM Spectrum Virtualize: hardware encryption and software encryption. Both methods of encryption protect against the potential exposure of sensitive user data that is stored on discarded, lost, or stolen media. Both can also facilitate the warranty return or disposal of hardware.

Which method that is used for encryption is chosen automatically by the system based on the placement of the data:

- ▶ Hardware encryption: Data is encrypted by using SAS hardware or self-encrypting drives, for example, if IBM FlashCore Module (FCM) drives are presented in the system, hardware-based data compression and self-encryption is used. Hardware encryption is used only for internal storage (drives).
- ▶ Software encryption: Data is encrypted by using the node's CPU (the encryption code uses the AES-NI CPU instruction set). Used only for external storage.

Note: Software encryption is available in IBM Spectrum Virtualize V7.6 and later.

Both methods of encryption use the same encryption algorithm, key management infrastructure, and license.

Note: The design for encryption is based on the concept that a system is encrypted or not encrypted. Encryption implementation is intended to encourage solutions that contain only encrypted volumes or only unencrypted volumes. For example, after encryption is enabled on the system, all new objects (for example, pools) are by default created as encrypted.

12.3.2 Encrypted data

IBM Spectrum Virtualize performs data-at-rest encryption, which is the process of encrypting data that is stored on the end devices, such as physical drives.

Data is encrypted or decrypted when it is written to or read from internal drives (hardware encryption) or external storage systems (software encryption).

So, data is encrypted when transferred across the storage area network (SAN) only between IBM Spectrum Virtualize systems and external storage. Data in transit is *not* encrypted when transferred on SAN interfaces under the following circumstances:

- ▶ Server-to-storage data transfer
- ▶ Remote Copy (RC) (for example, Global Mirror or Metro Mirror (MM))
- ▶ Intracluster (node-to-node) communication

Note: Only data-at-rest is encrypted. Host to storage communication and data that is sent over links that are used for Remote Mirroring are not encrypted.

Figure 12-1 shows an encryption example. Encrypted disks and encrypted data paths are marked in blue. Unencrypted disks and data paths are marked in red. The server sends unencrypted data to an SVC 2145-DH8 system, which stores hardware-encrypted data on internal disks. The data is mirrored to a remote Storwize V7000 Gen1 system by using RC. The data flowing through the RC link is not encrypted. Because the Storwize V7000 Gen1 (2076-324) system cannot perform any encryption activities, data on the Storwize V7000 Gen1 is not encrypted.

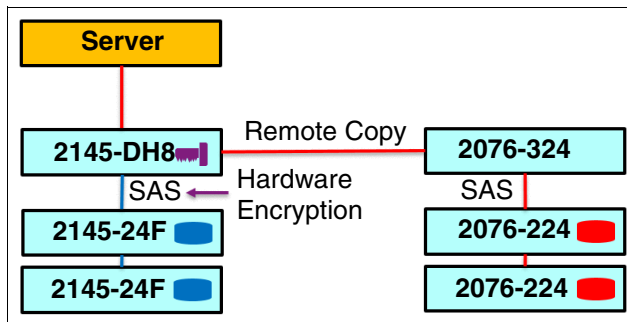


Figure 12-1 Encryption on a single site

To enable encryption of both data copies, the Storwize V7000 Gen1 system must be replaced by an encryption capable (with optional encryption enabled) IBM Spectrum Virtualize system, as shown in Figure 12-2. After the replacement, both copies of data are encrypted, but the RC communication between both sites remains unencrypted.

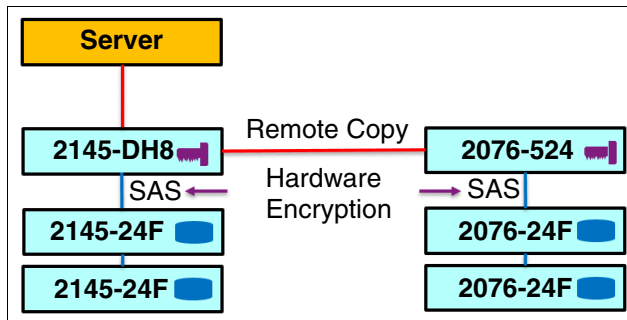


Figure 12-2 Encryption on both sites

Figure 12-3 shows an example configuration that uses software and hardware encryption. Software encryption is used to encrypt an external virtualized storage system (2076-324 in Figure 12-3). Hardware encryption is used for internal, SAS-attached disk drives.

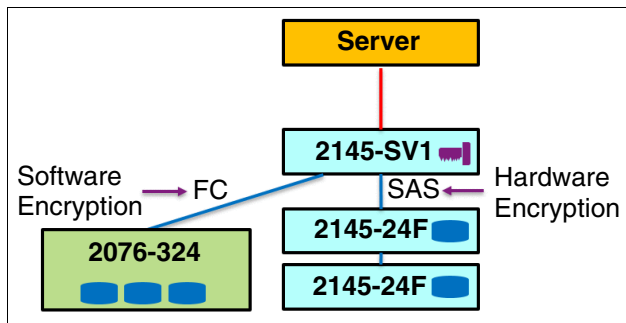


Figure 12-3 Example of software encryption and hardware encryption

The placement of hardware encryption and software encryption in the IBM Spectrum Virtualize code stack are shown in Figure 12-4. As compression is performed before encryption, it is possible to get benefits of compression for the encrypted data.

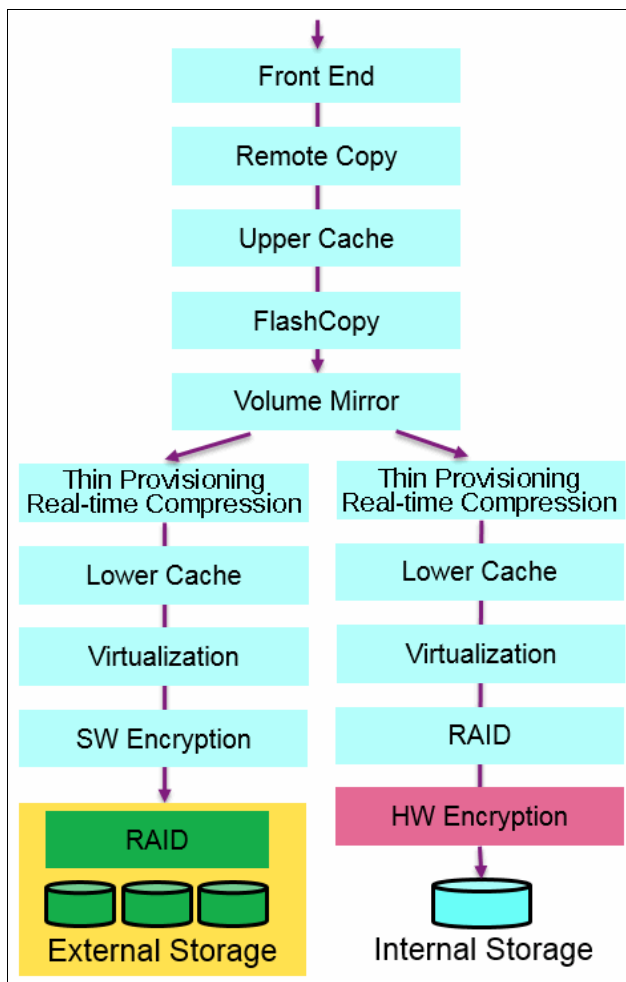


Figure 12-4 Encryption placement in the IBM Spectrum Virtualize Software stack (with IBM Real-time Compression Appliance)

Each volume copy can use different encryption methods (hardware and software). It also may have volume copies with different encryption status (encrypted versus unencrypted). The encryption method depends only on the pool that is used for the specific copy. You can migrate data between different encryption methods by using volume migration or volume mirroring.

12.3.3 Encryption keys

Hardware and software encryption use the same encryption key infrastructure. The only difference is the object that is encrypted by using the keys. The following objects can be encrypted:

- ▶ Pools (software encryption)
- ▶ Child pools (software encryption)
- ▶ Arrays (hardware encryption)

Consider the following points regarding encryption keys:

- ▶ Keys are unique for each object, and they are created when the object is created.
- ▶ Two types of keys are defined in the system:
 - Master access key:
 - The master access key is created when encryption is enabled.
 - The master access key can be stored on USB flash drives, key servers, or both. One master access key is created for each enabled encryption key provider.
 - It can be copied or backed up as necessary.
 - It is *not* permanently stored anywhere in the system.
 - It is required at boot time to unlock access to encrypted data.
 - Data encryption keys (one for each encrypted object):
 - Data encryption keys are used to encrypt data. When an encrypted object (such as an array, a pool, or a child pool) is created, a new data encryption key is generated for this object.
 - Managed disks (MDisks) that are not self-encrypting are automatically encrypted by using the data encryption key of the pool or child pool to which they belong.
 - MDisks that are self-encrypting are not reencrypted by using the data encryption key of the pool or child pool they belong to by default. You can override this default by manually configuring the MDisk as not self-encrypting.
 - Data encryption keys are stored in secure memory.
 - During cluster internal communication, data encryption keys are encrypted with the master access key.
 - Data encryption keys cannot be viewed.
 - Data encryption keys cannot be changed.
 - When an encrypted object is deleted, its data encryption key is discarded (*secure erase*).

Important: If all master access key copies are lost and the system must cold-restart, all encrypted data is gone. No method exists, even for IBM, to decrypt the data without the keys. If encryption is enabled and the system cannot access the master access key, all SAS hardware is offline, including unencrypted arrays.

Note: A self-encrypting MDisk is an MDisk from an encrypted volume in an external storage system.

12.3.4 Encryption licenses

Encryption is a licensed feature that uses key-based licensing. A license must be present for each control enclosure in the system before you can enable encryption.

If you add a control enclosure to a system that has encryption that is enabled, the control enclosure must also be licensed.

No trial license for encryption exists because when the trial runs out, the access to the data is lost. Therefore, you must purchase an encryption license before you activate encryption. Licenses are generated by IBM Data Storage Feature Activation (DSFA) based on the serial number (S/N) and the machine type and model (MTM) of the control enclosure.

You can activate an encryption license during the initial system setup (on the Encryption window of the initial setup wizard) or later on in the running environment.

To purchase an encryption license, contact your IBM marketing representative or IBM Business Partner.

12.4 Activating encryption

Encryption is enabled at a system level, and all the following prerequisites must be met to use encryption:

- ▶ You must purchase an encryption license before you activate the function.
If you did not purchase a license, contact an IBM marketing representative or IBM Business Partner to purchase an encryption license.
- ▶ At least three USB flash drives are required if you plan to *not* use a key management server. They are available as a feature code from IBM.
- ▶ You must activate the license that you purchased.
- ▶ Encryption must be enabled.

Activation of the license can be performed in one of two ways:

- ▶ **Automatic activation:** Used when you have the authorization code, and the workstation that is being used to activate the license has access to external network. In this case, you must enter only the authorization code. The license key is automatically obtained from the internet and activated in the IBM Spectrum Virtualize system.
- ▶ **Manual activation:** If you cannot activate the license automatically because any of the requirements are not met, you can follow the instructions that are provided in the GUI to obtain the license key from the web and activate it in the IBM Spectrum Virtualize system.

Both methods are available during the initial system setup and when the system is in use.

12.4.1 Obtaining an encryption license

You must purchase an encryption license before you activate encryption. If you did not purchase a license, contact an IBM marketing representative or IBM Business Partner to purchase an encryption license.

When you purchase a license, you receive a function authorization document with an authorization code that is printed on it. With this code, you may proceed with the automatic activation process.

If the automatic activation process fails or if you prefer to use the manual activation process, see [IBM Data Storage Feature Activation](#) to retrieve your license keys.

Ensure that you have the following information:

- ▶ Machine type (MT)
- ▶ Serial number (S/N)
- ▶ Machine signature
- ▶ Authorization code

For more information about how to retrieve the machine signature of a control enclosure, see 12.4.5, “Activating the license manually” on page 750.

12.4.2 Starting the activation process during the initial system setup

One of the steps in the initial setup enables the encryption license activation. The system asks “Was the encryption feature purchased for this system?”. To activate encryption at this stage, complete the following steps:

1. Select **Yes**, as shown in Figure 12-5.

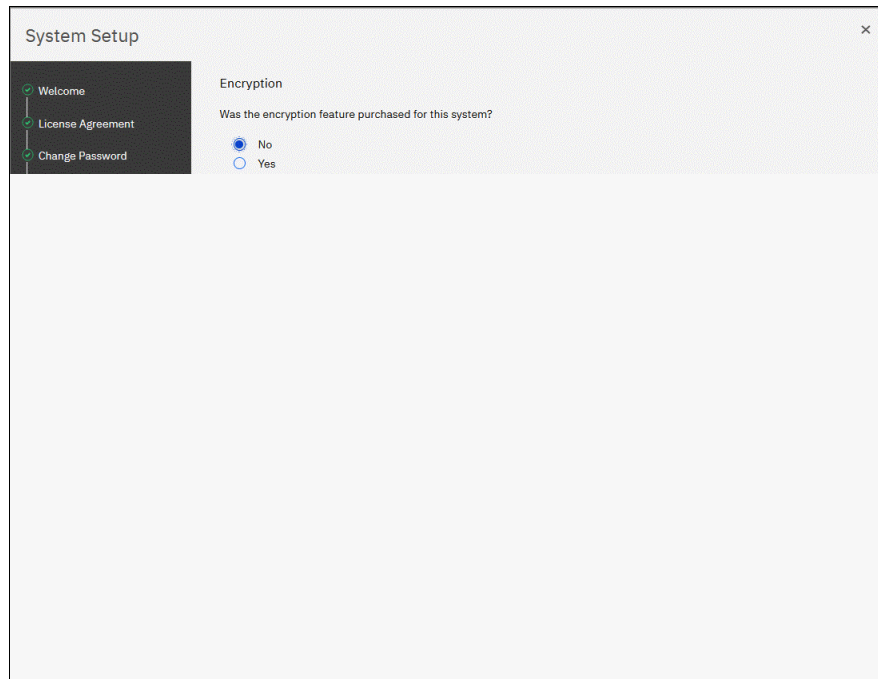


Figure 12-5 Encryption activation during the initial system setup

The Encryption window displays information about your storage system, as shown in Figure 12-6.

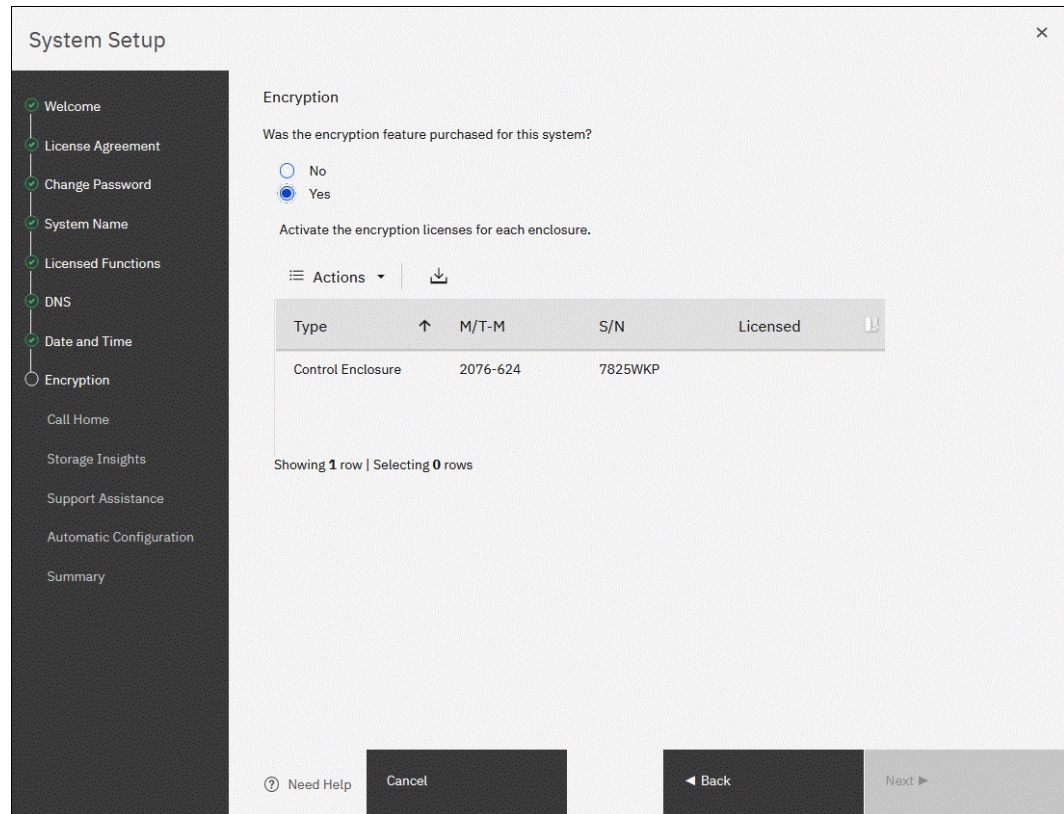


Figure 12-6 Information storage system during the initial system setup

2. Right-click the control enclosure to open a menu with two license activation options (**Activate License Automatically** and **Activate License Manually**), as shown in Figure 12-7. Use either option to activate encryption. For more information about how to complete the automatic activation process, see 12.4.4, “Activating the license automatically” on page 747. For more information about how to complete a manual activation process, see 12.4.5, “Activating the license manually” on page 750.

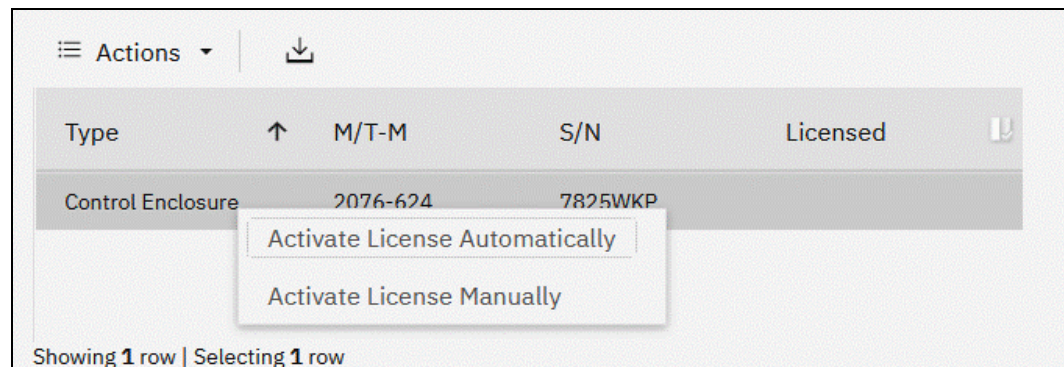


Figure 12-7 Selecting the license activation method

3. After either activation process is complete, you can see a green check mark in the column that is labeled **Licensed** next to a control enclosure for which the license was enabled. You can proceed with the initial system setup by clicking **Next**, as shown in Figure 12-8.

Note: Every enclosure needs an active encryption license before you can enable encryption on the system. Attempting to add a non-licensed enclosure to an encryption-enabled system fails.

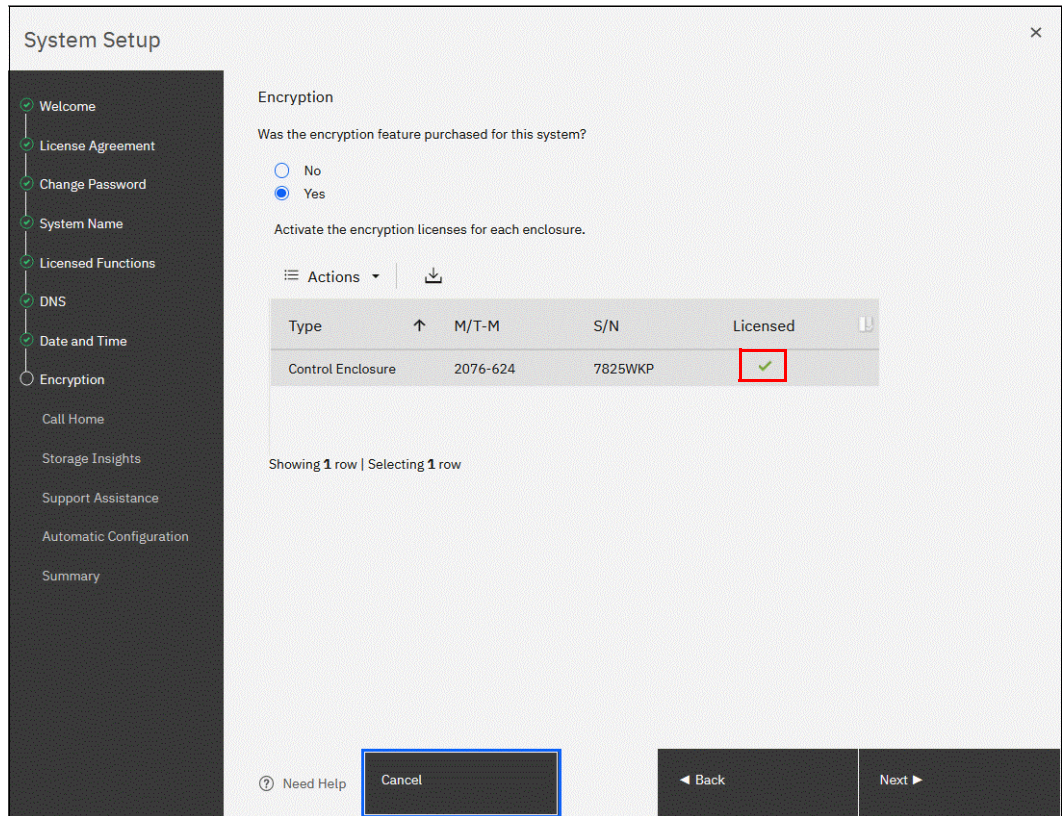


Figure 12-8 Successful encryption license activation during the initial system setup

12.4.3 Starting the activation process on a running system

To activate encryption on a running system, complete the following steps:

1. Select **Settings** → **System** → **Licensed Functions**.
2. Click **Encryption Licenses**, as shown in Figure 12-9.

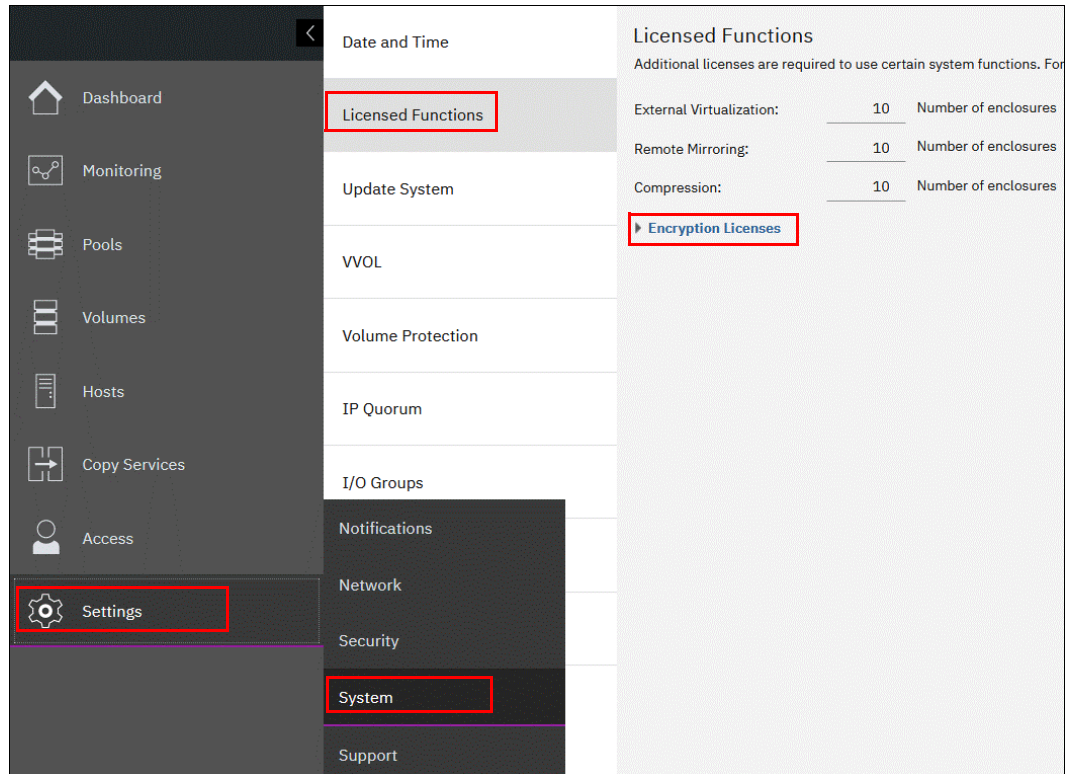


Figure 12-9 Expanding the Encryption Licenses section on the Licensed Functions window

3. The Encryption Licenses window displays information about your control enclosures. Right-click the enclosure on which you want to install an encryption license. This action opens a menu with two license activation options (**Activate License Automatically** and **Activate License Manually**), as shown in Figure 12-10. Use either option to activate encryption. For more information about how to complete an automatic activation process, see 12.4.4, “Activating the license automatically” on page 747. For more information about how to complete a manual activation process, see 12.4.5, “Activating the license manually” on page 750.

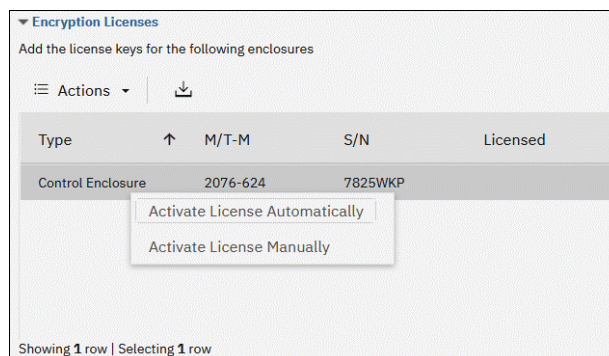
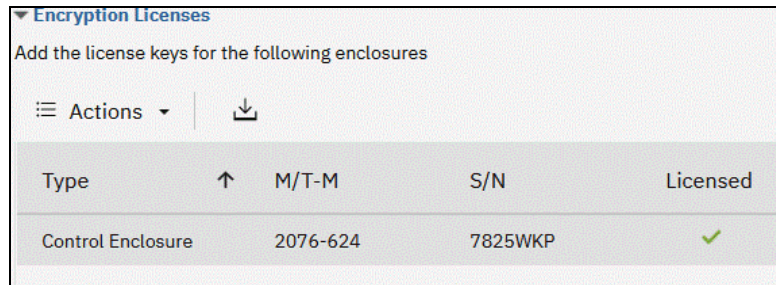


Figure 12-10 Selecting the Control Enclosure on which you want to enable the encryption

After either activation process is complete, you can see a green check mark in the column that is labeled **Licensed** for the control enclosure, as shown in Figure 12-11.



The screenshot shows a table titled "Encryption Licenses" with the instruction "Add the license keys for the following enclosures". The table has columns for "Type", "M/T-M", "S/N", and "Licensed". A row for "Control Enclosure" shows the M/T-M code "2076-624" and the S/N code "7825WKP", with a green checkmark in the "Licensed" column.

Type	M/T-M	S/N	Licensed
Control Enclosure	2076-624	7825WKP	✓

Figure 12-11 Successful encryption license activation on a running system

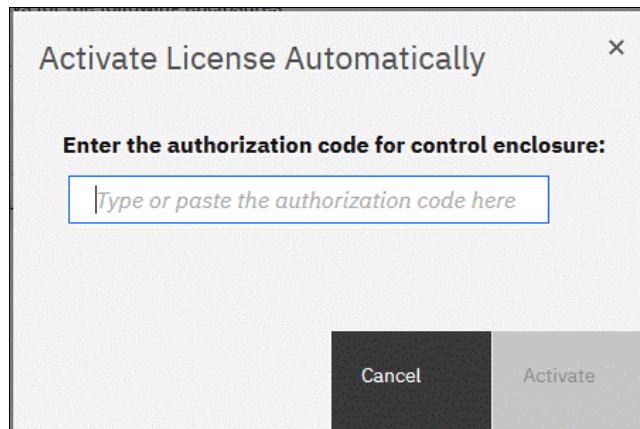
12.4.4 Activating the license automatically

The automatic license activation is the faster method to activate the encryption license for IBM Spectrum Virtualize. You need the authorization code and the workstation that is used to access the GUI that can access the external network.

Important: To perform this operation, the PC that was used to connect to the GUI and activate the license must connect to the internet.

To activate the encryption license for a control enclosure automatically, complete the following steps:

1. Click **Activate License Automatically** to open the Activate License Automatically window, as shown in Figure 12-12.



The dialog box is titled "Activate License Automatically" and contains the instruction "Enter the authorization code for control enclosure:". Below this is a text input field with the placeholder text "Type or paste the authorization code here". At the bottom of the dialog are two buttons: "Cancel" and "Activate".

Figure 12-12 Encryption license: Activate License Automatically window

2. Enter the authorization code that is specific to the control enclosure that you selected, as shown in Figure 12-13. Click **Activate**.

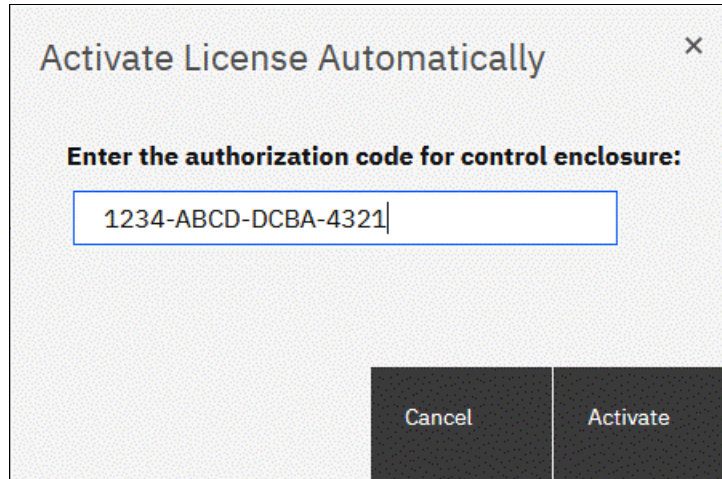


Figure 12-13 Entering an authorization code

The system connects to IBM to verify the authorization code and retrieve the license key. Figure 12-14 shows a window that is displayed during this connection. If everything works correctly, the procedure takes less than a minute.

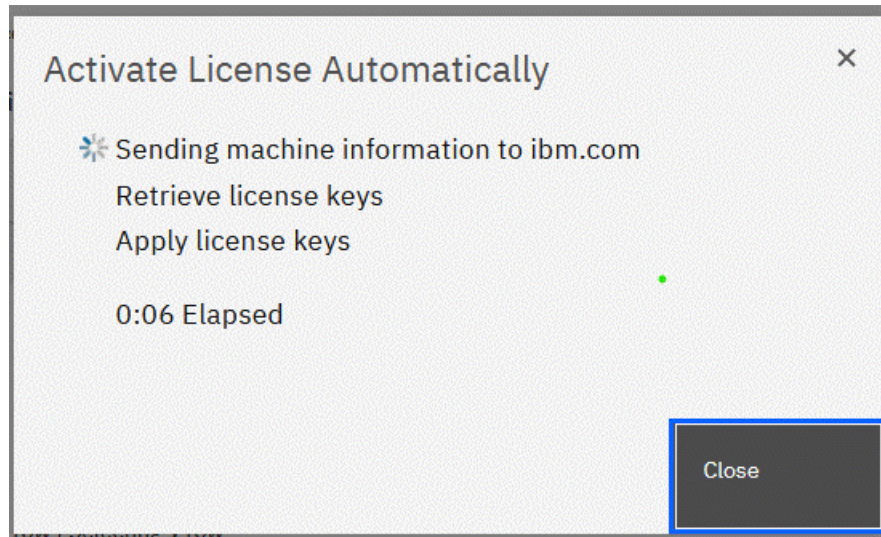


Figure 12-14 Activating encryption

After the license key is retrieved, it is automatically applied, as shown in Figure 12-15 on page 749.

▼ Encryption Licenses			
Add the license keys for the following enclosures			
☰ Actions ▾		↓	
Type	↑	M/T-M	S/N
Control Enclosure		2076-624	7825WKP ✓

Figure 12-15 Successful encryption license activation

Problems with automatic license activation

If connection problems occur with the automatic license activation procedure, the system times out after 3 minutes with an error, as shown in Figure 12-16.

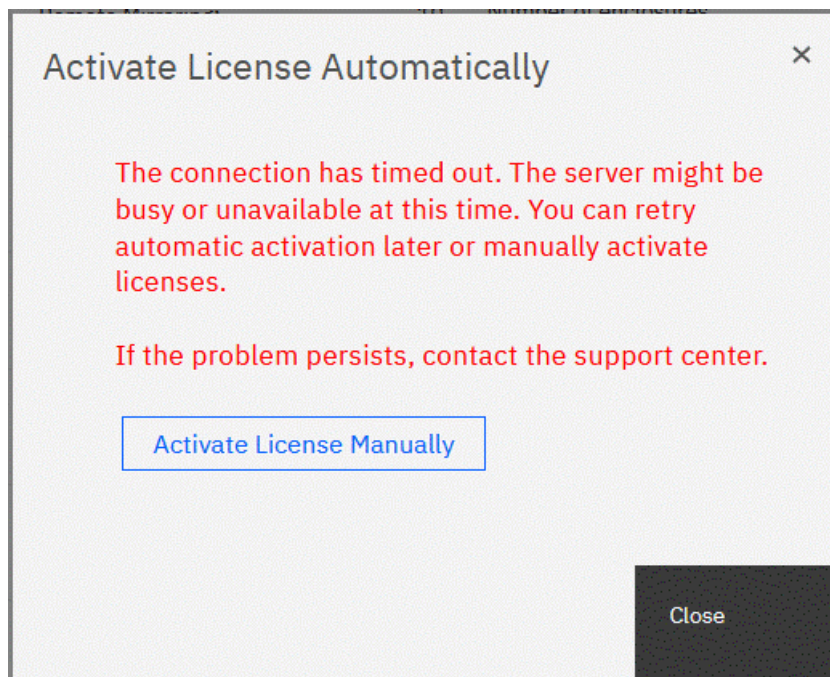


Figure 12-16 Authorization code failure

Check whether the PC that is used to connect to the IBM FlashSystem GUI and activate the license can access the internet. If you cannot complete the automatic activation procedure, use the manual activation procedure that is described in 12.4.5, “Activating the license manually” on page 750.

Although authorization codes and encryption license keys use the same format (four groups of four hexadecimal digits), you can use each of them only in the appropriate activation process. If you use a license key when the system expects an authorization code, the system displays the error message.

12.4.5 Activating the license manually

To manually activate the encryption license for a control enclosure, complete the following steps:

1. Click **Activate License Manually** to open the Manual Activation window, as shown in Figure 12-17.

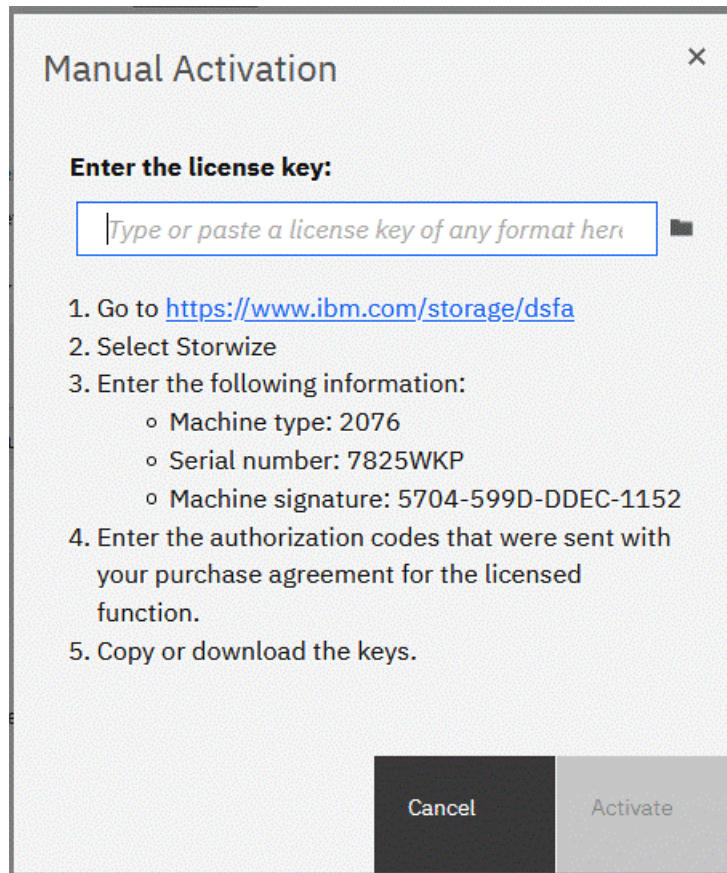


Figure 12-17 Manual encryption license activation window

2. If you have not done so, obtain the encryption license for the control enclosure. The information that is required to obtain the encryption license is displayed in the Manual Activation window. Use this data to follow the instructions in 12.4.1, “Obtaining an encryption license” on page 743.

3. You can enter the license key either by typing it, pasting it, or clicking the folder icon and uploading the license key file to the storage system that was downloaded from DSFA. In Figure 12-18, the sample key is entered. Click **Activate**.

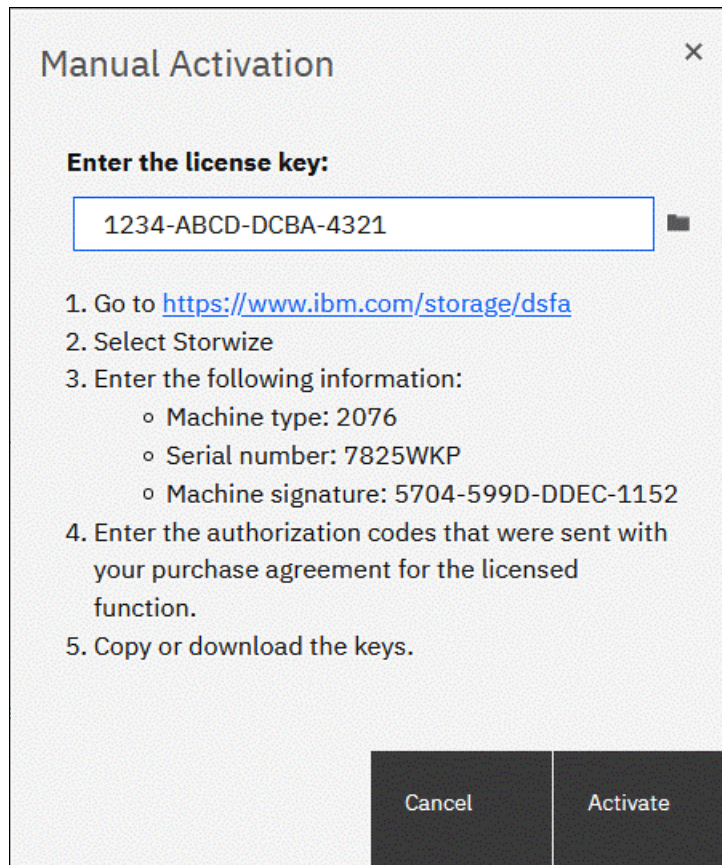


Figure 12-18 Entering an encryption license key

After the task completes successfully, the GUI shows that encryption is licensed for the specified control enclosure, as shown in Figure 12-19.

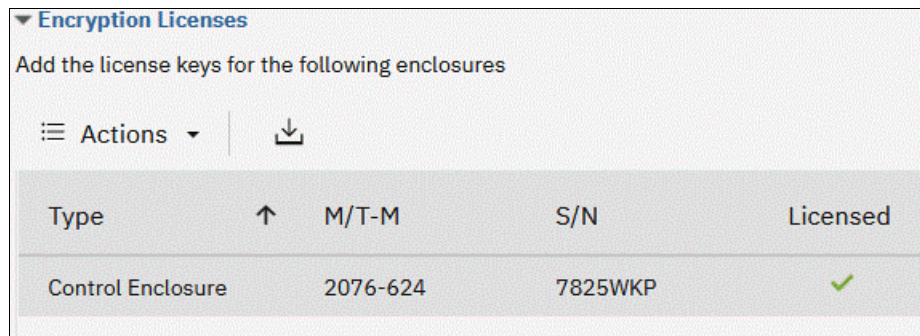


Figure 12-19 Successful encryption license activation

Problems with manual license activation

Although authorization codes and encryption license keys use the same format (four groups of four hexadecimal digits), you can use each of them only in the appropriate activation process. If you use an authorization code when the system expects a license key, the system displays an error message, as shown in Figure 12-20.

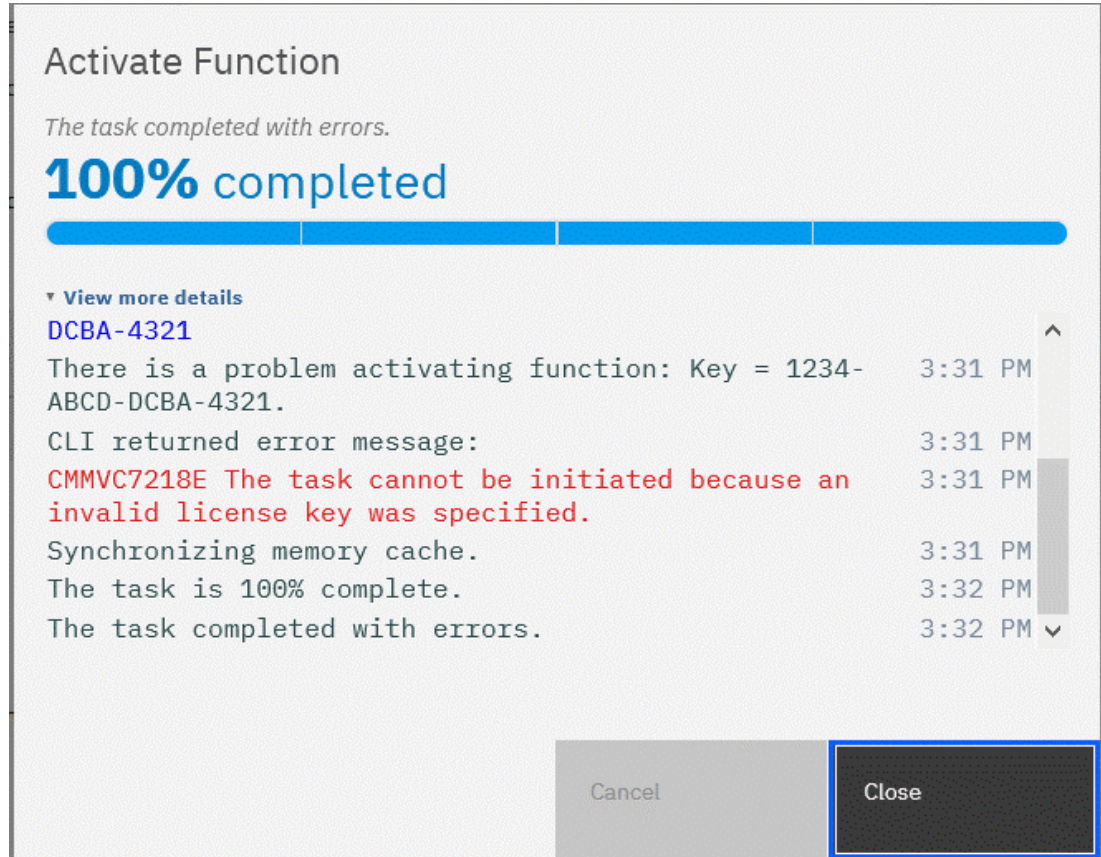


Figure 12-20 License key failure

12.5 Enabling encryption

This section describes the process to create and store system master access key copies, which are also known as *encryption keys*. These keys can be stored on any or both of two key providers: USB flash drives or a key server.

Two types of key servers are supported by IBM Spectrum Virtualize:

- ▶ IBM Security Key Lifecycle Manager, which was introduced in IBM Spectrum Virtualize V7.8.
- ▶ Gemalto SafeNet KeySecure, which was introduced in IBM Spectrum Virtualize V8.2.

For a list of supported key servers, see [Supported Key Servers - IBM Spectrum Virtualize](#).

IBM Spectrum Virtualize V8.1 introduced the ability to define up to four encryption key servers, which is a preferred configuration because it increases key provider availability. In this version, support for simultaneous use of both USB flash drives and key server was added.

Organizations that use encryption key management servers might consider parallel use of USB flash drives as a backup solution. During normal operation, such drives can be disconnected and stored in a secure location. However, during a catastrophic loss of encryption servers, the USB drives can still be used to unlock the encrypted storage.

The key server and USB flash drive characteristics that are described next might help you to choose the type of encryption key provider that you want to use.

Key servers can have the following characteristics:

- ▶ Physical access to the system is not required to perform a rekey operation.
- ▶ Support for businesses that have security requirements that preclude the use of USB ports.
- ▶ Possibility to use hardware security modules (HSMs) for encryption key generation.
- ▶ Ability to replicate keys between servers and perform automatic backups.
- ▶ Implementations follow an open standard (Key Management Interoperability Protocol (KMIP)) that aids in interoperability.
- ▶ Ability to audit operations that are related to key management.
- ▶ Ability to separately manage encryption keys and physical access to storage systems.

USB flash drives have the following characteristics:

- ▶ Physical access to the system might be required to process a rekey operation.
- ▶ No moving parts with almost no read or write operations to the USB flash drive.
- ▶ Inexpensive to maintain and use.
- ▶ Convenient and easy to have multiple identical USB flash drives available as backups.

Important: Maintaining confidentiality of the encrypted data hinges on the security of the encryption keys. Pay special attention to ensure secure creation, management, and storage of the encryption keys.

12.5.1 Starting the Enable Encryption wizard

After the license activation step completes on IBM FlashSystem Control Enclosures, you can now enable encryption. You can enable encryption after completing the initial system setup by using the GUI or command-line interface (CLI).

There are two ways in the GUI to start the Enable Encryption wizard:

- ▶ It can be started by clicking **Enable** next to **Enable Encryption** on the Suggested Tasks window, as shown in Figure 12-21.

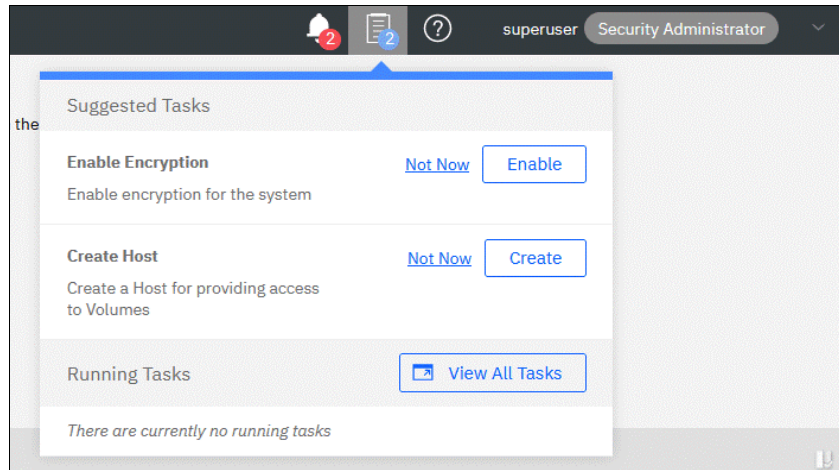


Figure 12-21 Enable Encryption from the Suggested Tasks window

- ▶ You can select **Settings** → **Security** → **Encryption**, and then click **Enable Encryption**, as shown in Figure 12-22.

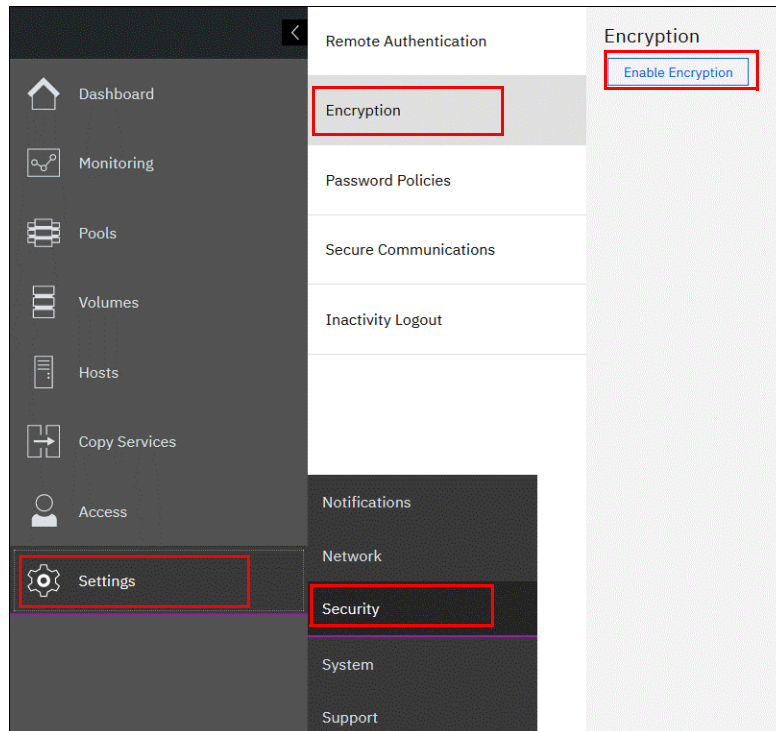


Figure 12-22 Enable Encryption from the Security window

The Enable Encryption wizard starts by prompting you to select the encryption key provider to use for storing the encryption keys, as shown in Figure 12-23 on page 755. You can enable either or both providers.

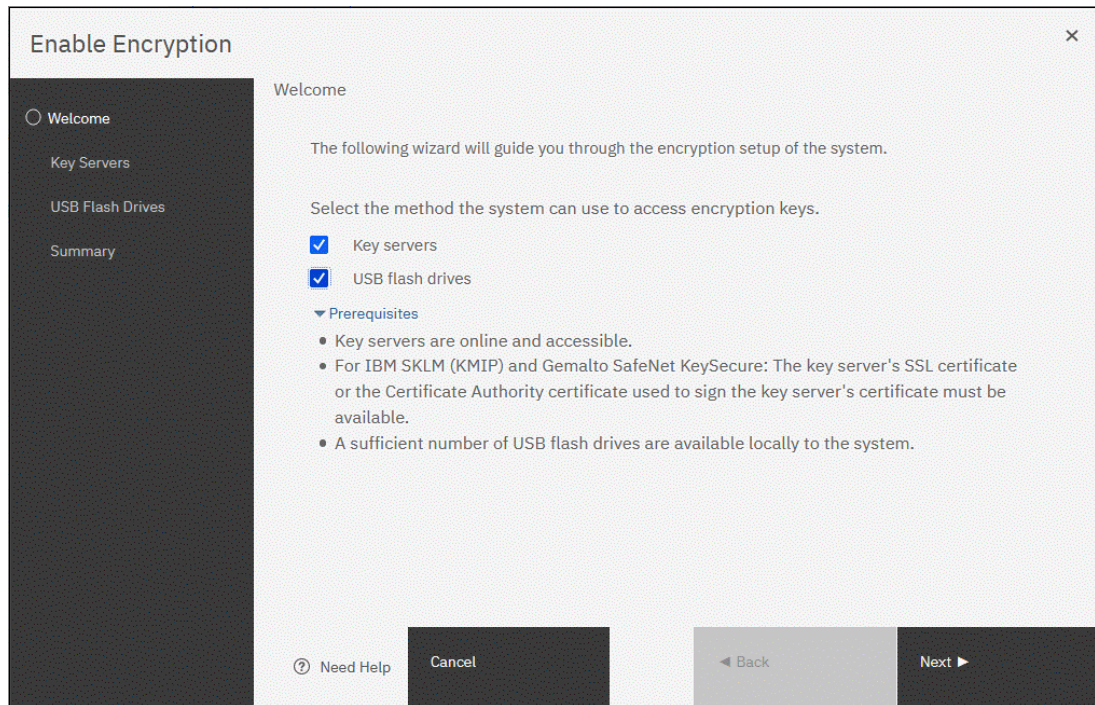


Figure 12-23 Enable Encryption wizard Welcome window

The next section presents a scenario in which both encryption key providers are enabled concurrently.

For more information about how to enable encryption by using only USB flash drives, see 12.5.2, “Enabling encryption by using USB flash drives” on page 755.

For more information about how to enable encryption by using key servers as the sole encryption key provider, see 12.5.3, “Enabling encryption by using key servers” on page 759.

12.5.2 Enabling encryption by using USB flash drives

Note: The system needs at least three USB flash drives before you can enable encryption by using this encryption key provider. IBM USB flash drives are preferred and can be obtained from IBM with the Feature Code Encryption USB flash drives (Four Pack). Other flash drives might also work. You can use any USB ports in any node of the cluster.

Using USB flash drives as the encryption key provider requires a minimum of three USB flash drives to store the generated encryption keys. Because the system attempts to write the encryption keys to any USB key that is inserted into a node port (N_Port), it is critical to maintain physical security of the system during this procedure.

While the system enables encryption, you are prompted to insert USB flash drives into the system. The system generates and copies the encryption keys to all available USB flash drives.

Ensure that each copy of the encryption key is valid before you write any user data to the system. The system validates any key material on a USB flash drive when it is inserted into the canister. If the key material is not valid, the system logs an error. If the USB flash drive is unusable or fails, the system does not display it as output. Figure 12-25 on page 757 shows an example where the system detected and validated three USB flash drives.

If your system is in a secure location with controlled access, one USB flash drive for each canister can remain inserted in the system. If a risk of unauthorized access exists, all USB flash drives with the master access keys must be removed from the system and stored in a secure place.

Securely store all copies of the encryption key. For example, any USB flash drives that are holding an encryption key copy that are not left plugged into the system can be locked in a safe. Similar precautions must be taken to protect any other copies of the encryption key that are stored on other media.

Notes: Generally, create at least one extra copy on another USB flash drive for storage in a secure location. You can also copy the encryption key from the USB drive and store the data on other media, which can provide extra resilience and mitigate risk that the USB drives used to store the encryption key come from a faulty batch.

Every encryption key copy must be stored securely to maintain confidentiality of the encrypted data.

A minimum of one USB flash drive with the correct master access key is required to unlock access to encrypted data after a system restart, such as a system-wide restart or power loss. No USB flash drive is required during a warm restart, such as node-exiting service mode or a single node restart. The data center power-on procedure must ensure that USB flash drives containing encryption keys are plugged into the storage system before it is powered on.

During power-on, insert the USB flash drives into the USB ports on two supported canisters to safeguard against failure of a node, node's USB port, or USB flash drive during the power-on procedure.

To enable encryption by using USB flash drives as the only encryption key provider, complete the following steps:

1. In the Enable Encryption wizard **Welcome** tab, select **USB flash drives** and click **Next**, as shown in Figure 12-24 on page 757.

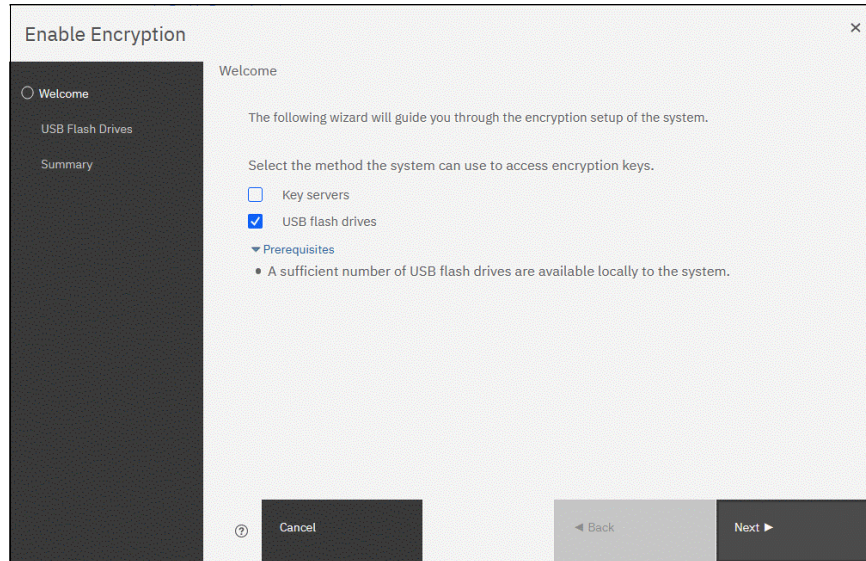


Figure 12-24 Selecting USB flash drives in the Enable Encryption wizard

2. If there are fewer than three USB flash drives that are inserted into the system, you are prompted to insert more drives. The system reports how many more drives must be inserted.

Note: The **Next** option remains disabled until at least three USB flash drives are inserted and the system detects them.

3. Insert the USB flash drives into the USB ports as requested.

After the minimum required number of drives is detected, the encryption keys are automatically copied onto the USB flash drives, as shown in Figure 12-25.

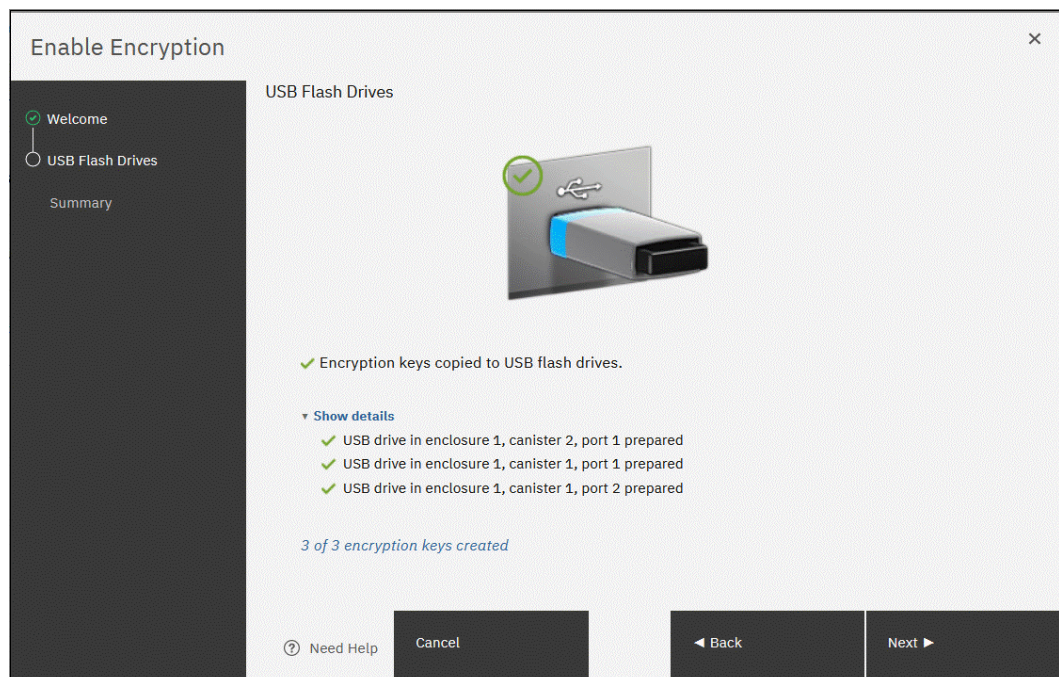


Figure 12-25 Writing the master access key to USB flash drives

You can keep adding USB flash drives or replacing the drives that are plugged in to create new copies. When done, click **Next**.

4. The number of keys that were created is shown in the **Summary** tab, as shown in Figure 12-26. Click **Finish** to finalize the encryption enablement.

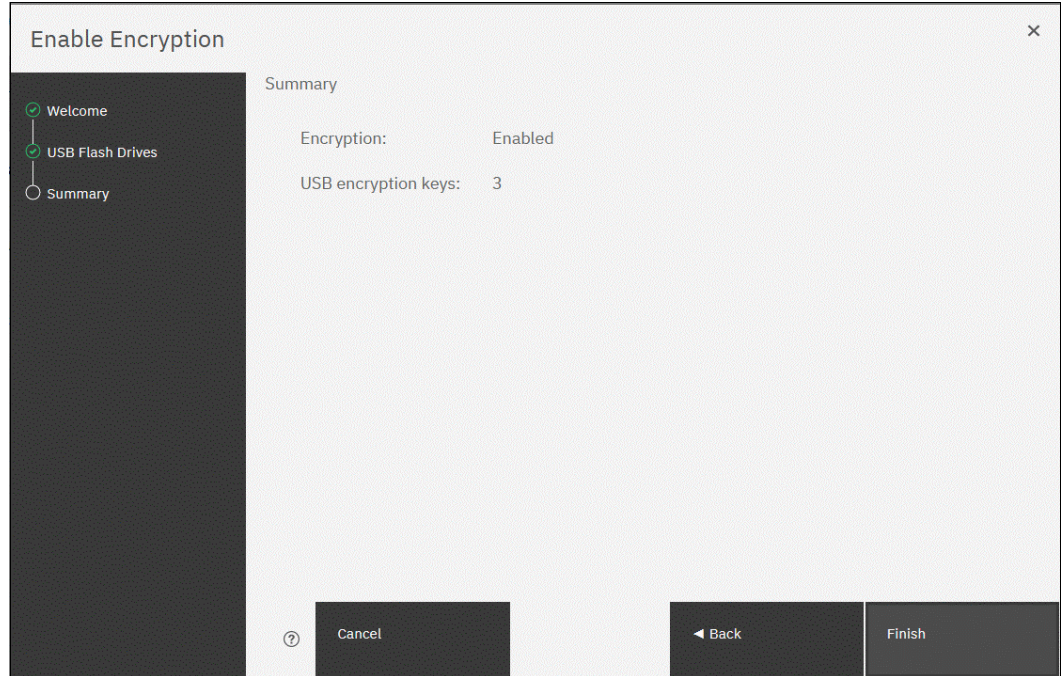


Figure 12-26 Committing the encryption enablement

You receive a message confirming that the encryption is now enabled on the system, as shown in Figure 12-27.

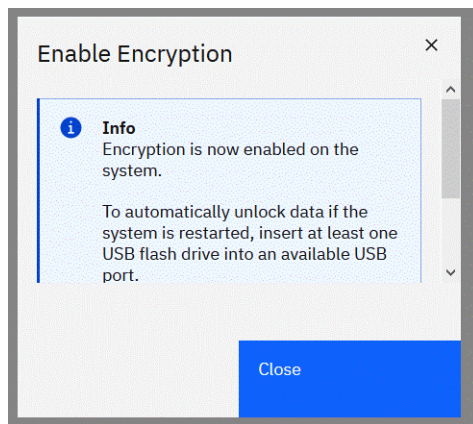


Figure 12-27 Encryption enabled message by using a USB flash drive

5. You can confirm that encryption is enabled and verify which key providers are in use by selecting **Settings** → **Security** → **Encryption**, as shown in Figure 12-28.

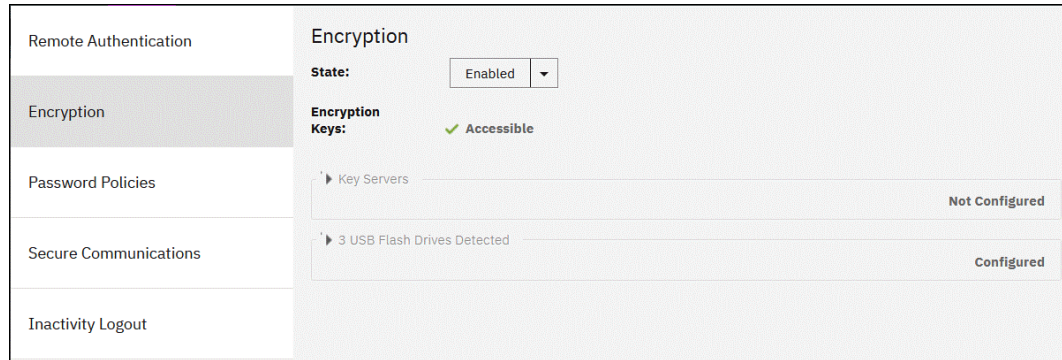


Figure 12-28 Encryption view that uses USB flash drives as the enabled provider

12.5.3 Enabling encryption by using key servers

A key server is a centralized system that receives and then distributes encryption keys to its clients, including IBM Spectrum Virtualize systems.

IBM Spectrum Virtualize supports the following key servers as encryption key providers:

- ▶ IBM Security Key Lifecycle Manager
- ▶ Gemalto SafeNet KeySecure

Note: Support for IBM Security Key Lifecycle Manager was introduced in IBM Spectrum Virtualize V7.8. Support for Gemalto SafeNet KeySecure was introduced in IBM Spectrum Virtualize V8.2.1.

IBM Security Key Lifecycle Manager and SafeNet KeySecure support KMIP, which is a standard for the management of cryptographic keys.

Note: Make sure that the key management server function is fully independent from the encrypted storage that has encryption that is managed by this key server environment. Failure to observe this requirement might create an encryption deadlock. An *encryption deadlock* is a situation in which none of key servers in the environment can become operational because some critical part of the data in each server is stored on a storage system that depends on one of the key servers to unlock access to the data.

IBM Spectrum Virtualize V8.1 and later supports up to four key server objects that are defined in parallel. But, only one key server type (IBM Security Key Lifecycle Manager or KeySecure) can be enabled at one time.

Another characteristic when working with key servers is that it is not possible to migrate from one key server type directly to another. If you want to migrate from one type to another, you first must migrate from your current key server to USB encryption, and then migrate from USB to the other type of key server.

Enabling encryption by using IBM Security Key Lifecycle Manager

Before you create a key server object in the storage system, the key server must be configured. Ensure that you complete the following tasks on the IBM Security Key Lifecycle Manager server before you enable encryption on the storage system:

- ▶ Configure the IBM Security Key Lifecycle Manager server to use Transport Layer Security V1.2. The default setting is TLSv1, but IBM Spectrum Virtualize supports only Version 1.2. So, set the value of security protocols to SSL_TLSv2 (which is a set of protocols that includes TLS V1.2) in the IBM Security Key Lifecycle Manager server configuration properties.
- ▶ Ensure that the database service is started automatically on startup.
- ▶ Ensure that there is at least one Secure Sockets Layer (SSL) certificate for browser access.
- ▶ Create an IBM Spectrum_VIRT device group for IBM Spectrum Virtualize systems.

For more information about completing these tasks, see [IBM Documentation](#).

Access to the key server that stores the correct master access key is required to enable access to encrypted data in the system after a system restart. A system restart might be a system-wide restart or power loss. Access to the key server is not required during a warm restart, such as node-exiting service mode or a single node restart. The data center power-on procedure must ensure key server availability before the storage system that uses encryption starts. If a system with encrypted data restarts and does not have access to the encryption keys, then the encrypted storage pools are offline until the encryption keys are detected.

To enable encryption by using an IBM Security Key Lifecycle Manager key server, complete the following steps:

1. Ensure that service IP addresses are configured on all your nodes.
2. In the Enable Encryption wizard **Welcome** tab, select **Key servers** and click **Next**, as shown in Figure 12-29.

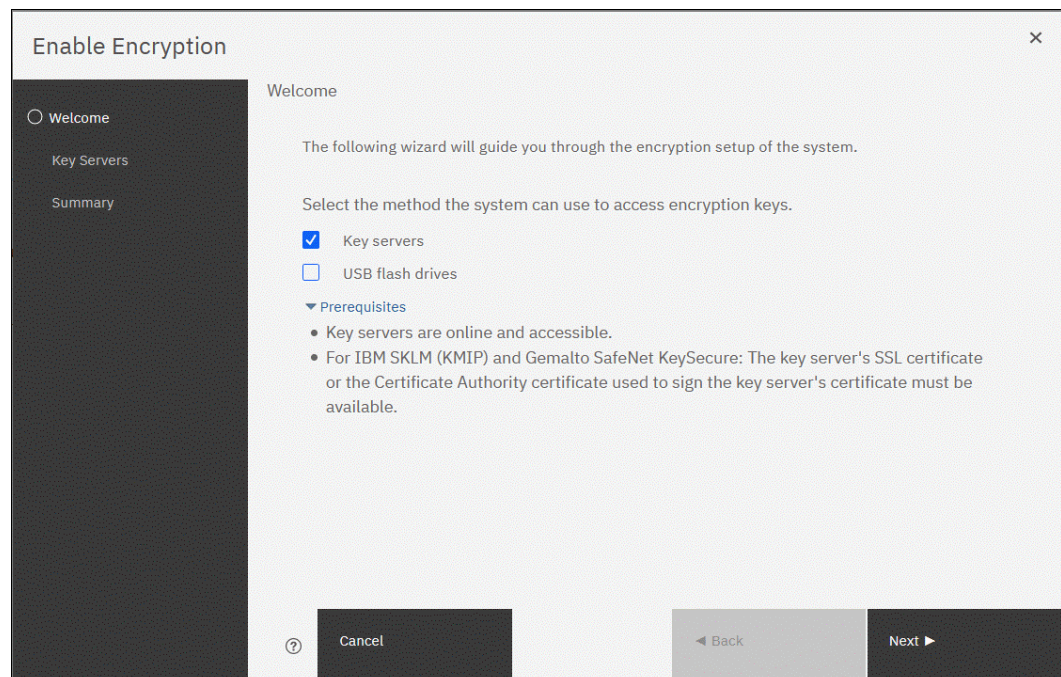


Figure 12-29 Selecting the key server as the only provider in the Enable Encryption wizard

3. Select **IBM SKLM (with KMIP)** as the key server type, as shown in Figure 12-30.

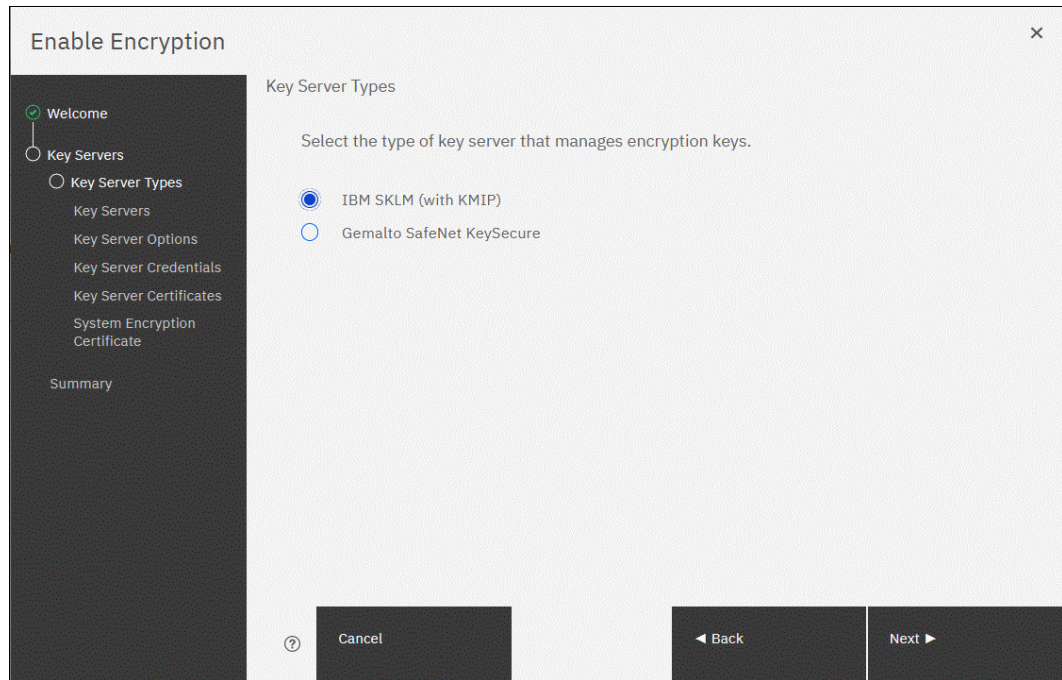


Figure 12-30 Selecting IBM Security Key Lifecycle Manager as the key server type

4. The wizard opens the Key Servers tab, as shown in Figure 12-31 on page 762. Enter the name and **Internet Protocol (IP) address** of the key servers. The first key server that is specified must be the primary IBM Security Key Lifecycle Manager key server.

Note: The supported versions of IBM Security Key Lifecycle Manager (up to Version 4.0, which was the latest code version that was available at the time of writing) differentiate between the primary and secondary key server role. The primary IBM Security Key Lifecycle Manager server as defined on the Key Servers window of the Enable Encryption wizard must be the server that is defined as the primary by IBM Security Key Lifecycle Manager administrators.

The key server name serves only as a label. Only the provided IP address is used to contact the server. If the key server's TCP port number differs from the default value for the KMIP protocol (that is, 5696), enter the port number.

An example of a complete primary IBM Security Key Lifecycle Manager configuration is shown in Figure 12-31.

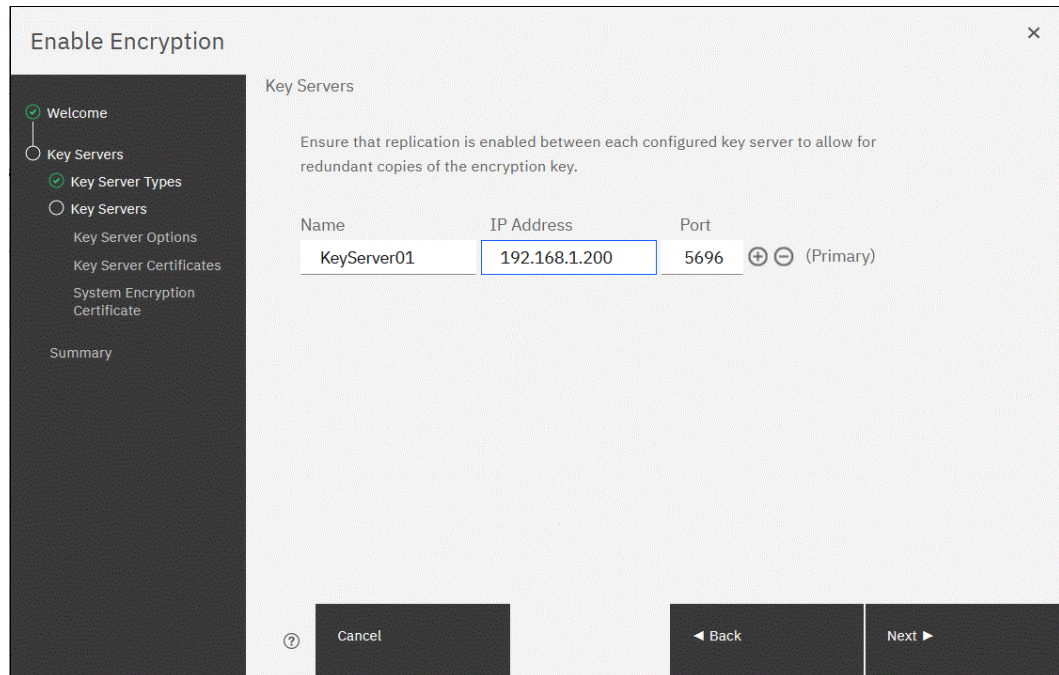


Figure 12-31 Configuring the primary IBM Security Key Lifecycle Manager server

5. If you want to add secondary IBM Security Key Lifecycle Manager servers, click the + symbol and enter the data for the secondary IBM Security Key Lifecycle Manager servers, as shown in Figure 12-32. You can define up to three extra IBM Security Key Lifecycle Manager servers. Click **Next** when you are done.

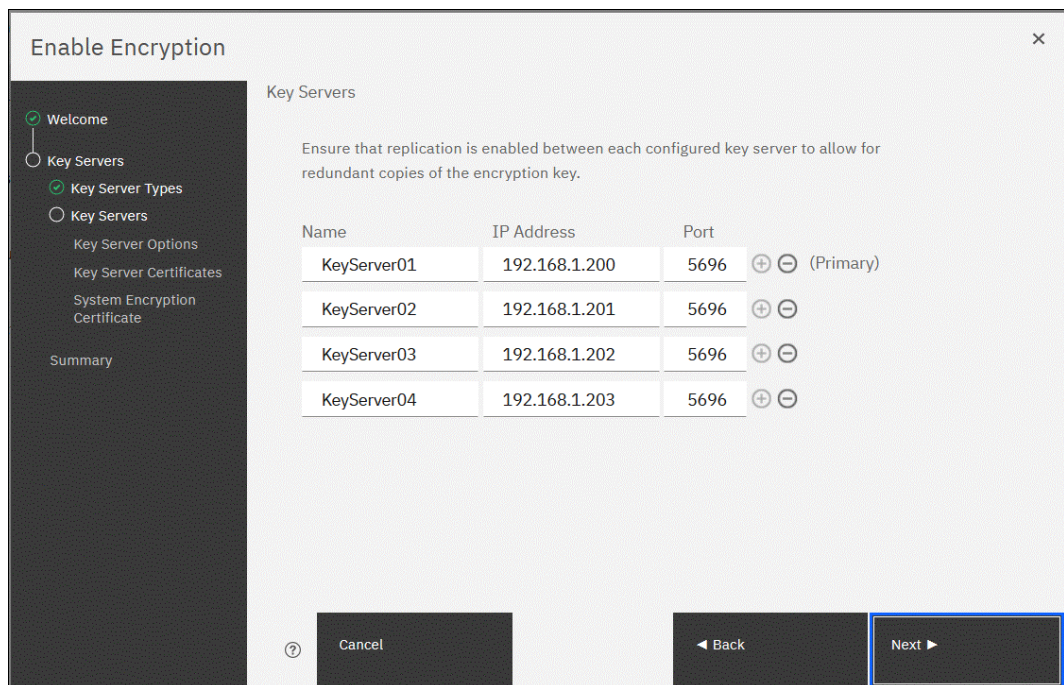


Figure 12-32 Configuring multiple IBM Security Key Lifecycle Manager servers

- The next window in the wizard is a reminder that the Spectrum_VIRT device group that is dedicated for IBM Spectrum Virtualize systems must exist on the IBM Security Key Lifecycle Manager key servers. Make sure that this device group exists and click **Next** to continue, as shown in Figure 12-33.

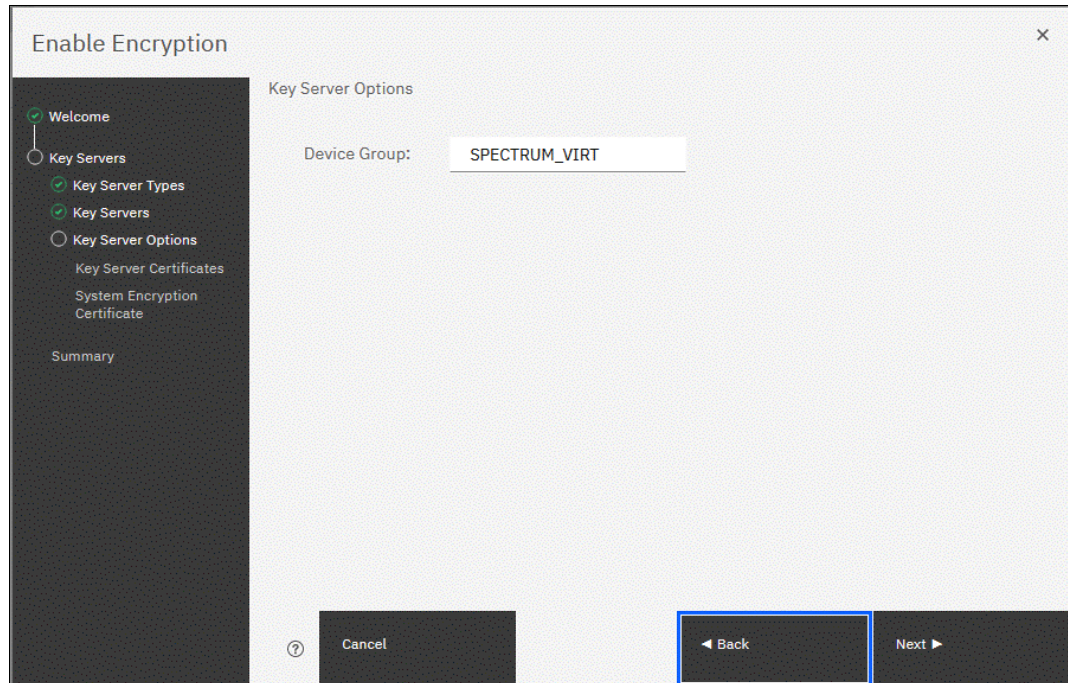


Figure 12-33 Checking the key server device group

- Enable secure communication between the IBM Spectrum Virtualize system and the IBM Security Key Lifecycle Manager key servers by uploading the key server certificate from a trusted third-party certificate authority (CA) or by using a self-signed certificate. The self-signed certificate can be obtained from each of the key servers directly.

After uploading the certificates in the window that is shown in Figure 12-34, click **Next**.

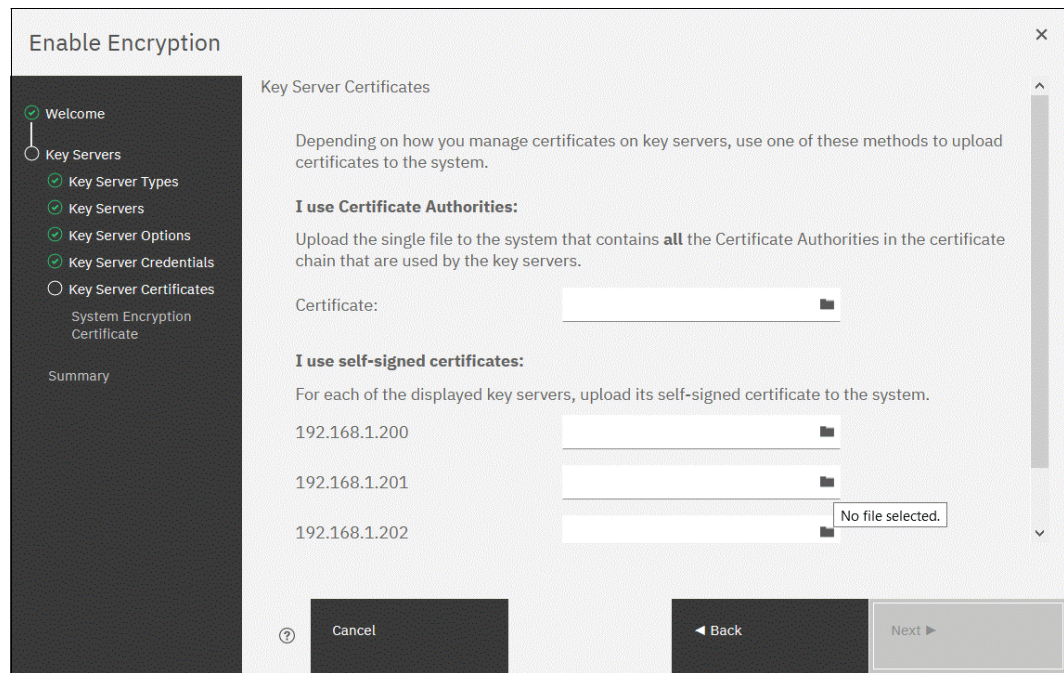


Figure 12-34 Uploading the key-server or certificate authority SSL certificate

8. Configure the IBM Security Key Lifecycle Manager key server to trust the public key certificate of the IBM Spectrum Virtualize system. You can download the IBM Spectrum Virtualize system public SSL certificate by clicking **Export Public Key**, as shown in Figure 12-35. Install this certificate in the IBM Security Key Lifecycle Manager key server in the Spectrum_VIRT device group.

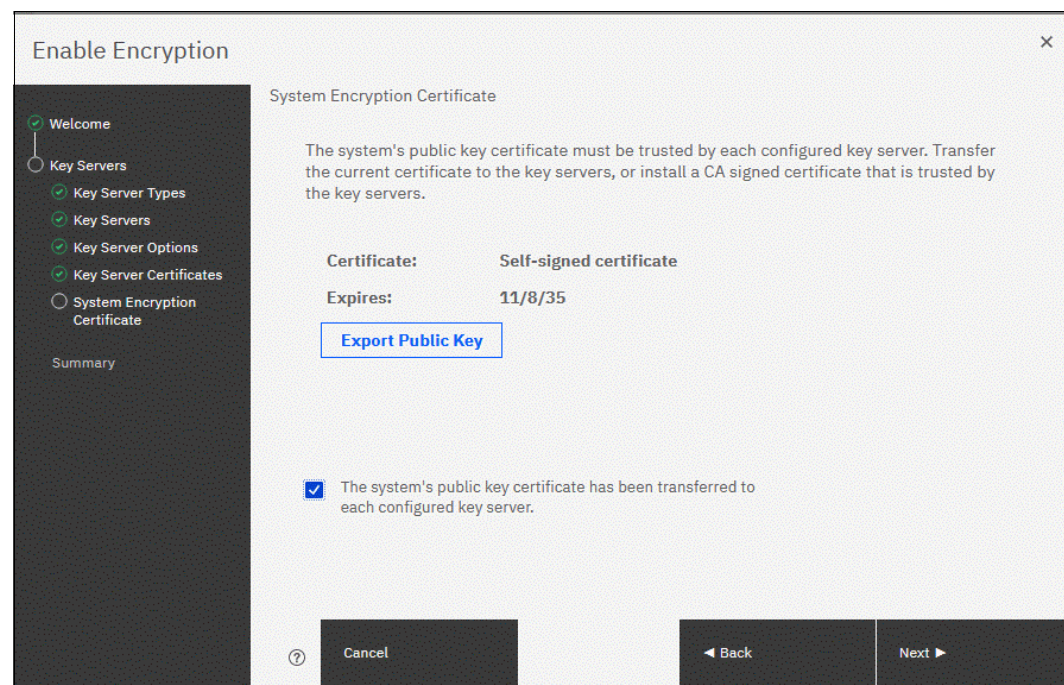


Figure 12-35 Downloading the IBM Spectrum Virtualize SSL certificate

9. When the IBM Spectrum Virtualize system public key certificate is installed on the IBM Security Key Lifecycle Manager key servers, acknowledge this installation by clicking the checkbox below the **Export Public Key** button and click **Next**.
10. The key server configuration is shown in the **Summary** tab, as shown in Figure 12-36. Click **Finish** to create the key server object and finalize the encryption enablement.

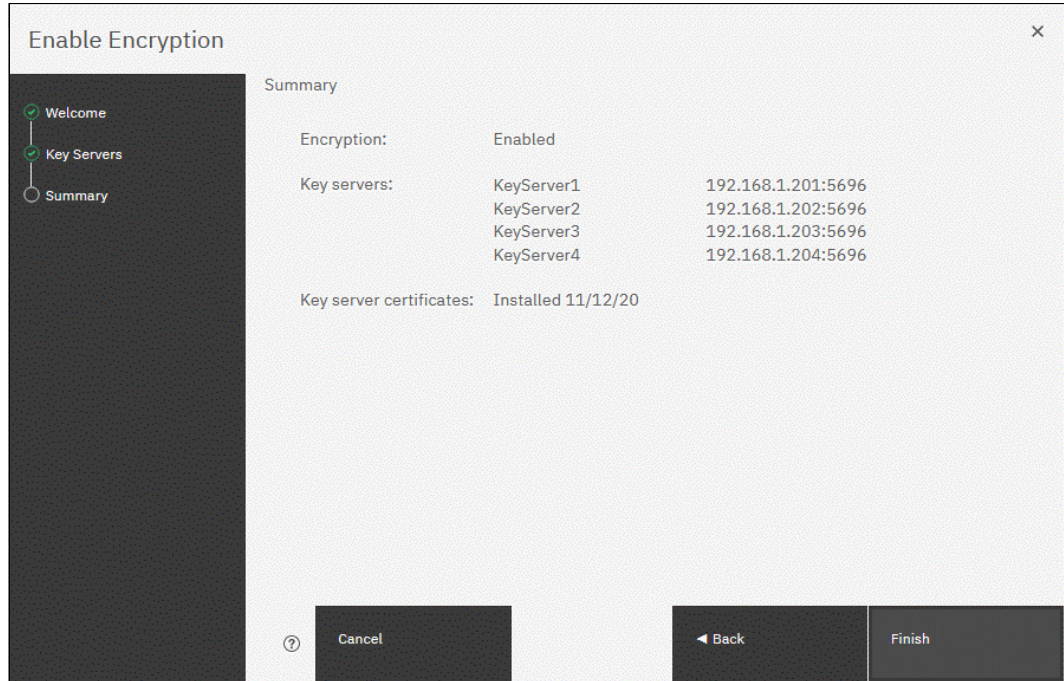


Figure 12-36 Finishing the enablement of encryption by using IBM Security Key Lifecycle Manager key servers

11. If no errors occur while the key server object is created, you receive a message that confirms that the encryption is now enabled on the system. Click **Close**.
12. Confirm that encryption is enabled by selecting **Settings** → **Security** → **Encryption**, as shown in Figure 12-37. The **Online** state that indicates which IBM Security Key Lifecycle Manager servers are detected as available by the system.

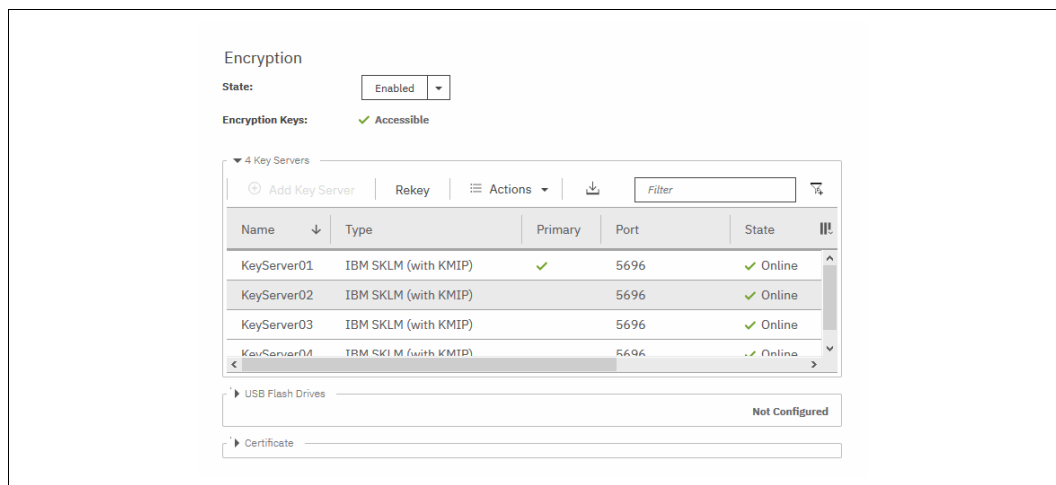


Figure 12-37 Encryption that is enabled with only IBM Security Key Lifecycle Manager servers as encryption key providers

Enabling encryption by using SafeNet KeySecure

IBM Spectrum Virtualize V8.2.1 introduced support for Gemalto SafeNet KeySecure, which is a third-party key management server. It can be used as an alternative to IBM Security Key Lifecycle Manager.

IBM Spectrum Virtualize supports Gemalto SafeNet KeySecure V8.3.0 and later, and uses only the KMIP protocol. It is possible to configure up to four SafeNet KeySecure servers in IBM Spectrum Virtualize for redundancy, and they can coexist with USB flash drive encryption.

It is not possible to have both SafeNet KeySecure and IBM Security Key Lifecycle Manager key servers that are configured concurrently in IBM Spectrum Virtualize. It is also not possible to migrate directly from one type of key server to another (from IBM Security Key Lifecycle Manager to SafeNet KeySecure or vice versa). If you want to migrate from one type to another, first migrate to USB flash drives encryption, and then migrate to the other type of key servers.

KeySecure uses an active-active clustered model. All changes to one key server are instantly propagated to all other servers in the cluster.

Although KeySecure uses the KMIP protocol like IBM Security Key Lifecycle Manager does, an option is available to configure the username and password for IBM Spectrum Virtualize and KeySecure server authentication, which is not possible when the configuration is performed with IBM Security Key Lifecycle Manager.

The certificate for client authentication in SafeNet KeySecure can be self-signed or signed by a CA.

To enable encryption in IBM Spectrum Virtualize by using a Gemalto SafeNet KeySecure key server, complete the following steps:

1. Ensure that the service IP addresses are configured on all your nodes.
2. In the Enable Encryption wizard **Welcome** tab, select **Key servers** and click **Next**, as shown in Figure 12-38 on page 767.

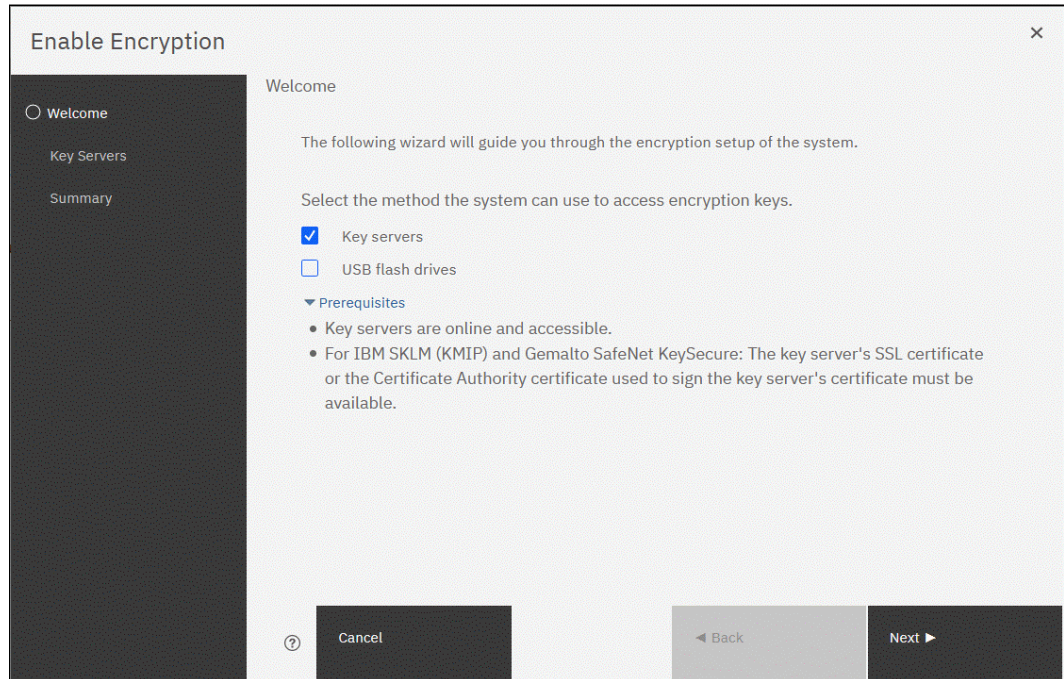


Figure 12-38 Selecting key servers as the only provider in the Enable Encryption wizard

3. In the next window, you can choose between the IBM Security Key Lifecycle Manager or Gemalto SafeNet KeySecure server types, as shown in Figure 12-39. Select **Gemalto SafeNet KeySecure** and click **Next**.

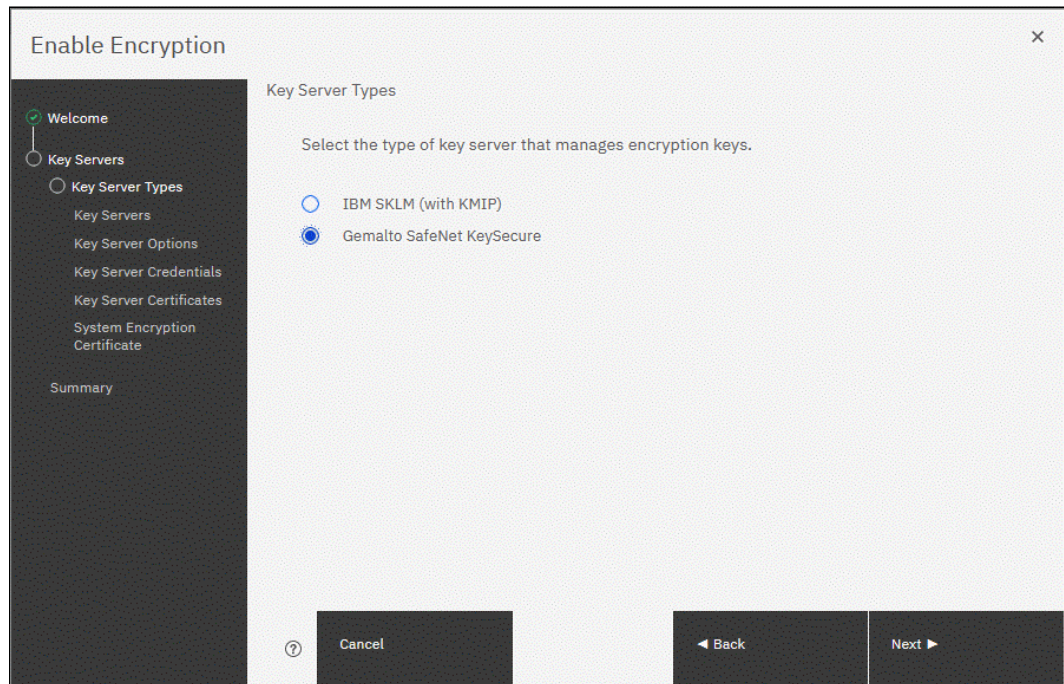


Figure 12-39 Selecting Gemalto SafeNet KeySecure as the key server type

4. Add up to four SafeNet KeySecure servers in the next wizard window, as shown in Figure 12-40. For each key server, enter the name, IP address, and TCP port for the KMIP protocol (the default value is 5696). The server name is only a label, so it does not need to be the real hostname of the server.

Although Gemalto SafeNet KeySecure uses an active-active clustered model, IBM Spectrum Virtualize asks for a primary key server. The primary key server represents only the KeySecure server that is used for key create and rekey operations. Therefore, any of the clustered key servers can be selected as the primary.

Selecting a primary key server is beneficial for load balancing. Any four key servers can be used to retrieve the master key.

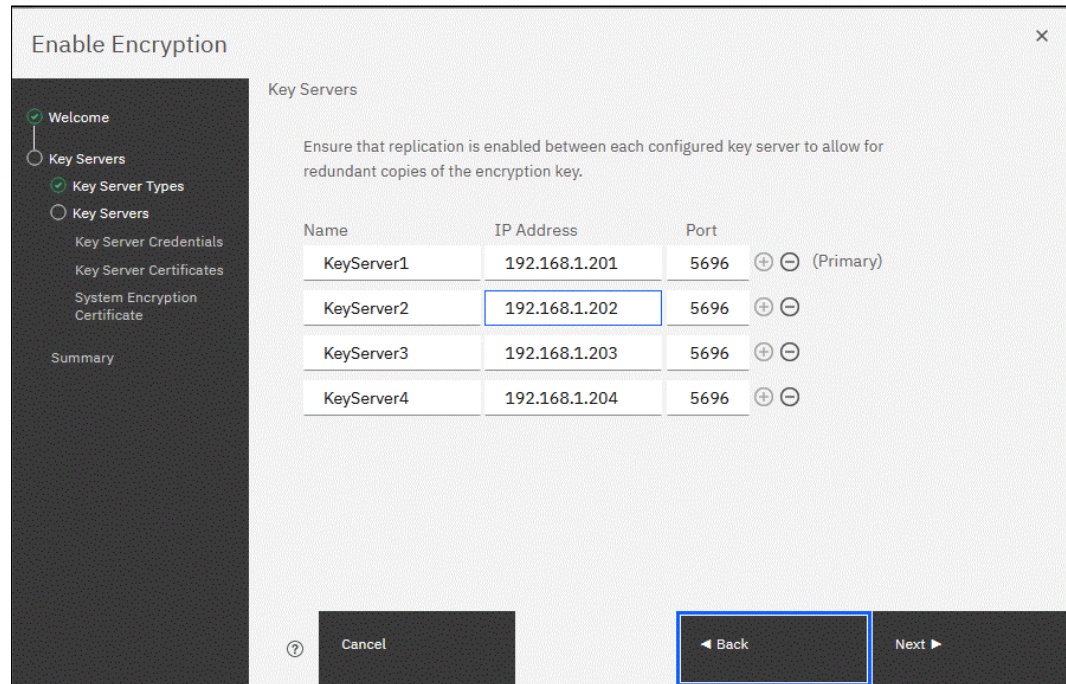


Figure 12-40 Configuring multiple SafeNet KeySecure servers

5. The next window in the wizard prompts for the key servers' credentials (username and password), as shown in Figure 12-41 on page 769. This setting is optional because it depends on how the SafeNet KeySecure servers are configured.

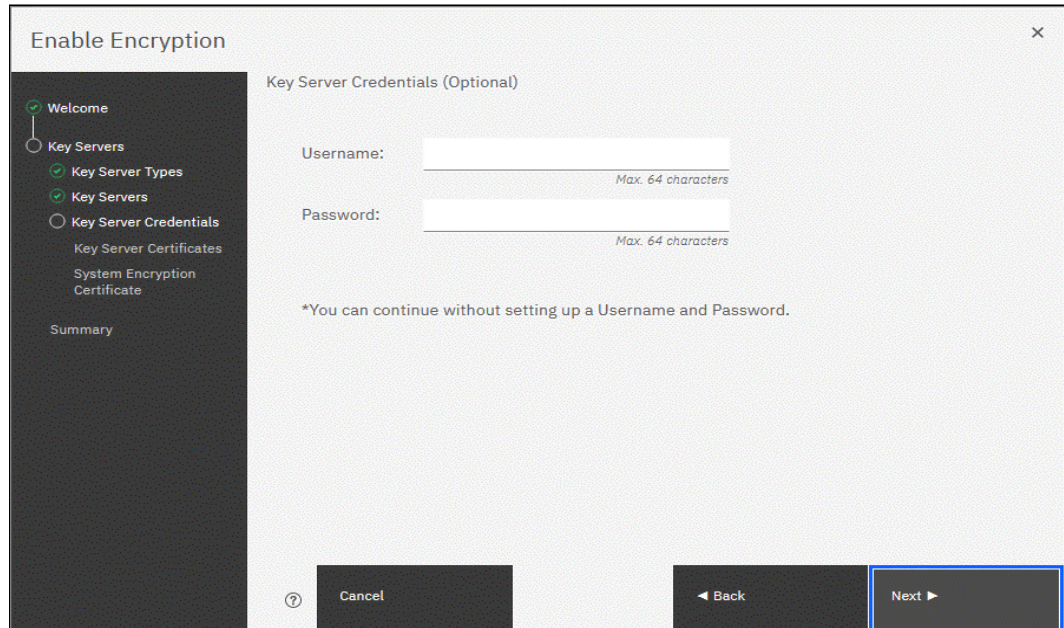


Figure 12-41 Key server credentials input (optional)

6. Enable secure communication between the IBM Spectrum Virtualize system and the SafeNet KeySecure key servers by uploading the key server certificate from a trusted third-party CA or by using a self-signed certificate. The self-signed certificate can be obtained from each of key servers directly. After uploading any of the certificates in the window that is shown in Figure 12-42, click **Next**.

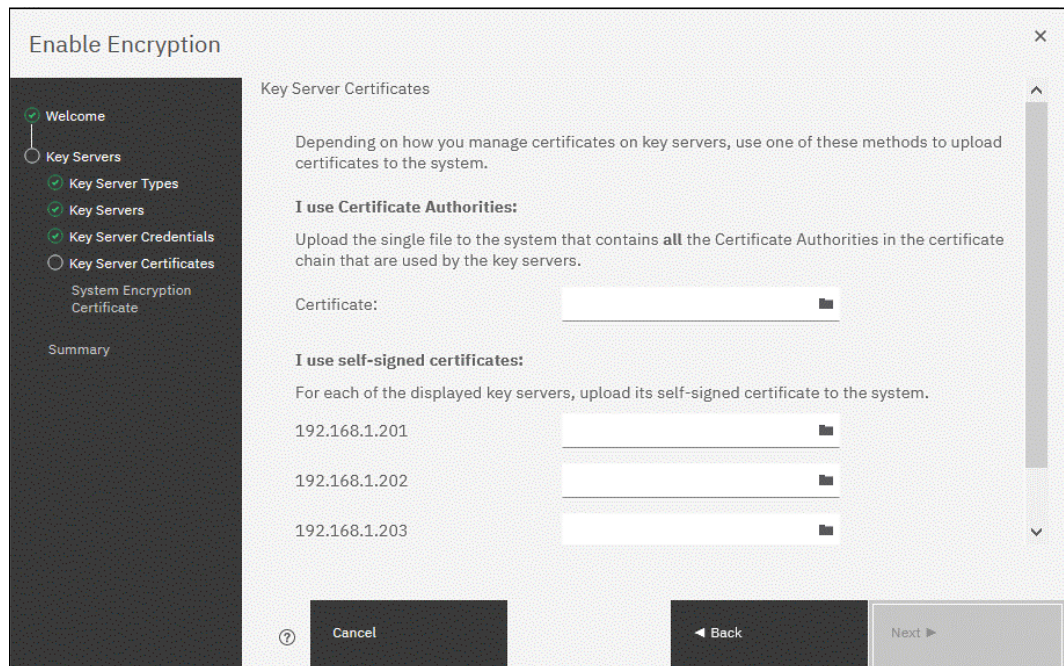


Figure 12-42 Uploading the SafeNet KeySecure key servers certificate

- Configure the SafeNet KeySecure key servers to trust the public key certificate of the IBM Spectrum Virtualize system. You can download the IBM Spectrum Virtualize system public SSL certificate by clicking **Export Public Key**, as shown in Figure 12-43. After adding the public key certificate to the key servers, select the checkbox and click **Next**.

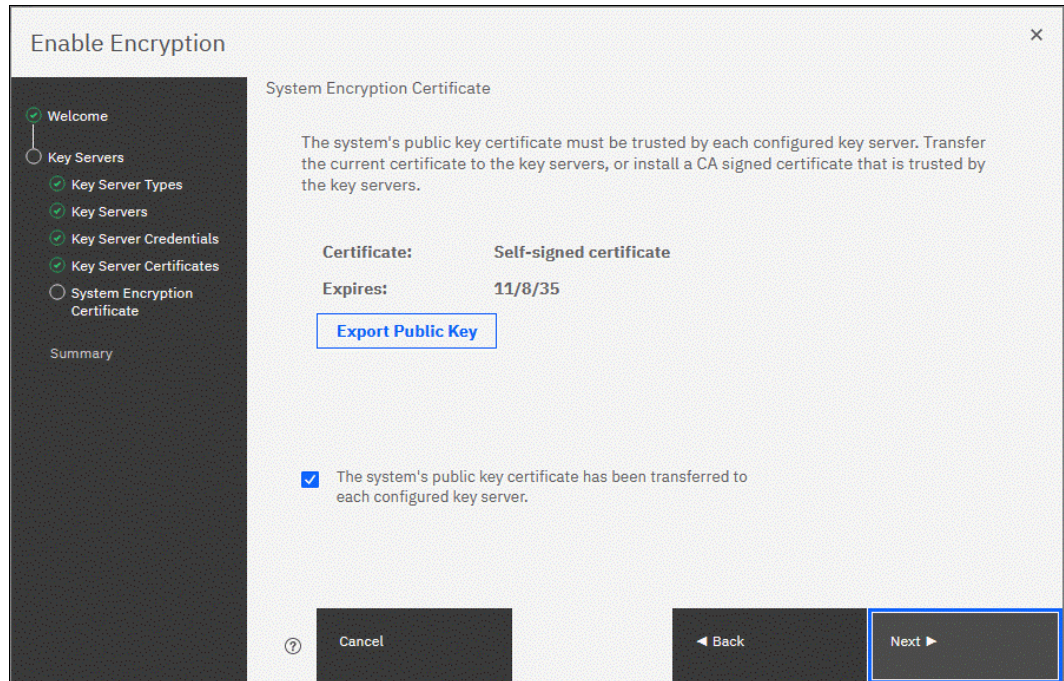


Figure 12-43 Downloading the IBM Spectrum Virtualize SSL certificate

- The key server configuration is shown in the **Summary** tab, as shown in Figure 12-44. Click **Finish** to create the key server object and finalize the encryption enablement.

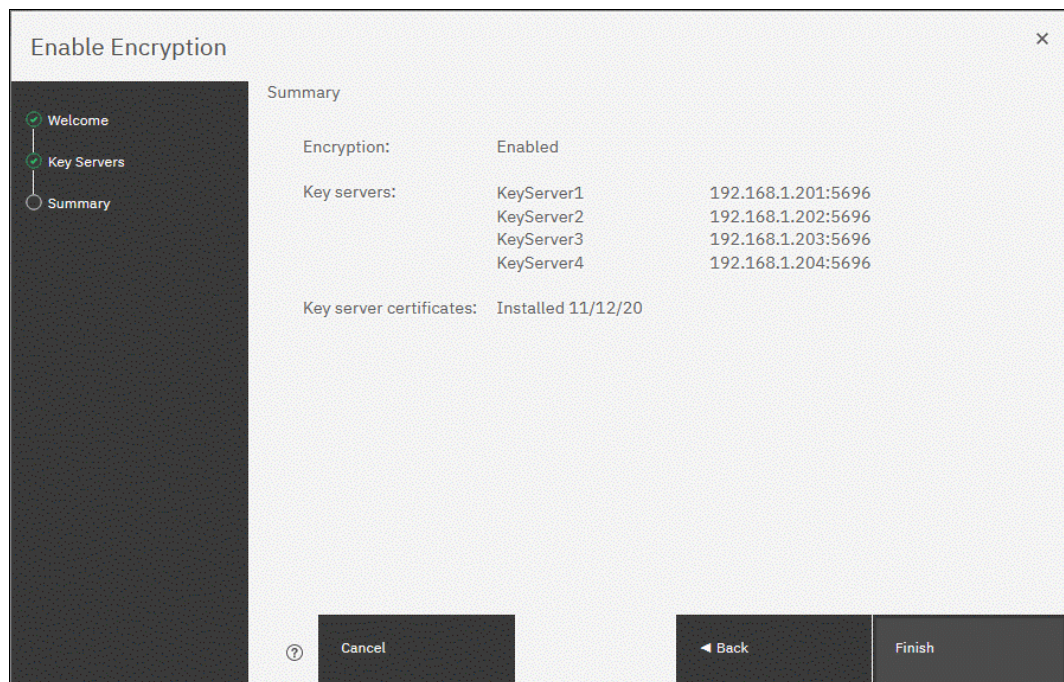


Figure 12-44 Finishing the enablement of encryption by using SafeNet KeySecure key servers

9. If no errors occurred while creating the key server object, you receive a message that confirms that the encryption is now enabled on the system. Click **Close**.
10. Confirm that encryption is enabled by selecting **Settings** → **Security** → **Encryption**, as shown in Figure 12-45. Check whether the four servers are shown as online, which indicates that all four SafeNet KeySecure servers are detected as available by the system.

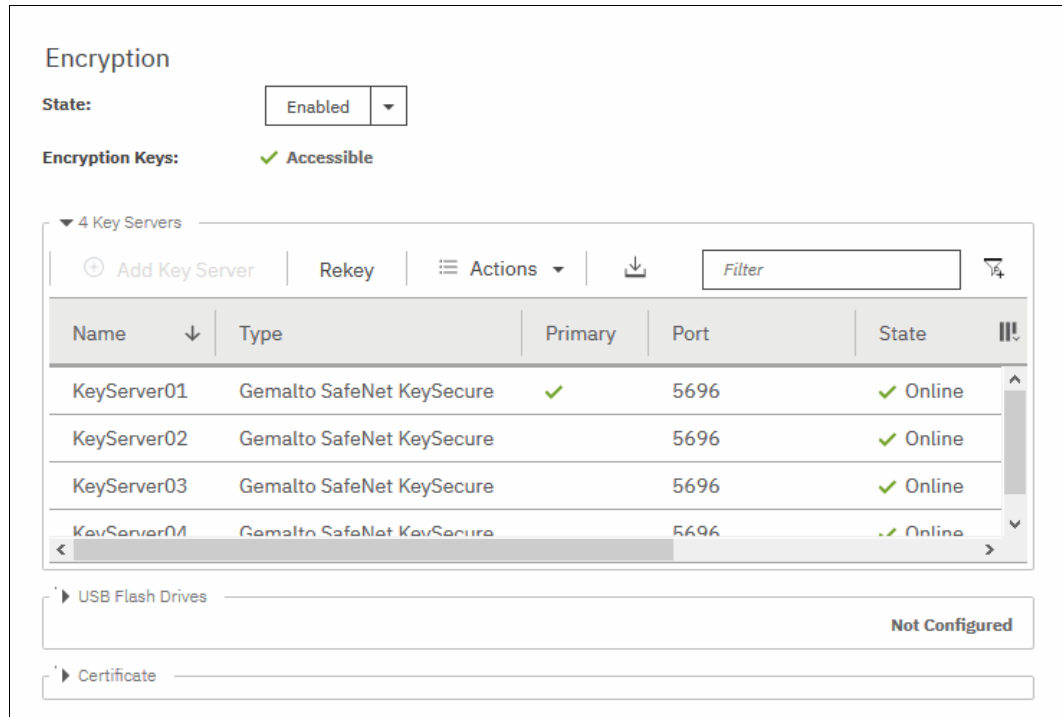


Figure 12-45 Encryption that is enabled with four SafeNet KeySecure key servers

12.5.4 Enabling encryption by using both providers

IBM Spectrum Virtualize enables parallel use of both a USB flash drive and one type of key server (IBM Security Key Lifecycle Manager or SafeNet KeySecure) as encryption key providers. It is possible to configure both providers in a single run of the Encryption Enable wizard. To perform this configuration, the system must meet requirements of both key server (IBM Security Key Lifecycle Manager or SafeNet KeySecure) and USB flash drive encryption key providers.

Note: Make sure that the key management server function is fully independent from encrypted storage that has encryption that is managed by this key server environment. Failure to observe this requirement might create an encryption deadlock. An encryption deadlock is a situation in which none of key servers in the environment can become operational because some critical part of the data in each server is stored on an encrypted storage system that depends on one of the key servers to unlock access to the data.

IBM Spectrum Virtualize V8.1 and later supports up to four key server objects that are defined in parallel.

Before you enable encryption by using both USB flash drives and key servers, confirm the requirements that are described in 12.5.2, “Enabling encryption by using USB flash drives” on page 755 and 12.5.3, “Enabling encryption by using key servers” on page 759.

To enable encryption by using a key server and USB flash drive, complete the following steps:

1. Ensure that you have service IP addresses that are configured on all your nodes.
2. In the Enable Encryption wizard **Welcome** tab, select **Key servers** and **USB flash drives** and click **Next**, as shown in Figure 12-46.

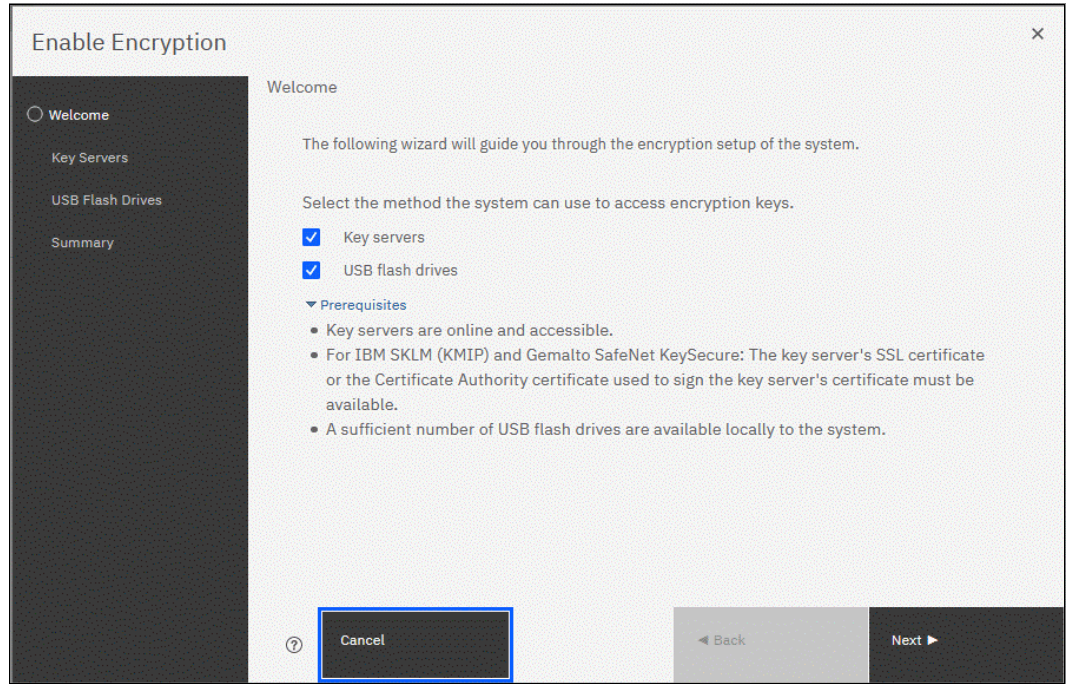


Figure 12-46 Selecting key servers and USB flash drives in the Enable Encryption wizard

3. The wizard opens the Key Server Types window, as shown in Figure 12-47. Select the key server type that manages the encryption keys.

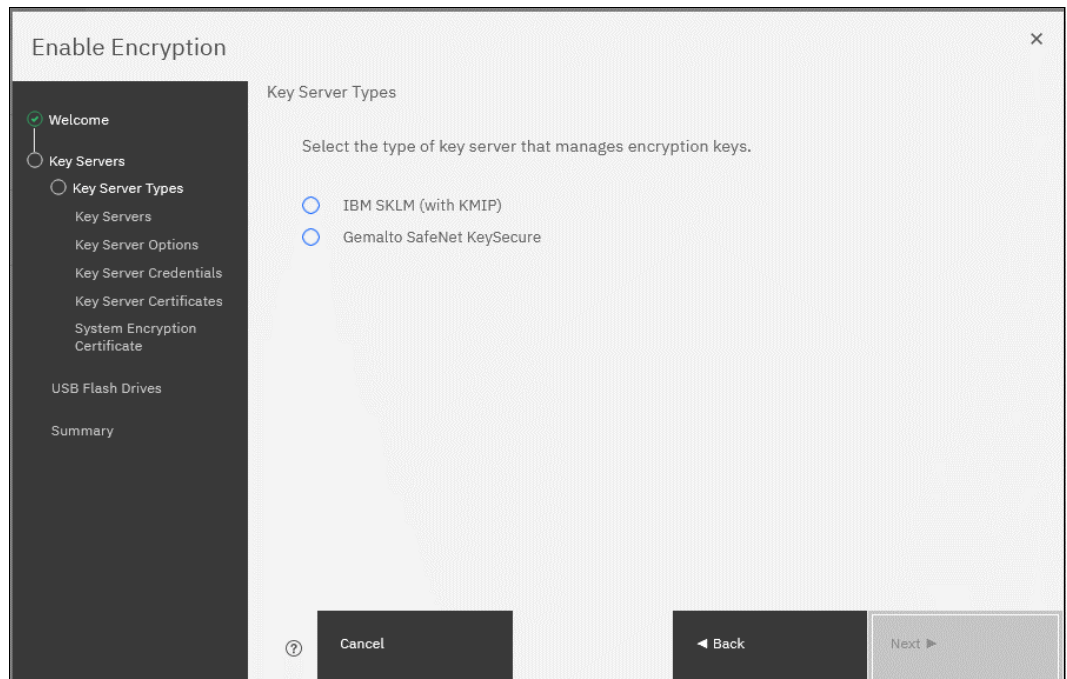


Figure 12-47 Selecting the key server type

The next flow of the actions is the same as described in 12.5.3, “Enabling encryption by using key servers” on page 759, depending on the type of key server that is selected.

When the key servers details are entered, the USB flash drive encryption configuration is displayed. In this step, the master encryption key copies are stored in the USB flash drives. If fewer than three drives are detected, the system requests that you plug in more USB flash drives. You cannot proceed until the required minimum number of USB flash drives is detected by the system.

After at least three USB flash drives are detected, the system writes the master access key to each of the drives, as shown in Figure 12-48. The system attempts to write the encryption key to any flash drive that it detects. Therefore, it is crucial to maintain the physical security of the system during this procedure.

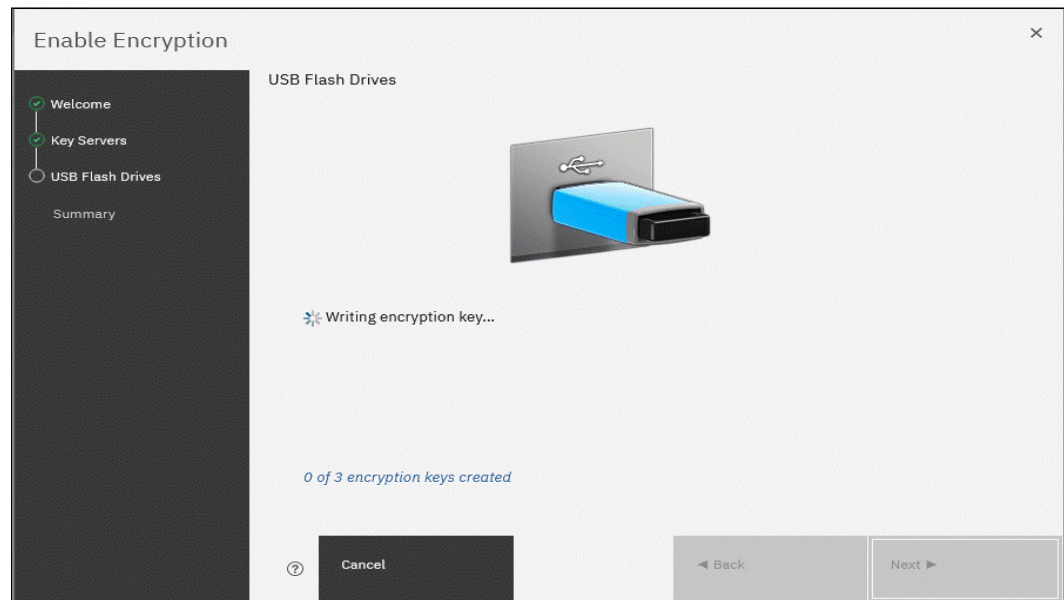


Figure 12-48 Master access key that is writing to the USB flash drives

4. After copying the encryption keys to USB flash drives, a window opens and shows a summary of the configuration that is implemented on the system. Click **Finish** to create the key server object and finalize the encryption enablement.

If no errors occur while creating the key server object, the system displays a window that confirms that the encryption is now enabled on the system and that both encryption key providers are enabled.

5. You can confirm that encryption is enabled and verify which key providers are in use by selecting **Settings** → **Security** → **Encryption**. Note the state *Online* of key servers and state *Validated* of USB ports where USB flash drives are inserted to make sure that they are properly configured.

12.6 Configuring more providers

After the system is configured with a single encryption key provider, a second provider can be added.

Note: If you set up encryption of your storage system when it was running a version of IBM Spectrum Virtualize earlier than Version 7.8.0, you must rekey the master encryption key before you can enable a second encryption provider when you upgrade to Version 8.1 or later.

12.6.1 Adding key servers as a second provider

If the storage system is configured with the USB flash drive provider, it is possible to configure IBM Security Key Lifecycle Manager or SafeNet KeySecure servers as a second provider. To enable key servers as a second provider, complete the following steps:

1. Select **Settings** → **Security** → **Encryption**. Expand the **Key Servers** section and click **Configure**, as shown in Figure 12-49. To enable a key server as a second provider, the system must detect at least one USB flash drive with a current copy of the master access key.

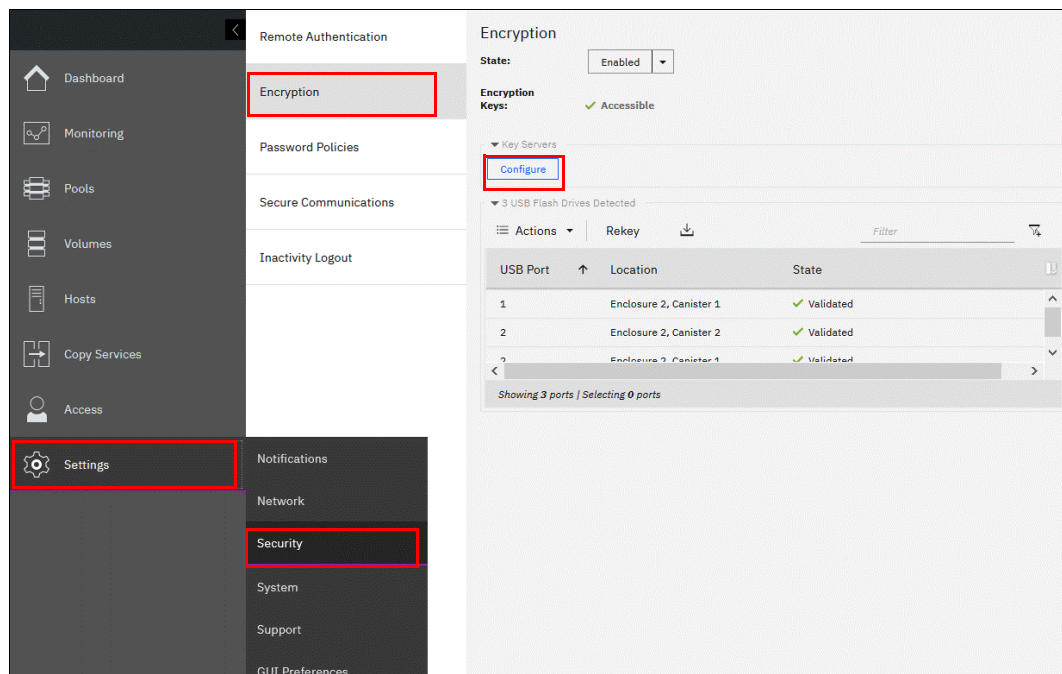


Figure 12-49 Enabling key servers as a second provider

2. Complete the steps that are required to configure the key server provider, as described in 12.5.3, “Enabling encryption by using key servers” on page 759. The difference in the process that is described in that section is that the wizard gives you an option to disable USB flash drive encryption, which aims to migrate from the USB flash drive to key server provider.

Select **No** to enable both encryption key providers, as shown in Figure 12-50 on page 775.

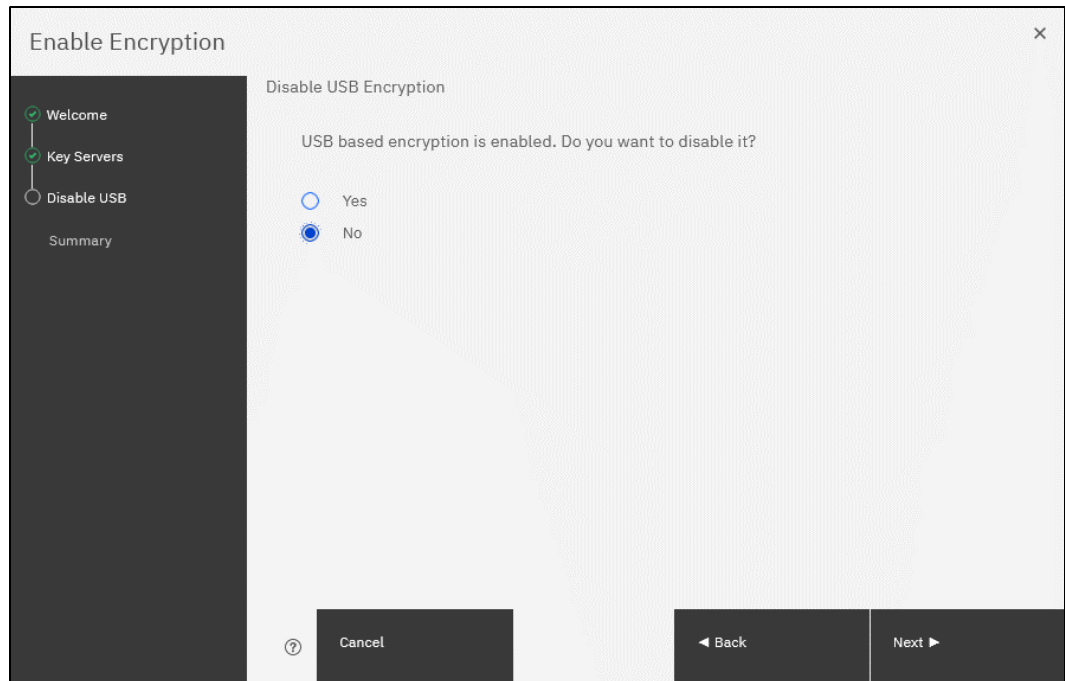


Figure 12-50 Do not disable the USB flash drive encryption key provider

This choice is confirmed on the summary window before the configuration is committed, as shown in Figure 12-51.

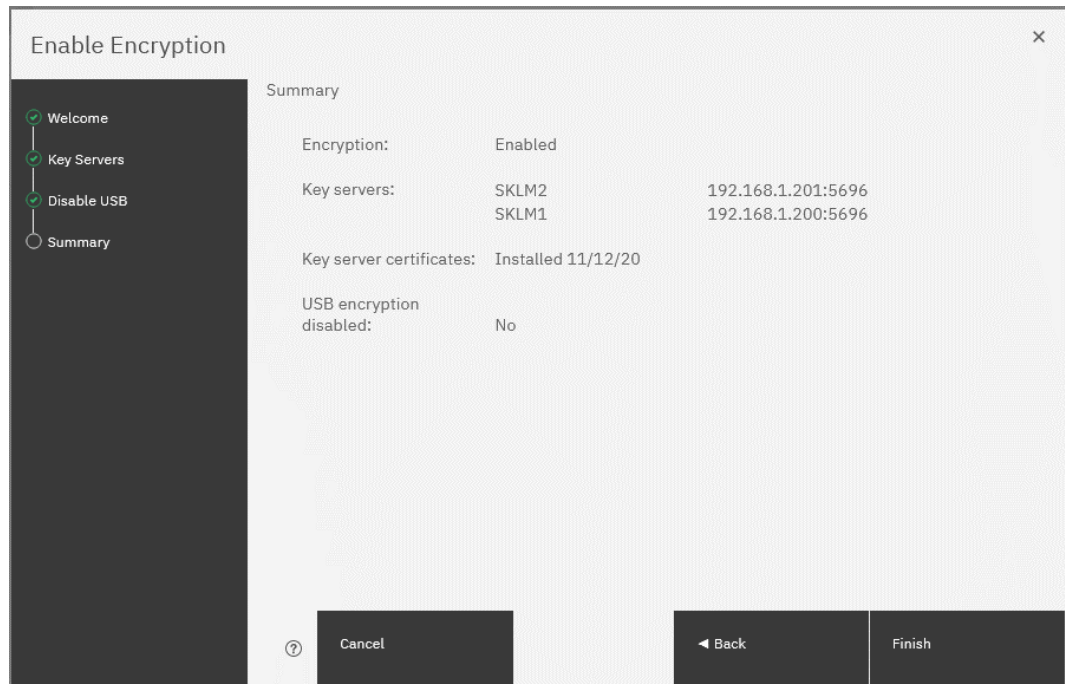


Figure 12-51 Configuration summary before committing

- After you click **Finish**, the system configures the keys servers as a second encryption key provider. Successful completion of the task is confirmed by a message. Click **Close**.

4. You can confirm that encryption is enabled and verify which key providers are in use by selecting **Settings** → **Security** → **Encryption**, as shown in Figure 12-52. Note the **Online** state of key servers and **Validated** state of USB ports where USB flash drives are inserted to make sure that they are properly configured.

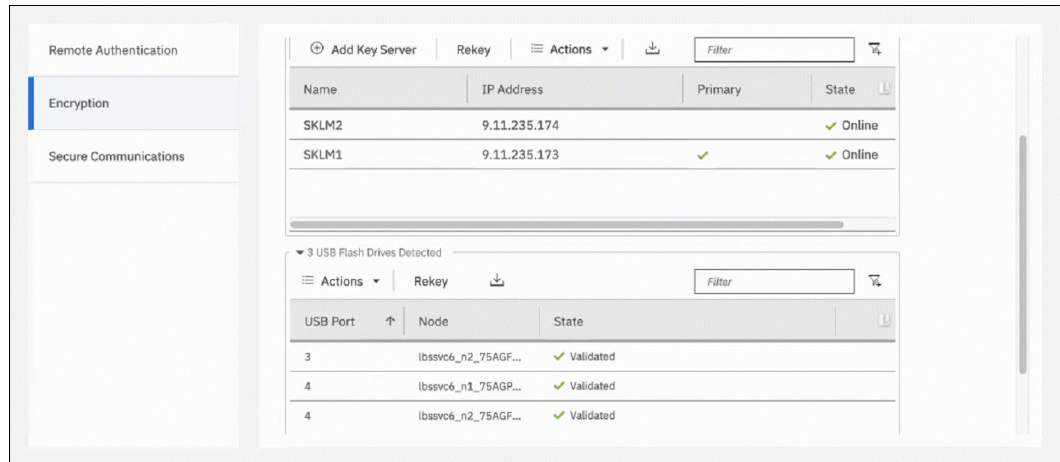


Figure 12-52 Encryption that is enabled with two key providers available

12.6.2 Adding USB flash drives as a second provider

If the storage system is configured with an IBM Security Key Lifecycle Manager or a SafeNet KeySecure encryption key provider, it is possible to configure USB flash drives as a second provider. To enable USB flash drives as a second provider, complete the following steps:

1. Select **Settings** → **Security** → **Encryption**. Expand the **USB Flash Drives** section and click **Configure**. To enable USB flash drives as a second provider, the system must access key servers by using the current master access key.
2. After you click **Configure**, you see a wizard like the one that is described in 12.5.2, “Enabling encryption by using USB flash drives” on page 755. You cannot disable key server providers during this process.

After successful completion of the process, you are presented with a message confirming that both encryption key providers are enabled.

3. You can confirm that encryption is enabled and verify which key providers are in use by selecting **Settings** → **Security** → **Encryption**, as shown in Figure 12-53 on page 777. Note the state **Online** state of key servers and **Validated** state of USB ports where USB flash drives are inserted to make sure that they are properly configured.

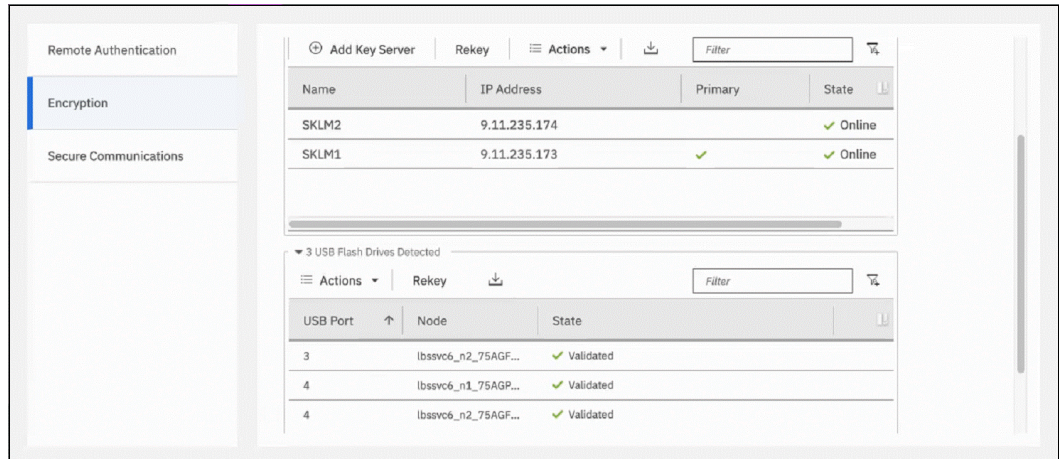


Figure 12-53 Encryption that is enabled with two key providers that are available

12.7 Migrating between providers

IBM Spectrum Virtualize V8.1 introduced support for simultaneous use of both USB flash drives and a key server as encryption key providers. The system also enables migration of a configuration by using only a USB flash drive provider to key servers provider path, and vice versa.

If you want to migrate from one key server type to another (for example, migrating from IBM Security Key Lifecycle Manager to SafeNet KeySecure or vice versa), direct migration is not possible. In this case, you first must migrate from the current key server type to a USB flash drive, and then migrate to the other type of key server.

12.7.1 Changing from a USB flash drive provider to an encryption key server

The system is designed to facilitate changing from a USB flash drives encryption key provider to an encryption key server provider. If you follow the steps that are described in 12.6.1, “Adding key servers as a second provider” on page 774, but when completing step 2 on page 774 you select **Yes** instead of **No** (see Figure 12-54). This action causes de-activation of the USB flash drives provider, and the procedure completes with only key servers that are configured as a key provider.

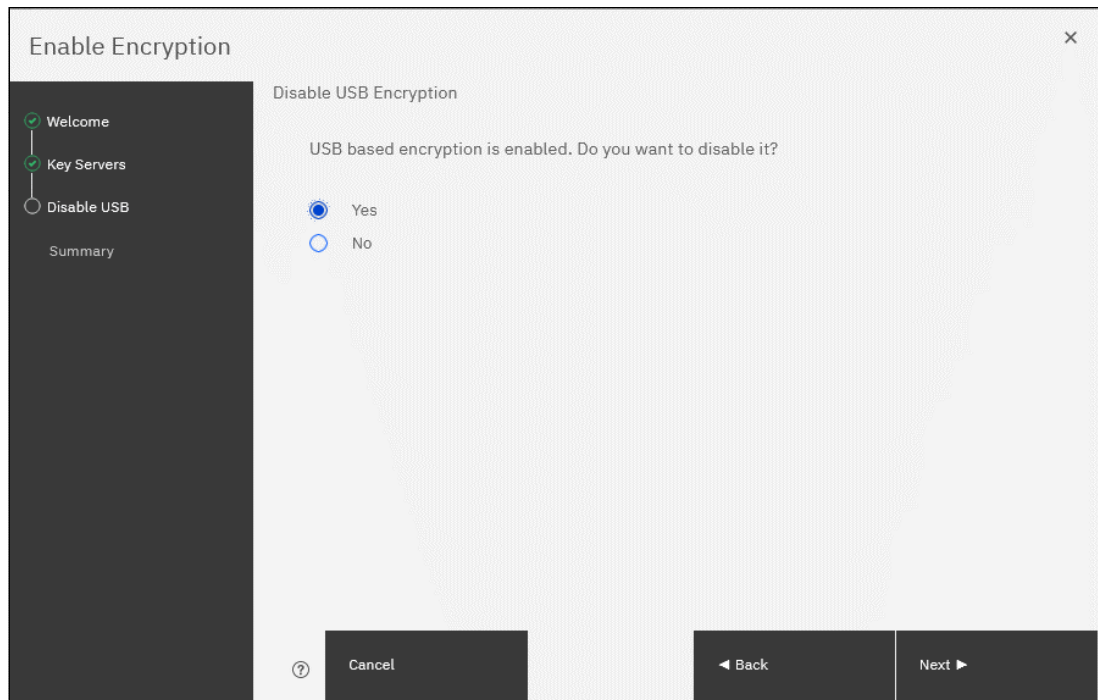


Figure 12-54 Disabling the USB flash drive provider while changing to the IBM Security Key Lifecycle Manager provider

12.7.2 Changing from an encryption key server to a USB flash drive provider

Changing from using encryption key servers provider to a USB flash drives provider is not possible by using only the GUI.

To change the direction, add USB flash drives as a second provider by completing the steps that are described in 12.6.2, “Adding USB flash drives as a second provider” on page 776.

Then, run the following command in the CLI:

```
chencryption -usb validate
```

To make sure that the USB drives contain the correct master access key, disable the encryption key server provider by running the following command:

```
chencryption -keyserver disable
```

This command disables the encryption key server provider, which effectively migrates your system from an encryption key server to a USB flash drive provider.

12.7.3 Migrating between different key server types

The migration between different key server types cannot be performed directly from one type of key server to another. USB flash drives encryption must be used to facilitate this task.

If you want to migrate from one type of key server to another, you first must migrate from your current key servers to USB encryption, and then migrate from USB to the other type of key servers.

The procedure to migrate from one key server type to another is shown here. In this example, we migrate an IBM Spectrum Virtualize system that is configured with IBM Security Key Lifecycle Manager key server (as shown in Figure 12-55) to SafeNet KeySecure servers.

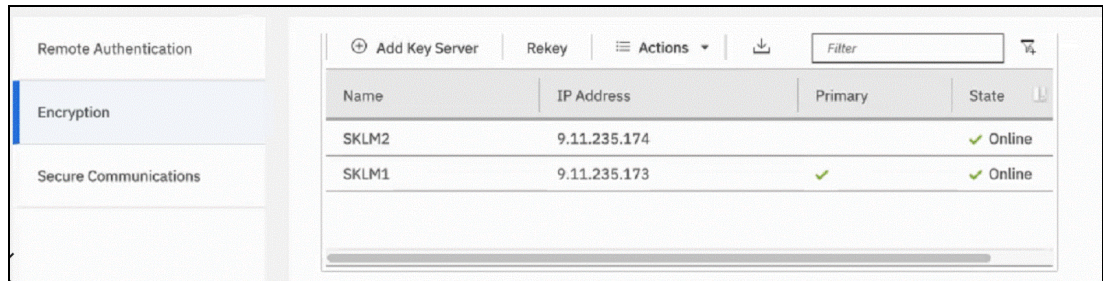


Figure 12-55 IBM Spectrum Virtualize encryption that is configured with IBM Security Key Lifecycle Manager servers

To migrate to Gemalto SafeNet KeySecure, complete the following steps:

1. Migrate from key server encryption to USB flash drives encryption, as described in 12.7.2, “Changing from an encryption key server to a USB flash drive provider” on page 778. After this step, only USB flash drives encryption is configured, as shown in Figure 12-56.

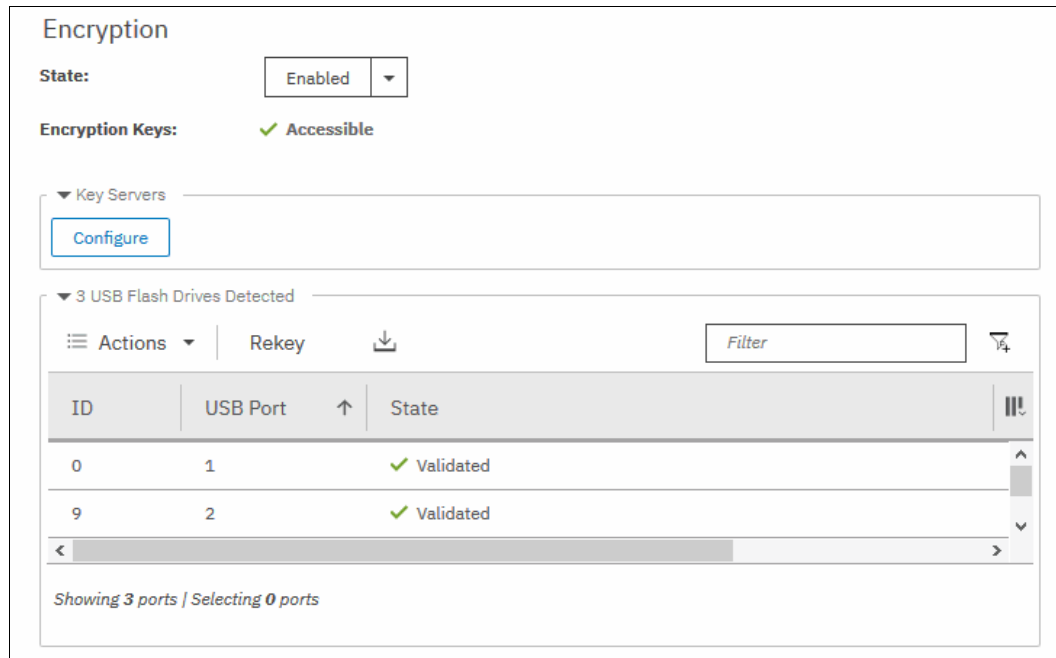


Figure 12-56 IBM FlashSystem encryption that is configured with USB flash drives

2. Migrate from USB flash drives encryption to another key server type encryption (in this example, Gemalto SafeNet KeySecure) by following the steps that are described in 12.7.1, “Changing from a USB flash drive provider to an encryption key server” on page 778. After completing this step, the other key server type is configured as an encryption provider in IBM Spectrum Virtualize, as shown in Figure 12-57.

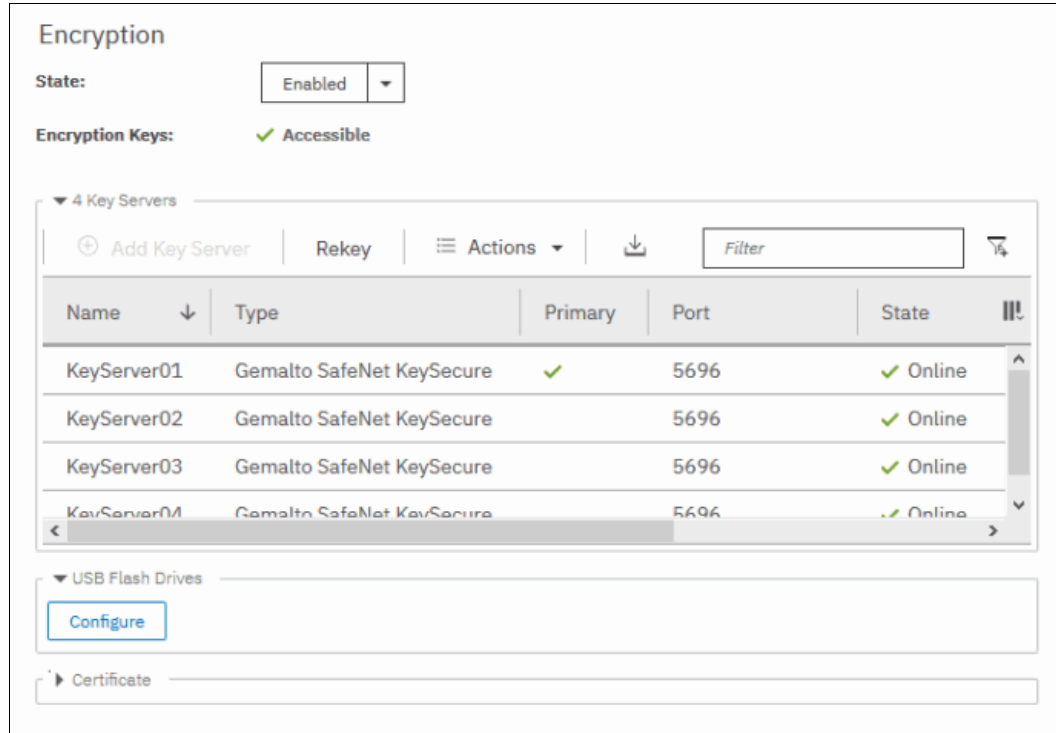


Figure 12-57 IBM FlashSystem encryption that is configured with SafeNet KeySecure

12.8 Recovering from a provider loss

If both encryption key providers are enabled and you lose one of them (by losing all copies of the encryption key that are kept on the USB flash drives or by losing all IBM Security Key Lifecycle Manager servers), you can recover from this situation by disabling the provider to which you lost the access. To disable the provider, you *must have access to a valid master access key* on the *remaining provider*.

If you lose access to the encryption key server provider, run the following command:

```
chencryption -keyserver disable
```

If you lose access to the USB flash drives provider, run the following command:

```
chencryption -usb disable
```

If you want to restore the configuration with both encryption key providers, follow the instructions that are described in 12.6, “Configuring more providers” on page 774.

Note: If you lose access to all encryption key providers that are defined in the system, no method is available to recover access to the data that is protected by the master access key.

12.9 Using encryption

The design for encryption is based on the concept that a system is fully encrypted or not encrypted. Encryption implementation is intended to encourage solutions that contain only encrypted volumes or only unencrypted volumes. For example, after encryption is enabled on the system, all new objects (for example, pools) are by default created as encrypted.

Some unsupported configurations are actively policed in code. For example, no support exists for creating unencrypted child pools from encrypted parent pools. However, exceptions exist:

- ▶ During the migration of volumes from unencrypted to encrypted volumes, a system might report both encrypted and unencrypted volumes.
- ▶ It is possible to create unencrypted arrays from CLI by manually overriding the default encryption setting.

Notes: Encryption support for distributed redundant array of independent disks (DRAID) is available in IBM Spectrum Virtualize V7.7 and later.

You must decide whether to encrypt or not encrypt an object when it is created. You cannot change this setting later. To change the encryption state of stored data, you must migrate from an encrypted object (for example, a pool) to an unencrypted one, or vice versa. Volume migration is the only way to encrypt any volumes that were created before enabling encryption on the system.

12.9.1 Encrypted pools

For more information about how to open the Create Pool window, see Chapter 5, “Storage pools”. After encryption is enabled, any new pool is created by default as encrypted, as shown in Figure 12-58.

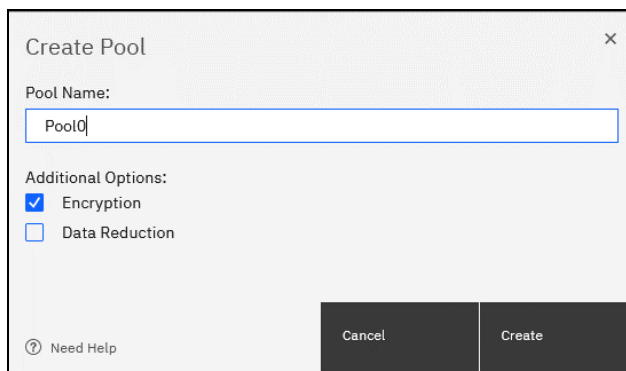


Figure 12-58 Create Pool window

You can click **Create** to create an encrypted pool. All storage that is added to this pool is encrypted.

You can customize the Pools view in the management GUI to show the pool encryption status. Select **Pools** → **Pools**, and then select **Actions** → **Customize Columns** → **Encryption**, as shown in Figure 12-59.

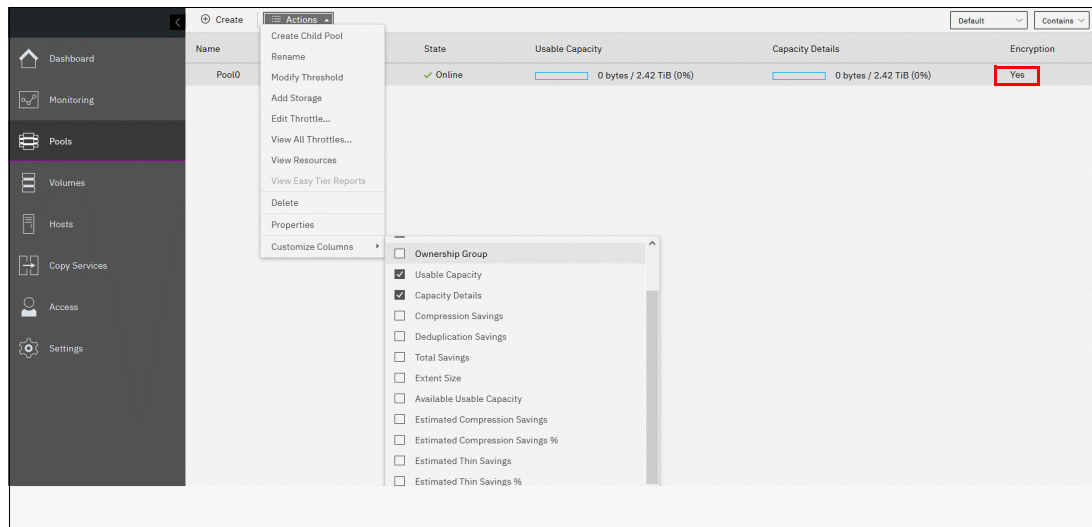


Figure 12-59 Pool encryption state

If you create an unencrypted pool but add only encrypted arrays or self-encrypting MDisks to the pool, the pool is reported as encrypted because all extents in the pool are encrypted. The pool reverts to the unencrypted state if you add an unencrypted array or MDisk. By default, if encryption is enabled on the storage, newly added internal MDisks (arrays) are created encrypted and the pool is reported as encrypted unless there is any unencrypted MDisks in the pool.

Important: Unencrypted pools allow encrypted and unencrypted MDisks to be added in one pool. If you remove all unencrypted MDisks from an unencrypted pool so that the pool contains only encrypted MDisks, the pool is reported as encrypted. Data is encrypted as MDisks are encrypted, but the pool still allows unencrypted MDisks to be added. Therefore, it is possible to mix the configuration if unencrypted MDisks are added, and the pool reverts to the unencrypted state. In this case, data ends up on an unencrypted MDisk. You must be cautious when encrypting previously unencrypted pools by adding encrypted MDisks, and removing unencrypted pools and expanding them later with new MDisks.

You can mix and match storage encryption types in a pool. Figure 12-60 on page 783 shows an example of an encrypted pool that contains storage by using different encryption methods.

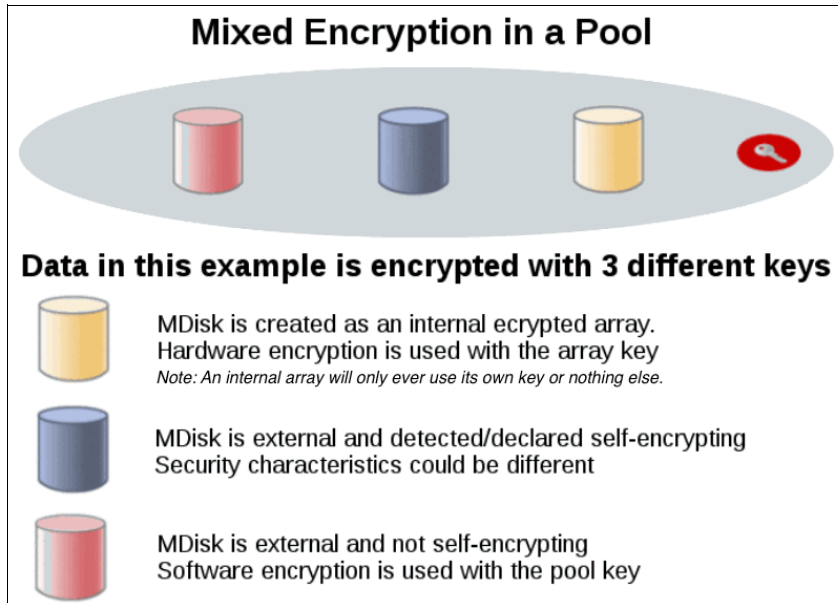


Figure 12-60 Mixing and matching encryption in a pool

12.9.2 Encrypted child pools

For more information about how to open the Create Child Pool window, see Chapter 5, “Storage pools” on page 237. If the parent pool is encrypted, every child pool also must be encrypted. The GUI enforces this requirement by automatically selecting **Encryption Enabled** in the Create Child Pool window and preventing changes to this setting, as shown in Figure 12-61.

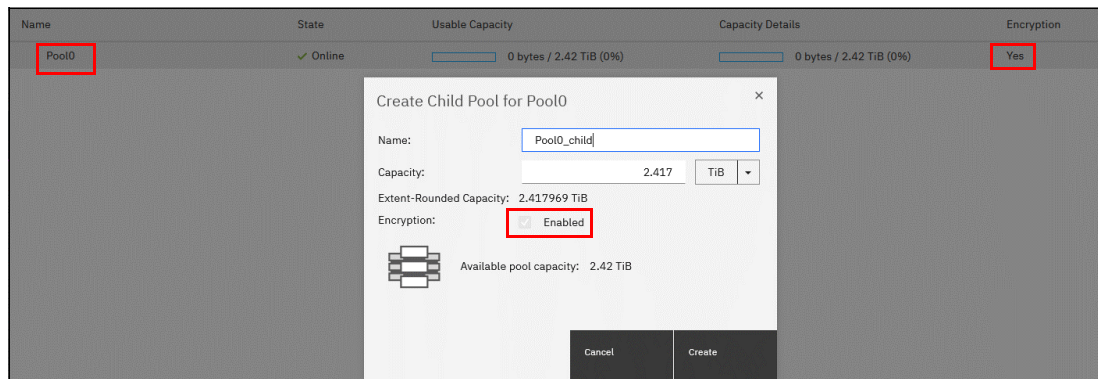


Figure 12-61 Creating a child pool of an encrypted parent pool

However, if you want to create encrypted child pools from an unencrypted storage pool that contains a mix of internal arrays and external MDisks, the following restrictions apply:

- ▶ The parent pool must not contain any unencrypted internal arrays. If any unencrypted internal arrays are in the unencrypted pool, when you try to create a child pool and select the option to set it as encrypted, it is created as unencrypted.
- ▶ All IBM FlashSystem Control Enclosures in the system must support software encryption and have the encryption license activated.

Note: An encrypted child pool that is created from an unencrypted parent storage pool reports as unencrypted if the parent pool contains any unencrypted internal arrays. Remove these arrays to ensure that the child pool is fully encrypted.

12.9.3 Encrypted arrays

For more information about how to add internal storage to a pool, see Chapter 5, “Storage pools” on page 237. After encryption is enabled, all newly built arrays are hardware encrypted by default. In this case, the GUI does not allow you to create an unencrypted array. To create an unencrypted array, the CLI must be used. Example 12-1 shows how to create an unencrypted array by using the CLI.

Example 12-1 Creating an unencrypted array by using the CLI with IBM FlashSystem

```
IBM_SAN:ITS0-V7k:superuser>svctask mkarray -drive 6:4 -level raid1 -sparegoal 0
-strip 256 -encrypt no Pool2
MDisk, id [2], successfully created
IBM_SAN:ITS0-V7k:superuser>
```

Note: It is not possible to add unencrypted arrays to an encrypted pool.

You can customize the MDisks by Pools view to show the array encryption status. Select **Pools** → **MDisk by Pools**, and then select **Actions** → **Customize Columns** → **Encryption**. You also can right-click the table header to customize columns and select **Encryption**, as shown in Figure 12-62.

The screenshot shows the IBM FlashSystem GUI interface. On the left, a navigation menu is open to 'Pools' > 'MDisk by Pools'. The main area displays a table of MDisk information. A red box highlights the 'Encryption' column header in the table.

Name	State	Usable Capacity	Written Capacity Limit	RAID	Encryption
Unassigned MDisk (0)					
Pool0	✓ Online	0 bytes / 4.84 TiB (0%)	2.42 TiB / 4.84 TiB (50%)		Yes
MDisk1	✓ Online	2.42 TiB	2.42 TiB	Distributed RAID 6	Yes
MDisk2	✓ Online	2.42 TiB	2.42 TiB	Distributed RAID 6	Yes
Volumes by Pool	✓ Online	0 bytes / 832.00 GiB (0%)	0 bytes / 832.00 GiB (0%)		No
Internal Storage	✓ Online	832.00 GiB	837.86 GiB	RAID 1	No
External Storage					
MDisks by Pools					
System Migration					

Figure 12-62 Array/MDisk view with the added Encryption column

You can also check the encryption state of an array by reviewing its drives by selecting **Pools** → **Internal Storage**. The internal drives that are associated with an encrypted array are assigned an encrypted property that you can view, as shown in Figure 12-63 on page 785.

Drive ID	Written C...	Use	Status	MDisk Name	Member ID	Enclosure ID	Slot ID	Encrypted
5	837.86 GiB	Member	✓ Online	MDisk1	0	2	6	✓
11	837.86 GiB	Member	✓ Online	MDisk1	1	2	22	✓
13	837.86 GiB	Member	✓ Online	MDisk1	2	2	11	✓
14	837.86 GiB	Member	✓ Online	MDisk1	3	2	4	✓
15	837.86 GiB	Member	✓ Online	MDisk1	4	2	9	✓
16	837.86 GiB	Member	✓ Online	MDisk1	5	2	2	✓
17	837.86 GiB	Member	✓ Online	MDisk2	0	2	17	✓
18	837.86 GiB	Member	✓ Online	MDisk2	1	2	8	✓
19	837.86 GiB	Member	✓ Online	MDisk2	2	2	24	✓
20	837.86 GiB	Member	✓ Online	MDisk2	3	2	5	✓
21	837.86 GiB	Member	✓ Online	MDisk2	4	2	21	✓
22	837.86 GiB	Member	✓ Online	MDisk2	5	2	7	✓

Figure 12-63 Drive encryption state

12.9.4 Encrypted MDisks

For more information about how to add external storage to a pool, see Chapter 5, “Storage pools” on page 237. Each MDisk that belongs to external storage that is added to an encrypted pool or child pool is automatically encrypted by using the pool or child pool key unless the MDisk is detected as self-encrypting.

The user interface gives no method to see which extents contain encrypted data and which do not. However, if a volume is created in a correctly configured encrypted pool, all data that is written to this volume is encrypted.

You can use the MDisk by Pools view to view the object encryption state by selecting **Pools** → **MDisk by Pools**. Figure 12-64 shows an example in which a self-encrypting MDisk is in an unencrypted pool, where the pool is reported as unencrypted.

Name	State	Usable Capacity	Written Capacity Limit	Encryption
Unassigned MDisks (0)				
Pool0	✓ Online	0 bytes / 4.84 TiB (0%)	2.42 TiB / 4.84 TiB (50%)	Yes
MDisk1	✓ Online	2.42 TiB	2.42 TiB	Yes
MDisk2	✓ Online	2.42 TiB	2.42 TiB	Yes
Pool1	✓ Online	0 bytes / 1.63 TiB (0%)	0 bytes / 1.63 TiB (0%)	No
mdisk0	✓ Online	832.00 GiB	837.86 GiB	No
mdisk1	✓ Online	837.00 GiB	837.86 GiB	Yes

Figure 12-64 MDisk encryption state

When working with MDisk encryption, take extra care when configuring the MDisks and pools.

If the MDisk was earlier used for storage of unencrypted data, the extents can contain stale unencrypted data. This issue occurs because file deletion marks disk space only as free. The data is *not* removed from the storage. Therefore, if the MDisk is not self-encrypting and was a part of an unencrypted pool and later was moved to an encrypted pool, the MDisk still contains stale data from its previous state.

Another mistake that can occur is misconfiguring an external MDisk as self-encrypting while in reality it is not self-encrypting. In that case, the data that is written to this MDisk is not encrypted by IBM FlashSystem because IBM FlashSystem expects that the storage system hosting the MDisk encrypts the data. Concurrently, the MDisk does not encrypt the data because it is not self-encrypting, so the system ends up with unencrypted data on an extent in an encrypted storage pool.

However, all data that is written to any MDisk that is a part of correctly configured encrypted storage pool is going to be encrypted.

Self-encrypting External MDisks

When adding external storage to a pool, be exceptionally diligent when declaring the MDisk as self-encrypting. Correctly declaring an MDisk as self-encrypting avoids wasting resources, such as CPU time. However, when used incorrectly, it might lead to unencrypted data-at-rest.

To declare an MDisk as self-encrypting, select **Externally encrypted** in the Assign Storage view when adding external storage, as shown in Figure 12-65.

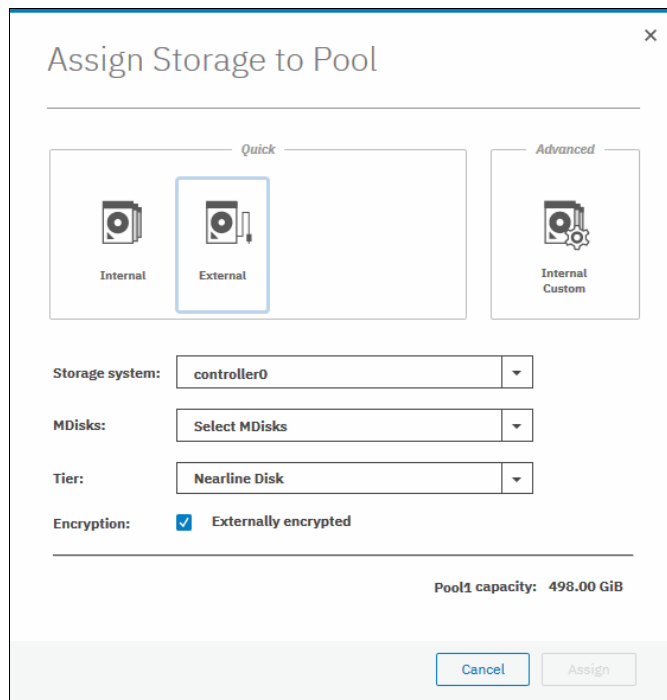


Figure 12-65 Declaring a MDisk as externally encrypted

IBM Spectrum Virtualize products can detect that an MDisk is self-encrypting by using the SCSI Inquiry page C2. MDisks that are provided by other IBM Spectrum Virtualize products report this page correctly. The **Externally encrypted** checkbox is selected for those MDisks.

Note: You can override the external encryption setting of a detected MDisk as self-encrypting and configure it as unencrypted by running `chmdisk -encrypt no`. However, run this command only if you plan to decrypt the data on the back end or if the back end uses inadequate data encryption.

To check whether an MDisk was declared as self-encrypting, select **Pools** → **MDisk by Pools** and verify the information in the Encryption column, as shown in Figure 12-66.

Name	State	Usable Capacity	Written Capacity Limit	Encryption
> Unassigned MDisks (5)				
> MigrationPool_1024	✓ Online		23.00 GiB / 23.00 GiB (100%)	Yes
> MigrationPool_4096	● Offline		0 bytes / 40.00 GiB (0%)	Yes
Pool1	✓ Online			Yes
mdisk3	✓ Online		1.00 GiB	Yes
mdisk7	✓ Online		1.00 GiB	No
> Pool2	● Offline	40.00 GiB / 1.59 TiB (2%)	1.02 TiB / 1.59 TiB (64%)	Yes

Figure 12-66 MDisk self-encryption state

The value that is shown in the Encryption column shows the property of objects in respective rows, which means that in the configuration that is shown in Figure 12-66, Pool1 is encrypted, so every volume that is created from this pool is encrypted. However, that pool is formed by two MDisks, out of which one is self-encrypting and one is not. Therefore, a value of No next to mdisk7 does not imply that the encryption of Pool1 is in any way compromised. It indicates that encryption of the data that is placed on mdisk7 is done only by using software encryption. Data that is placed on mdisk3 is encrypted by the back-end storage that is providing these MDisks.

Note: You can change the self-encrypting attribute of an MDisk that is unmanaged or a member of an unencrypted pool. However, you cannot change the self-encrypting attribute of an MDisk after it is added to an encrypted pool.

12.9.5 Encrypted volumes

For more information about how to create and manage volumes, see Chapter 6, “Volumes” on page 299. The encryption status of a volume depends on the pool encryption status. Volumes that are created in an encrypted pool are automatically encrypted.

You can modify the Volumes view to show whether the volume is encrypted. Select **Volumes** → **Volumes**, and then select **Actions** → **Customize Columns** → **Encryption** to customize the view to show the volume’s encryption status, as shown in Figure 12-67.

Name	State	Synchronized	Pool	UID	Capacity	Encryption
SVCVolume0	✓ Online		Pool0	600507640086031DD800000000000000	1.00 GiB	Yes
SVCVolume1	✓ Online		Pool0	600507640086031DD800000000000001	1.00 GiB	Yes
SVCVolume2	✓ Online		Pool0	600507640086031DD800000000000002	1.00 GiB	Yes
SVCVolume3	✓ Online		Pool0	600507640086031DD800000000000003	1.00 GiB	Yes
SVCUnencrypted0	✓ Online		Pool1	600507640086031DD800000000000004	1.00 GiB	No
SVCUnencrypted1	✓ Online		Pool1	600507640086031DD800000000000005	1.00 GiB	No
SVCUnencrypted2	✓ Online		Pool1	600507640086031DD800000000000006	1.00 GiB	No
UnencryptedVolumes0	✓ Online (formatting)		Pool1	600507640086031DD800000000000007	1.00 GiB	No
UnencryptedVolumes1	✓ Online (formatting)		Pool1	600507640086031DD800000000000008	1.00 GiB	No

Figure 12-67 Volume view customization

A volume is reported as encrypted only if all the volume copies are encrypted, as shown in Figure 12-68.

Name	State	Synchronized	Pool	UID	Capacity	Encryption
SVCVolume0	✓ Online		Pool0	600507640086031DD800000000000000	1.00 GiB	Yes
SVCVolume1	✓ Online		Pool0	600507640086031DD800000000000001	1.00 GiB	Yes
SVCVolume2	✓ Online		Pool0	600507640086031DD800000000000002	1.00 GiB	Yes
✓ SVCVolume3	✓ Online		Pool0	600507640086031DD800000000000003	1.00 GiB	Yes
Copy 0*	✓ Online	Yes	Pool0	600507640086031DD800000000000003	1.00 GiB	Yes
Copy 1	✓ Online	No	Pool0	600507640086031DD800000000000003	1.00 GiB	Yes
✓ SVCUnencrypted0	✓ Online		Pool1	600507640086031DD800000000000004	1.00 GiB	No
Copy 0*	✓ Online	Yes	Pool1	600507640086031DD800000000000004	1.00 GiB	No
Copy 1	✓ Online	No	Pool0	600507640086031DD800000000000004	1.00 GiB	Yes

Figure 12-68 Volume encryption status depending on volume copies encryption

When creating volumes, make sure to select encrypted pools to create encrypted volumes, as shown in Figure 12-69.

Figure 12-69 Creating an encrypted volume by selecting an encrypted pool

You cannot change an existing unencrypted volume to an encrypted version of itself dynamically. However, this conversion is possible by using two migration options:

- ▶ Migrate a volume to an encrypted pool or child pool.
- ▶ Mirror a volume to an encrypted pool or child pool and delete the unencrypted copy.

For more information about these methods, see Chapter 6, “Volumes” on page 299.

12.9.6 Restrictions

The following restrictions apply to encryption:

- ▶ Image mode volumes cannot be in encrypted pools.
- ▶ You cannot add external non-self-encrypting MDisks to encrypted pools unless all control enclosures in the system support encryption.

12.10 Rekeying an encryption-enabled system

Changing the master access key is a security requirement. *Rekeying* is the process of replacing the current master access key with a newly generated one. The rekey operation works whether encrypted objects exist. The rekeying operation requires access to a valid copy of the original master access key on an encryption key provider that you plan to rekey. Use the rekey operation according to the schedule that is defined in your organization's security policy and whenever you suspect that the key might be compromised.

If you have both USB and key servers that are enabled, rekeying is done separately for each of the providers.

Important: Before you create a master access key, ensure that all nodes are online and that the current master access key is accessible.

No method is available to directly change data encryption keys. If you must change the data encryption key that is used to encrypt data, the only available method is to migrate that data to a new encrypted object (for example, an encrypted child pool). Because the data encryption keys are defined per encrypted object, such migration forces a change of the key that is used to encrypt that data.

12.10.1 Rekeying by using a key server

Ensure that all the configured key servers can be reached by the system and that service IP addresses are configured on all your nodes.

To rekey the master access key that is kept on the key server provider, complete the following steps:

1. Select **Settings** → **Security** → **Encryption**. Ensure that **Encryption Keys** shows that all configured IBM Security Key Lifecycle Manager servers are reported as **Accessible**. Click **Key Servers** to expand the section.
2. Click **Rekey**, as shown in Figure 12-70.

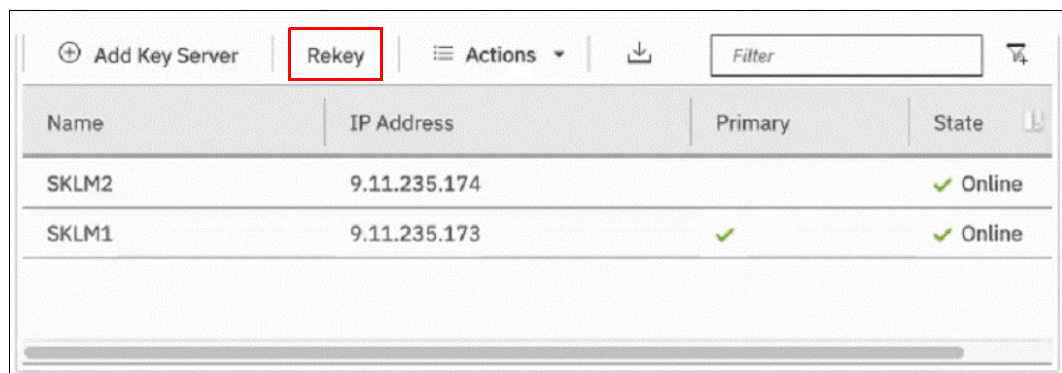


Figure 12-70 Starting the rekey on the IBM Security Key Lifecycle Manager key server

3. In the next window, confirm the rekey operation.

Note: The rekey operation is performed on only the primary key server that is configured in the system. If more key servers are configured apart from the primary key, they do not hold the updated encryption key until they obtain it from the primary key server. To restore encryption key provider redundancy after a rekey operation, replicate the encryption key from the primary key server to the secondary key servers.

You receive a message confirming that the rekey operation was successful.

12.10.2 Rekeying by using USB flash drives

During the rekey process, new keys are generated and copied to the USB flash drives. These keys are then used instead of the current keys. The rekey operation fails if at least one of the USB flash drives does not contain the current key. To rekey the system, you need at least three USB flash drives to store the master access key copies.

After the rekey operation is complete, update all other copies of the encryption key, including copies that are stored on other media. Take the same precautions to securely store all copies of the new encryption key as when you enabled encryption for the first time.

To rekey the master access key on USB flash drives, complete the following steps:

1. Select **Settings** → **Security** → **Encryption**. Click **USB Flash Drives** to expand the section.
2. Verify that all USB drives that are plugged into the system are detected and show as **Validated**, as shown in Figure 12-71. Click **Rekey**. You need at least three USB flash drives, with at least one reported as **Validated** to process a rekey.

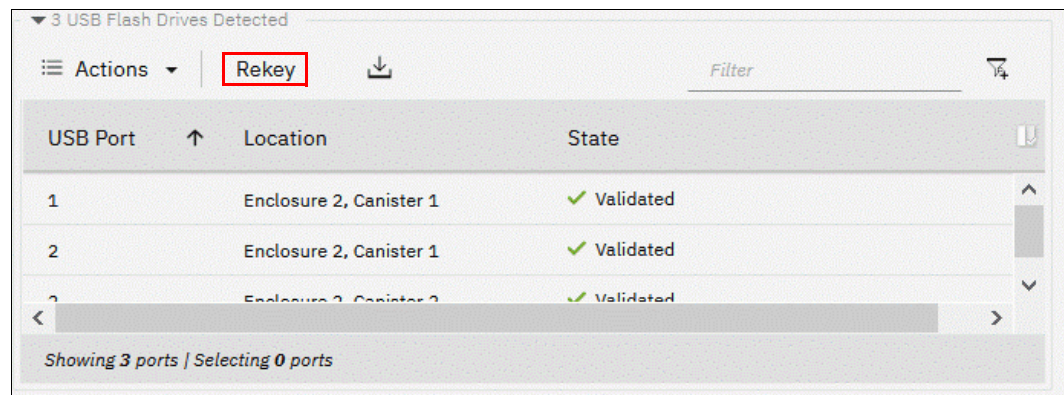


Figure 12-71 Starting a rekey on a USB flash drives provider

3. If the system detects a validated USB flash drive and at least three available USB flash drives, new encryption keys are automatically copied on the USB flash drives, as shown in Figure 12-72 on page 791. Click **Commit** to finalize the rekey operation.

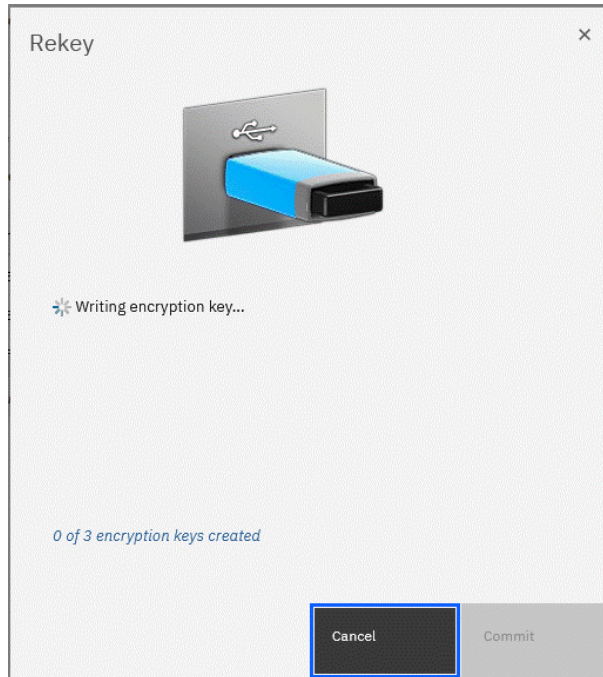


Figure 12-72 Writing new keys to USB flash drives

4. You receive a message confirming that the rekey operation was successful, as shown in Figure 12-73. Click **Close**.

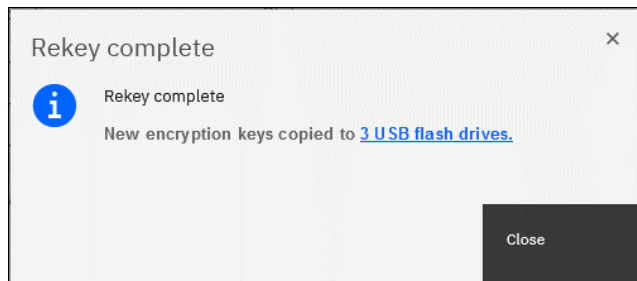


Figure 12-73 Successful rekey operation by using USB flash drives

12.11 Disabling encryption

You are prevented from disabling encryption if any encrypted objects are defined apart from self-encrypting MDisk. You can disable encryption in the same way whether you use USB flash drives, a key server, or both providers.

To disable encryption, complete the following steps:

1. Select **Settings** → **Security** → **Encryption**, and click **Enabled**. If no encrypted objects exist, a menu is displayed. Click **Disabled** to disable encryption on the system. Figure 12-74 shows an example for a system with both encryption key providers configured.

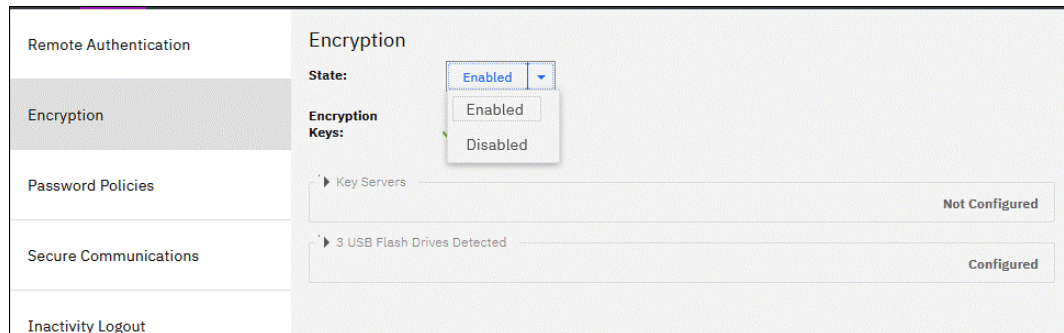


Figure 12-74 Disabling encryption on a system with both providers

2. You receive a message confirming that encryption was disabled. Figure 12-75 shows the message when a key server is used.

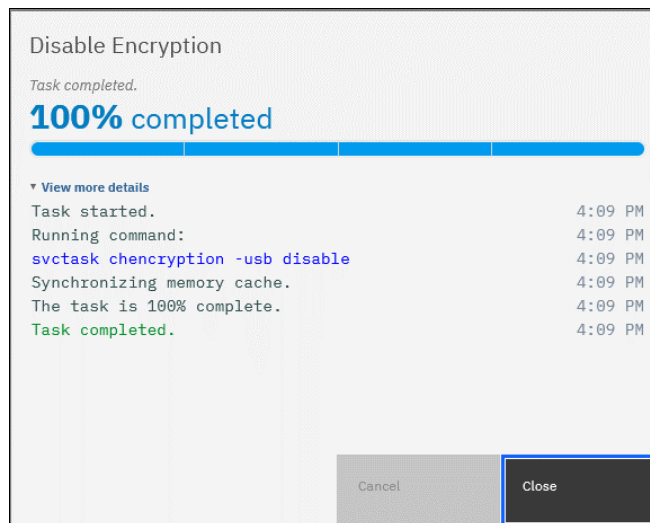


Figure 12-75 Encryption disabled



Reliability, availability, and serviceability, monitoring and logging, and troubleshooting

This chapter introduces useful and common procedures to maintain the system. It includes the following topics:

- ▶ 13.1, “Reliability, availability, and serviceability” on page 794
- ▶ 13.2, “Shutting down the IBM FlashSystem” on page 805
- ▶ 13.3, “Removing or adding a node from or to the system” on page 805
- ▶ 13.4, “Configuration backup” on page 808
- ▶ 13.5, “Software update” on page 812
- ▶ 13.6, “Health checker feature” on page 827
- ▶ 13.7, “Troubleshooting and fix procedures” on page 828
- ▶ 13.8, “Monitoring” on page 834
- ▶ 13.9, “Audit log” on page 852
- ▶ 13.10, “Collecting support information by using the GUI, CLI, and USB” on page 855
- ▶ 13.11, “Service Assistant Tool” on page 862
- ▶ 13.12, “IBM Storage Insights monitoring” on page 865

13.1 Reliability, availability, and serviceability

Reliability, availability, and serviceability (RAS) are important concepts in the design of the IBM Spectrum Virtualize system. Hardware features, software features, design considerations, and operational guidelines all contribute to make the IBM FlashSystem systems reliable.

Fault tolerance and high levels of availability are achieved by using the following methods:

- ▶ The distributed redundant array of independent disks (DRAID) capabilities of the underlying disks.
- ▶ IBM FlashSystem nodes clustering that use a *Compass* architecture.
- ▶ Auto-restart of hung nodes.
- ▶ Integrated battery backup units (BBUs) to provide memory protection if a site power failure occurs.
- ▶ Host system failover capabilities by using N_Port ID Virtualization (NPIV).
- ▶ Deploying advanced multi-site configurations, such as IBM HyperSwap and stretched clusters.

High levels of serviceability are available through the following methods:

- ▶ Cluster error logging
- ▶ Asynchronous error notification
- ▶ Automatic dump capabilities to capture software-detected issues
- ▶ Concurrent diagnostic procedures
- ▶ Directed maintenance procedures (DMPs) with guided online replacement processes
- ▶ Concurrent log analysis and memory dump data recovery tools
- ▶ Concurrent maintenance of IBM FlashSystem components
- ▶ Concurrent upgrade of IBM Spectrum Virtualize Software and firmware
- ▶ Concurrent addition or deletion of node canisters in the clustered system
- ▶ Automatic software version leveling when replacing a node
- ▶ Detailed status and error conditions that are displayed by light-emitting diode (LED) indicators
- ▶ Error and event notification through Simple Network Management Protocol (SNMP), syslog, and email
- ▶ Optional Remote Support Assistant

The heart of the IBM FlashSystem system is a pair of *node canisters*. These two canisters share the read and write data workload from the attached hosts and to the disk arrays. This section examines the RAS features of the systems, monitoring, and troubleshooting.

13.1.1 Node canisters

Two node canisters are contained in the control enclosure that work as a clustered system that runs the IBM Spectrum Virtualize Software. As shown in Figure 13-1, the top node canister is inverted above the bottom one. The control enclosure also contains two power supply units (PSUs) that operate independently of each other. The PSUs are visible from the back of the control enclosure.

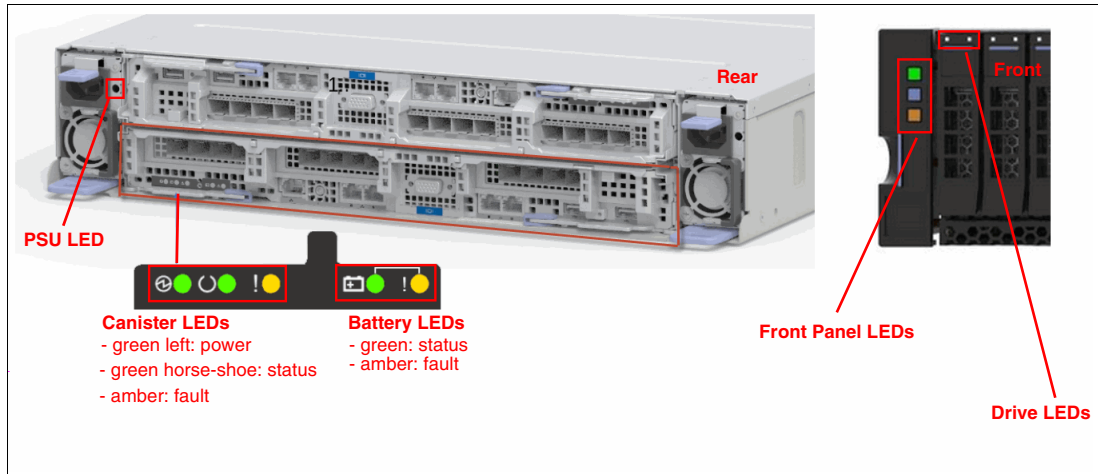


Figure 13-1 LEDs on each node canister

The connections of a single node canister (bottom) are shown in Figure 13-2.

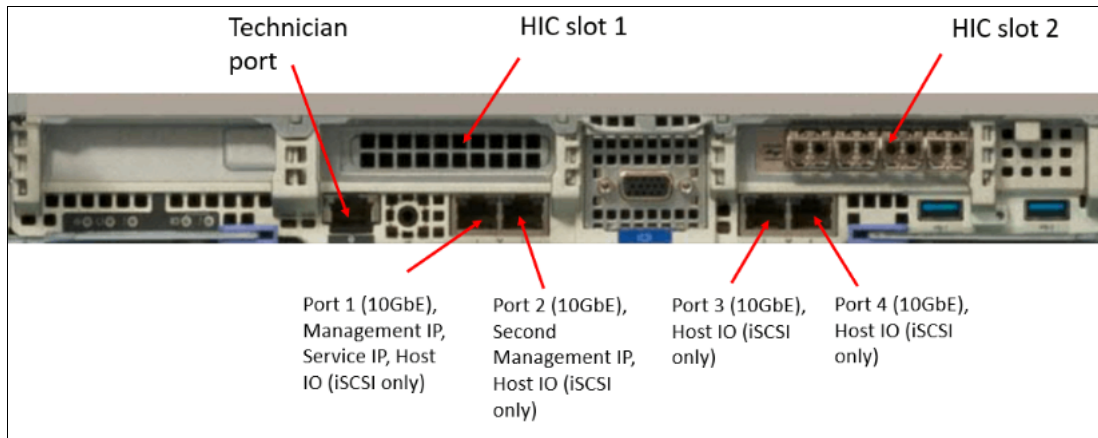


Figure 13-2 Typical node canister connections

Host interface cards

Each canister (apart from the IBM FlashSystem 5100) has three host interface card (HIC) slots. Depending on the system, there might already be a 4-port serial-attached Small Computer System Interface (SCSI) (SAS) card that is installed in each node, leaving two HIC slots that can be populated with a range of cards, as shown in Table 13-1. Nodes in the same I/O group must have the same HIC configuration.

Table 13-1 Supported card configurations for IBM FlashSystem 7xxx / IBM FlashSystem 9xxx systems

Supported number of cards	Ports	Protocol	Slot positions	Note
0 - 3	4	16 Gb Fibre Channel (FC)	1, 2, 3	
0 - 3	4	32 Gb FC	1, 2, 3	Works with a 16 Gb switch.
0 - 3	2	25 Gb Ethernet (GbE) (internet Wide Area Remote Direct Memory Access (RDMA) Protocol (iWARP))	1, 2, 3	
0 - 3	2	25 GbE (RDMA over Converged Ethernet (RoCE))	1, 2, 3	
0 - 1	2	12 Gb SAS Expansion	3	<ul style="list-style-type: none"> ▶ Four-port card, but only two are active. ▶ Expansion only, no SAS host attachment.

Note: The systems have onboard compression cards. There are no compression-assist cards like in previous models.

For V5100, there are only two card slots, and the following card configurations are supported (Table 13-2).

Table 13-2 Supported card configurations for IBM FlashSystem 5100 / IBM FlashSystem 5xxx systems

Supported number of cards	Ports	Protocol	Slot positions	Note
0 - 1	4	16 Gb FC	2	
0 - 1	4	32 Gb FC	2	Works with a 16 Gb switch.
0 - 1	2	25 GbE (iWARP)	2	
0 - 1	2	25 GbE (RoCE)	2	
0 - 1	2	12 Gb SAS Expansion	1	<ul style="list-style-type: none"> ▶ Four-port card, but only two are active. ▶ Expansion only, no SAS host attachment.

Note: For IBM FlashSystem 5100 and IBM FlashSystem 5000 systems, Peripheral Component Interconnect Express (PCIe) slot 1 has a blanking plate, so this slot cannot be used, and slots 2 and 3 become slots 1 and 2. The fabric attach card goes only in slot 2 (far right when the canister is in the lower canister) so that you can better use the direct connection of slot 2 to the CPU. Slot 1 (middle position) is connected through the PCIe switch and accepts only the optional (and slower) SAS card.

The FC card is required to add other control enclosures to the system (0 - 2). Using an FC card, you can connect the IBM FlashSystem 9xxx or 7xxx control enclosure to up to three more systems (for a maximum of eight nodes). For the IBM FlashSystem 5100 system, you can connect only one extra control enclosure (for a maximum of four nodes). For FC configurations, the meaning of the port LEDs is explained in Table 13-3.

Table 13-3 Fibre Channel link LED statuses

Port LED	Color	Meaning
Link status	Green	Link is up, connection established.
Speed	Amber	Link is not up or speed fault.

USB ports

Two active USB connectors are available in the horizontal position to the right of the node. They have no numbers, and no indicators are associated with them. These ports can be used for initial cluster setup, encryption key backup, and node status or log collection.

Ethernet and LED status

Four 10 GbE ports and one 1-Gigabit Ethernet port are on each canister. However, not all ports are equal, and their functions are described in Table 13-4. Figure 13-2 on page 795 shows the location of the technician port on a node canister.

Table 13-4 Ethernet ports and their functions

Onboard Ethernet port	Speed	Function
1	10 GbE	Management IP, Service IP, and Host I/O (internet Small Computer Systems Interface (iSCS) only)
2	10 GbE	Secondary Management IP, and Host I/O (internet Small Computer Systems Interface (iSCSI) only)
3	10 GbE	Host I/O (iSCSI only)
4	10 GbE	Host I/O (iSCSI only)
T	1 GbE	Technician Port: DHCP / domain name server (DNS) for direct attach service management

Each port has two LEDs, and their status values are listed in Table 13-5. However, the T port is strictly dedicated to technician actions (initial and emergency configuration by local support personnel).

Table 13-5 Ethernet LED statuses

LED	Color	Meaning
Link state	Green	It is on when there is an Ethernet link.
Activity	Amber	It is flashing when there is activity on the link.

Serial-attached SCSI ports

When a 4-port SAS interface card is installed, it is possible to connect the 2U and 5U expansion enclosures. However, only ports 1 and 3 are used for SAS connections, with the SAS chain from port 1 installed below the lower node canister, and the SAS chain from port 3 installed above the upper node canister, as shown in Table 13-6. The SAS card must be installed in PCIe slot 3 of each node canister. There are two LEDs for each SAS port with statuses, as shown in Table 13-6.

Table 13-6 SAS LED statuses

LED	Meaning
Green	Link is connected and up.
Orange	Fault on the SAS link (disconnected, wrong speed, and errors).

Node canister status LEDs

There are three LEDs in a row at the left of the canister that indicates the status and the functions of the node (see Table 13-7).

Table 13-7 Node canister LEDs

Position	Color	Name	State	Meaning
Left	Green	Power	On	The node is started and active. It might not be safe to remove the canister. If the fault LED is off, the node is an active member of a cluster or candidate. If the fault LED is also on, the node is in the service state or in error, which prevents the software to start.
			Flashing (2 Hz)	Canister is started and in standby mode.
			Flashing (4 Hz)	Node is running a power-on self-test (POST).
			Off	No power to the canister or it is running on battery.
Middle	Green	Status	On	The node is a member of a cluster.
			Flashing (2 Hz)	The node is a candidate for or in a service state.
			Flashing (4 Hz)	The node is performing a fire hose dump. Never unplug the canister during this time.
			Off	No power to the canister or the canister is in standby mode.

Position	Color	Name	State	Meaning
Right	Amber	Fault	On	The canister is in a service state or in error, for example, a POST error that is preventing the software from starting.
			Flashing (2 Hz)	Canister is being identified.
			Off	Node is either in the candidate or active state.

Battery LEDs

Immediately to the right of the canister LEDs, with a short gap between them, are the Battery LEDs, which provide the status of the battery (see Table 13-8).

Table 13-8 Battery LEDs

Position	Color	Name	State	Meaning
Left	Green	Status	On	Indicates that the battery is fully charged and has sufficient charge to complete two fire hose dumps.
			Flashing (2 Hz)	Indicates that the battery has sufficient charge to complete a single fire hose dump.
			Flashing (4 Hz)	Indicates that the battery is charging and has insufficient charge to complete a single fire hose dump.
			Off	Indicates that the battery is not available for use (for example, it is missing or contains a fault).
Right	Amber	Fault	On	Indicates that a battery has a fault or a condition occurred. The node enters the service state.
			Off	Indicates that there are no known battery faults or conditions. An exception is when a battery has insufficient charge to complete a single fire hose dump. Refer to the Status LED.

13.1.2 Expansion canisters

As Figure 13-3 shows, two 12 gigabits per second (Gbps) SAS ports are side by side on the canister of every enclosure. They are numbered 1 on the left and 2 on the right. Like the controller canisters, expansion canisters are also installed in the enclosure side by side in a vertical position.

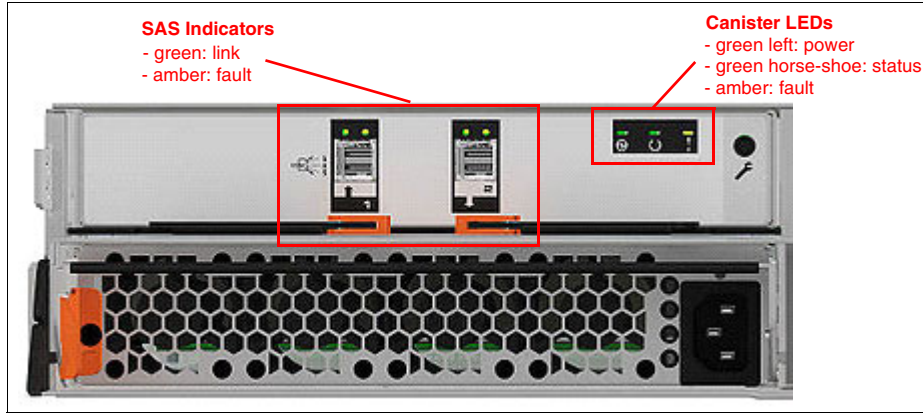


Figure 13-3 Expansion canister status LEDs

The interpretation of the SAS status LED indicators has the same meaning as the LED indicators of SAS ports in the control enclosure (Table 13-6 on page 798).

Table 13-9 lists the LED status values of the expansion canister.

Table 13-9 Expansion canister LEDs statuses

Position	Color	Name	State	Meaning
Left	Green	Power	On	The canister is powered on.
			Off	No power is available to the canister.
Middle	Green	Status	On	The canister is operating normally.
			Flashing	There is an error with the vital product data (VPD).
Right	Amber	Fault	On	There is an error that is logged against the canister or the system is not running.
			Flashing	Canister is being identified.
			Off	No fault, canister is operating normally.

13.1.3 Dense Drawer Enclosures LED

As Figure 13-4 on page 801 shows, two 12 Gbps SAS ports are side by side on the canister of every enclosure. They are numbered 1 on the right and 2 on the left. Each Dense Drawer has two canisters side by side, although they are inverted when compared to the 2U enclosures.

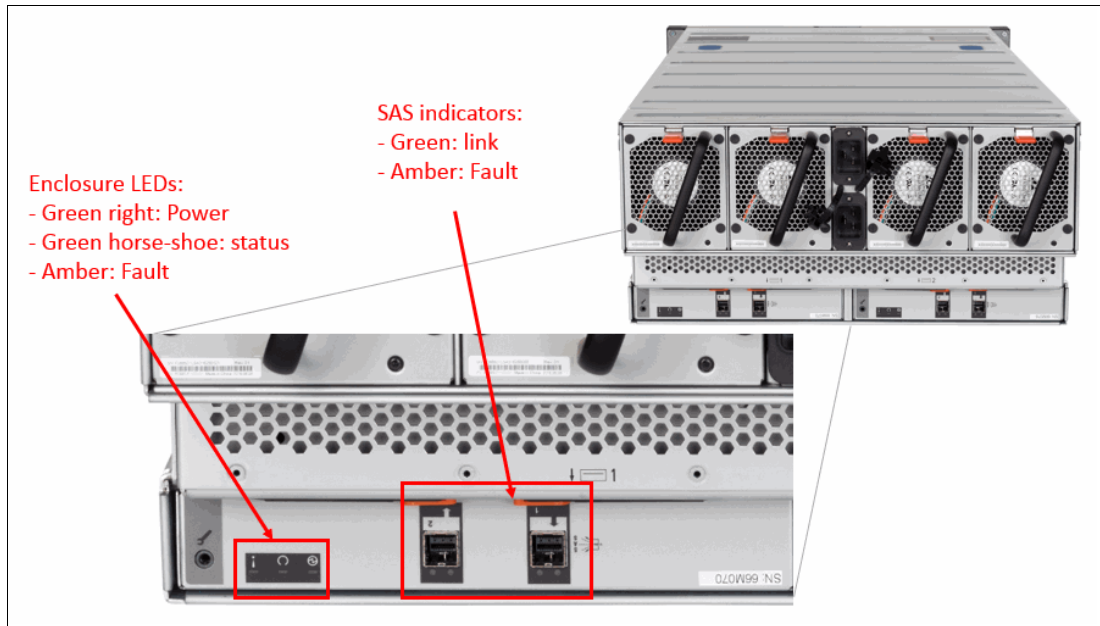


Figure 13-4 Dense Drawer LEDs

The interpretation of SAS status LED indicators has the same meaning as the LED indicators of SAS ports that are mentioned in the previous section (see Table 13-9 on page 800).

Table 13-10 shows the LED status values of the expansion canister.

Table 13-10 Expansion canister LEDs statuses

Position	Color	Name	State	Meaning
Right	Green	Power	On	The canister is powered on.
			Off	No power is available to the canister.
Middle	Green	Status	On	The canister is operating normally.
			Flashing	There is an error with the VPD.
Left	Amber	Fault	On	There is an error that is logged against the canister or the system is not running (OSES).
			Flashing	Canister is being identified.
			Off	No fault, canister is operating normally.

13.1.4 Enclosure SAS cabling

Expansion enclosures are attached to control enclosures through 12 Gbps SAS cables. The IBM FlashSystem control enclosure attaches up to 20 expansion enclosures or up to eight Dense Drawer enclosures.

A *strand* starts with an SAS initiator chip inside an IBM FlashSystem node canister and progresses through SAS expanders, which connect disk drives. Each canister contains an expander. Each drive has two ports, each connected to a different expander and strand. This configuration ensures that both nodes in the input/output (I/O) group have direct access to each drive, and that no single point of failure (SPOF) exists.

Figure 13-5 shows how the SAS connectivity works inside the node and expansion canisters.

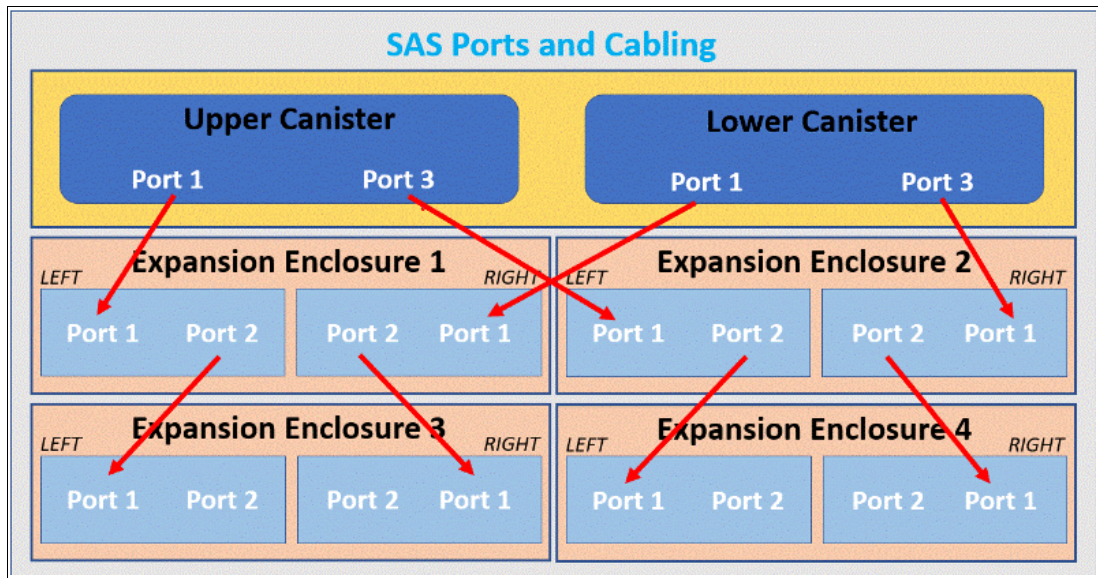


Figure 13-5 Concept of SAS chaining

Note: The last expansion enclosure in a chain must not have cables in port 2 of canister 1 or port 2 of canister 2. So, if you add another two enclosures to the setup that is shown in Figure 13-5, you connect a cable to port 2 of the existing enclosure canisters and port 1 of the new enclosure canisters.

A *chain* consists of a set of enclosures that are correctly interconnected (Figure 13-6 on page 803). Chain 1 of an I/O group is connected to SAS port 1 of both node canisters. Chain 2 is connected to SAS port 3. This configuration means that chain 2 includes the SAS expander and drives of the control enclosure.

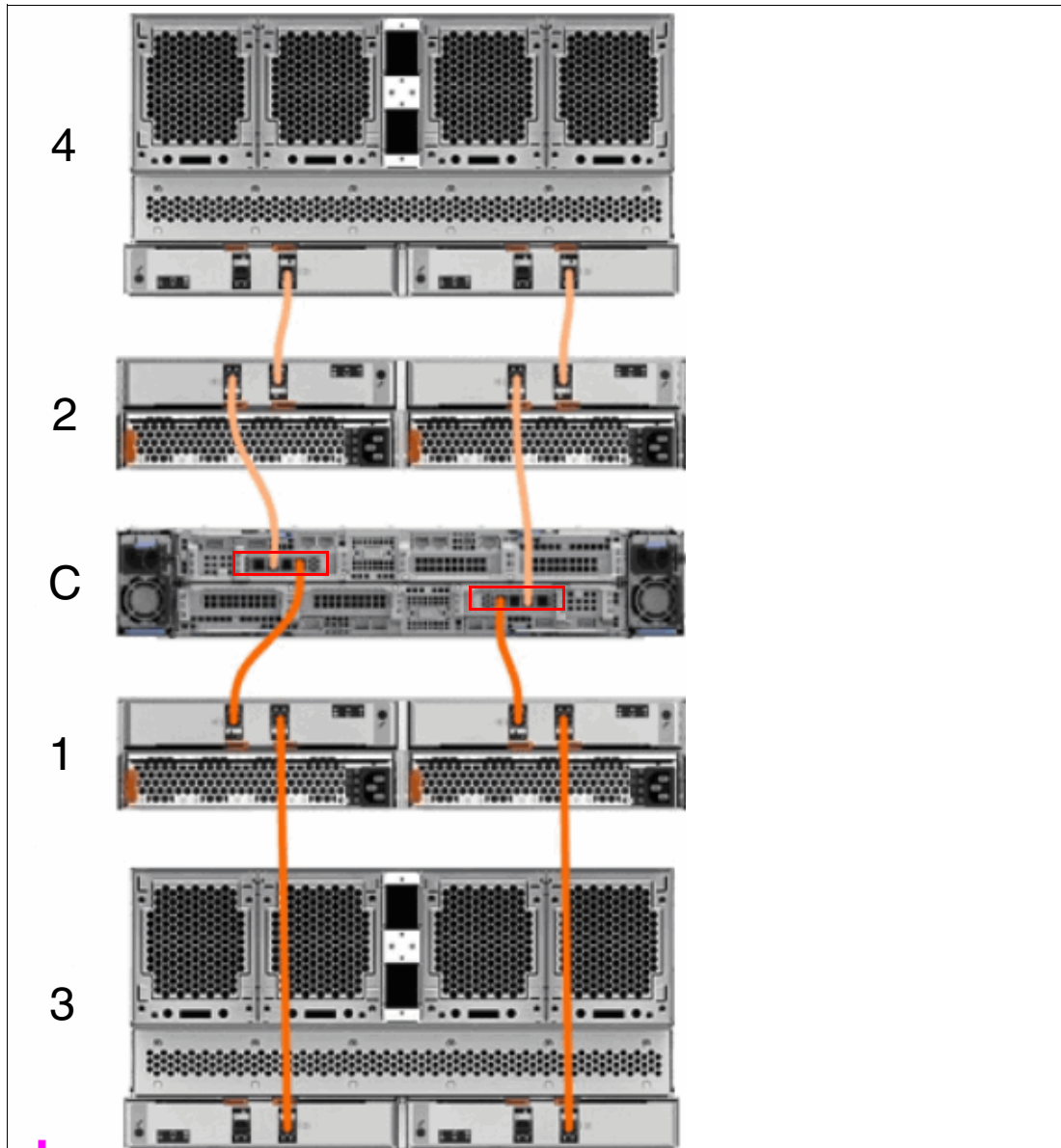


Figure 13-6 SAS cabling with numbered enclosures

At system initialization, when devices are added to or removed from strands, the system performs a discovery process to update the state of the drive and enclosure objects.

13.1.5 IBM FlashCore Module drives

There are two main considerations for RAS when talking about the new IBM FlashCore Module (FCM) drives:

- ▶ *Never reseal an FCM.* because when it is resealed, it performs an automatic reformat, which means all data on that drive can be lost.
- ▶ When removing an array, FCM drives might show as offline for some time due to formatting. They automatically come back online after the task finishes.

13.1.6 Power

All enclosures accommodate two PSUs for normal operation. A single PSU can supply the entire enclosure for redundancy. For this reason, it is highly advised to supply AC power to each PSU from different power distribution units (PDUs).

There is a power switch on the power supply and indicator LEDs. The switch must be on for the PSU to be operational. If the power switch is turned off, the PSU stops providing power to the system.

For control enclosure PSUs, the battery that is integrated in the node canister continues to supply power to the node. It supports the power outage for 5 seconds before initiating safety procedures. A fully charged battery can perform two fire hose dumps. A *fire hose dump* is a process where a node stores cache and system data to an internal drive in the event of a power failure.

Figure 13-7 shows two PSUs that are present in the control and expansion enclosure. The controller PSU has one LED that can be green or amber, depending on the status of the PSU. If the LED is off, that means there is no AC power to the entire enclosure.

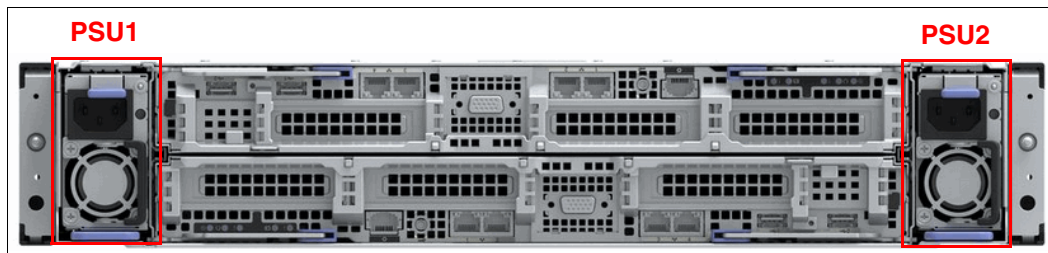


Figure 13-7 Controller and expansion enclosure LED status indicator

Figure 13-8 presents the rear overview of the enclosure canister with a PSU. The enclosure is powered on by the direct attachment of a power cable.



Figure 13-8 Expansion enclosure power supply unit

Power supplies in both control and expansion enclosures are hot-swappable and replaceable without needing to shut down a node or cluster. If the power is interrupted in one node for less than 5 seconds, the canister does not perform a fire hose dump and continues operation from the battery. This feature is useful for a case of, for example, maintenance of UPS systems in the data center or replugging the power to a different power source or PDU unit. A fully charged battery can perform two fire hose dumps.

13.2 Shutting down the IBM FlashSystem

You can safely shut down the system by using the GUI or command-line interface (CLI).

Important: Never shut down your system by powering off the PSUs, removing both PSUs, or removing both power cables from a running system. These actions can lead to inconsistency or loss of the data that is staged in the cache.

Before shutting down the IBM FlashSystem system, stop all hosts that allocated volumes from the device. This step can be skipped for hosts that have volumes that are also provisioned with mirroring (host-based mirroring) from different storage devices. However, doing so incurs errors that are related to lost storage paths and disks on the host error log.

You can shut down a single node canister, or you can shut down the entire cluster. When you shut down only one node canister, all activities remain active. When you shut down a canister or the entire cluster, you must power on locally to start the canister or system.

13.2.1 Shutting down and powering on a complete infrastructure

When you shut down or power on the entire infrastructure (storage, servers, and applications), follow a particular sequence for both the shutdown and the power-on actions. Next, we describe an example sequence of a shutdown, and then a power-on of an infrastructure that includes an IBM FlashSystem system.

Shutting down

To shut down the infrastructure, complete the following steps:

1. Shut down your servers and all applications.
2. Shut down your IBM FlashSystem systems:
 - a. Shut down the IBM FlashSystem by using the GUI or CLI.
 - b. Power off both switches of the controller enclosure.
 - c. Power off both switches of all the expansion enclosures.
3. Shut down your storage area network (SAN) switches.

Powering on

To power on your infrastructure, complete the following steps:

1. Power on your SAN switches and wait until the start completes.
2. Power on your storage systems by completing the following steps:
 - a. Power on both power supplies of all the expansion enclosures.
 - b. Power on both power supplies of the control enclosure.
 - c. When the storage systems are up, power on your servers and start your applications.

13.3 Removing or adding a node from or to the system

There are situations where IBM Support might prompt you to remove a node from the system briefly. One typical use case is when a node becomes stuck during a code upgrade. You can remove the node from the cluster briefly to commit the upgrade and complete (or cancel) the procedure depending on how many nodes upgraded so far. This procedure should be done only under the direction of IBM Support.

The easiest way to do this task is running `svcinfolnode` to display all nodes and their ID and status, as shown in Example 13-1. You can make sure that each IOgroup has two nodes online (or that if you remove a node, that one node remains in the IOgroup to continue serving I/O).

Example 13-1 The `lsnode` output

```
IBM FlashSystem 7200:superuser>svcinfolnode
id name UPS_serial_number WWNN status IO_group_id IO_group_name config_node UPS_unique_id hardware iscsi_name
iscsi_alias panel_name enclosure_id canister_id enclosure_serial_number site_id site_name
1 node1 500507680B00E6C5 online 0 io_grp0 yes 600
iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node1 Canister 1 02-2 2 2 7825WKP
2 node2 500507680B00E6C4 online 0 io_grp0 no 600
iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node2 Canister 2 02-1 2 1 7825WKP
IBM FlashSystem 7200:superuser>
```

In this example, we remove node 1 from the cluster. Run the `svctask rmnode 1` command, as shown in Example 13-2.

Example 13-2 The `rmnode` command

```
IBM FlashSystem 7200:superuser>>svctask rmnode 1
IBM FlashSystem 7200:superuser>>
```

A node can also be removed by using the GUI. Complete the following steps:

1. Select **Monitoring** → **System**, and then select the relevant control enclosure that the node you want to remove is on, which opens the Enclosure Details window. Select the node and either right-click it and click **Remove**, or use the menu in the Components Details to remove it, as shown in Figure 13-9, which opens a confirmation window.

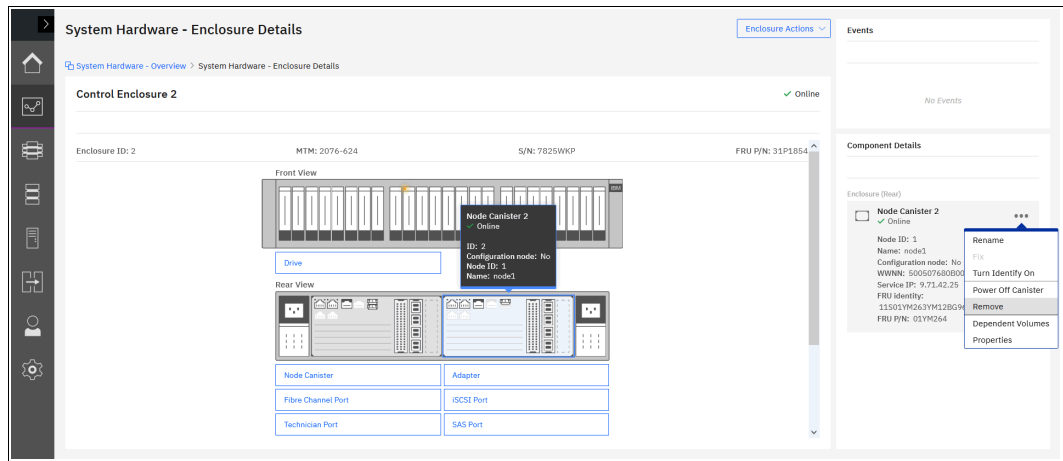


Figure 13-9 Removing a node by using the GUI

After you remove the node, if you rerun `svcinfolnode`, you see that it disappeared from the cluster, as shown in Example 13-3. The Service Assistant Tool (SAT) and GUI also reflect that there is now only one node in the cluster.

Example 13-3 The `lsnode` output after removing a node

```
IBM FlashSystem 7200:superuser>svcinfolnode
id name UPS_serial_number WWNN status IO_group_id IO_group_name config_node UPS_unique_id hardware iscsi_name
iscsi_alias panel_name enclosure_id canister_id enclosure_serial_number site_id site_name
1 node1 500507680B00E6C5 online 0 io_grp0 yes 600
iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node1 Canister 1 02-2 2 2 7825WKP
IBM FlashSystem 7200:superuser>
```

Note: By default, the cache is flushed before the node is deleted to prevent data loss if a failure occurs on the other node in the I/O group. This flush incurs a delay after you remove a node to when it comes back up as candidate status.

2. After a brief period, check the SAT, which shows that the node that you removed is in the service or candidate status, as shown in Figure 13-10.

The screenshot shows the Service Assistant Tool interface. The current node is 'node1' with status 'Active'. The interface includes a navigation menu on the left and a main content area with several sections:

- Home:** Provides detailed status and error summary for the current node.
- Change Node:** A table showing node details.

Node Name	Node Status	Error	Panel	System	Site	Relationship
node1	Active		02-2	IBM FlashSystem 7200		Local
	Service	090	7825WKP-1			Partner
- Node Errors:** A section for viewing errors.
- Node Detail:** A table with columns for Node, Hardware, Access, Location, and Ports.

Node	Hardware	Access	Location	Ports
Node ID:	1			
Node Name:	node1			
Node Status:	Active			
Part Identity:	115011M0263YM128096802L			
Node FRU:	01W0264			
Configuration Node:	Yes			
Model:	690			
System:	IBM FlashSystem 7200			
Site Name:				
System Software Build:	152.16.2009101641000			
Software Version:	8.4.0.0			
Software Build:	152.16.2009101641000			
Console IP:	9.71.42.30:443			
Max File Module Key:	No			

Figure 13-10 Service Assistant Tool post-node removal

3. Select the radio button for the node that is in service and then select **Exit Service State** from the **Actions** menu. Click **GO**, and a confirmation window opens, as shown in Figure 13-11.

The dialog box has a title bar that says "Are you sure you want to exit service state?". It contains a warning icon and the following text:

You have selected to exit service state. This action releases the node from the held-in-service state. If there are no critical node errors, the node enters candidate state. If possible, the node then becomes active in a system.

Note: If you are exiting from service state on the connected node, the connection to the service assistant is lost.

Do you want to continue?

Buttons: OK, Cancel

Figure 13-11 Exit service state

4. A confirmation window opens and shows that the node exited the service state. Click **OK**, or close the window and click **Refresh** under the list of the nodes.
5. The node should automatically read itself to the system. If not, look at the numbers in the Panel column and go back to your CLI session. Run the **addnode** command and specify the panel ID to add the node back into the cluster, as shown in Example 13-4.

Example 13-4 The addnode command

```
IBM FlashSystem 7200:superuser>svctask addnode -panelname 7825WKP-1 -iogrp io_grp0
```

- Run `svcinfolnode` again or check the SAT to ensure that the node was added back, as shown in Example 13-5.

Example 13-5 The `svcinfolnode` command

```
IBM FlashSystem 7200:superuser>svcinfolnode
id name UPS_serial_number WWNN status IO_group_id IO_group_name config_node UPS_unique_id hardware
iscsi_name iscsi_alias panel_name enclosure_id canister_id
enclosure_serial_number site_id site_name
1 node1 500507680B00E6C5 online 0 io_grp0 yes 600
iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node1 Canister 1 02-2 2 7825WKP
2 node2 500507680B00E6C4 online 0 io_grp0 no 600
iqn.1986-03.com.ibm:2145.ibmIBM FlashSystem7200.node2 Canister 2 02-1 2 1 7825WKP
IBM FlashSystem 7200:superuser>
```

Note: If you want to remove an entire control enclosure from the cluster to reduce the size of the cluster or to decommission it, you can do this task by using the GUI. Go to the Enclosure Overview window, as shown in Figure 13-9 on page 806, but instead of selecting a node, select **Enclosure Actions** and then **Remove**. A confirmation window opens. This action runs the `rmnode` command against both nodes in the control enclosure. For more information about removing an enclosure, see [IBM Documentation](#) and search for “Removing a control enclosure and its expansion enclosures”.

13.4 Configuration backup

You can download and save the configuration backup file by using the IBM FlashSystem GUI or CLI. On an *ad hoc* basis, manually perform this procedure because it can save the file directly to your workstation. The CLI option requires you to log in to the system and download the dumped file by using specific Secure Copy Protocol (SCP). The CLI option is a best practice for an automated backup of the configuration.

Important: Generally, perform a daily backup of the IBM FlashSystem configuration backup file, for which the best approach is to automate this task. Always perform another backup before any critical maintenance task, such as an update of the IBM Spectrum Virtualize Software version.

The backup file is updated by the cluster every day. Saving it after any changes to your system configuration is important. It contains configuration data of arrays, pools, volumes, and other items. The backup does not contain any data from the volumes.

To successfully perform the configuration backup, the following prerequisites must be met:

- ▶ All nodes are online.
- ▶ No independent operations that change the configuration can be running in parallel.
- ▶ No object name can begin with an underscore.

Important: *Ad hoc* backup of configuration can be done only from the CLI by using the `svconfig backup` command. Then, the output of the command can be downloaded by using SCP or GUI.

13.4.1 Backing up by using the CLI

You can use the CLI to trigger configuration backups manually or by a regular automated process. The **svcconfig backup** command generates a new backup file. Triggering a backup by using the GUI is not possible. However, you might choose to save the automated 1 AM cron backup if you have not made any configuration changes.

Example 13-6 shows how to use the **svcconfig backup** command to generate an *ad hoc* backup of the current configuration.

Example 13-6 Saving the configuration by using the CLI

```
IBM FlashSystem 7200:superuser>svcconfig backup
.....
.....
CMMVC6155I SVCCONFIG processing completed successfully
IBM FlashSystem 7200:superuser>
```

The **svcconfig backup** command generates three files that provide information about the backup process and cluster configuration. These files are dumped into the `/tmp` directory on the configuration node. Run the **lsdumps** command to list them (see Example 13-7).

Example 13-7 Listing the backup files by using the CLI

```
IBM FlashSystem 7200:superuser>lsdumps |grep backup
55 svc.config.backup.bak_7825WKP-1
56 svc.config.backup.log_7825WKP-1
57 svc.config.backup.xml_7825WKP-1
58 svc.config.backup.sh_7825WKP-1
IBM FlashSystem 7200:superuser>
```

Note: The `svc.config.backup.bak` file is a previous copy of the configuration, and not part of the current backup.

Table 13-11 lists the three files that are created by the backup process.

Table 13-11 Files that are created by the backup process

File name	Description
<code>svc.config.backup.xml</code>	This file contains your cluster configuration data.
<code>svc.config.backup.sh</code>	This file contains the names of the commands that ran to create the backup of the cluster.
<code>svc.config.backup.log</code>	This file contains details about the backup, including any error information that might have been reported.

Save the current backup to a secure and safe location. The files can be downloaded by running **scp** (UNIX) or **pscp** (Microsoft Windows), as shown in Example 13-8. Replace the IP address with the cluster IP address of your system and specify a local folder on your workstation. In this example, we save to `C:\V7000Backup`.

Example 13-8 Saving the config backup files to your workstation

```
C:\putty>pscp -unsafe
superuser@9.72.42.30:/dumps/svc.config.backup.* c:\FS7200backup
Using keyboard-interactive authentication.
Password:
svc.config.backup.bak_782 | 133 kB | 33.5 kB/s | ETA: 00:00:00 | 100%
```

```
svc.config.backup.log_782 | 16 kB | 16.8 kB/s | ETA: 00:00:00 | 100%
svc.config.backup.sh_7822 | 5 kB | 5.9 kB/s | ETA: 00:00:00 | 100%
svc.config.backup.xml_782 | 105 kB | 52.8 kB/s | ETA: 00:00:00 | 100%
```

```
C:\putty>
```

```
C:\>dir FS7200backup
```

```
Volume in drive C has no label.
Volume Serial Number is 0608-239A
```

```
Directory of C:\FS7200backup
```

```
24.10.2018 10:57 <DIR> .
24.10.2018 10:57 <DIR> ..
24.10.2018 10:57          137.107 svc.config.backup.bak_7822DFF-1
24.10.2018 10:57          17.196 svc.config.backup.log_7822DFF-1
24.10.2018 10:57           6.018 svc.config.backup.sh_7822DFF-1
24.10.2018 10:58          108.208 svc.config.backup.xml_7822DFF-1
          4 File(s)          268.529 bytes
          2 Dir(s) 48.028.662.272 bytes free
```

```
C:\>
```

Using the **-unsafe** option enables you to use the wildcard for downloading all the `svc.config.backup` files with a single command.

Tip: If you encounter the Fatal: Received unexpected end-of-file from server error, when running the **pscp** command, consider upgrading your version of PuTTY.

13.4.2 Saving the backup by using the GUI

Although it is not possible to generate an *ad hoc* backup, you can save the backup files by using the GUI. To do so, complete the following steps:

1. Select **Settings** → **Support** → **Support Package**.
2. Click the **Manual Download Instructions** drop-down menu.
3. Click **Download Existing Package**, as shown in Figure 13-12 on page 811.

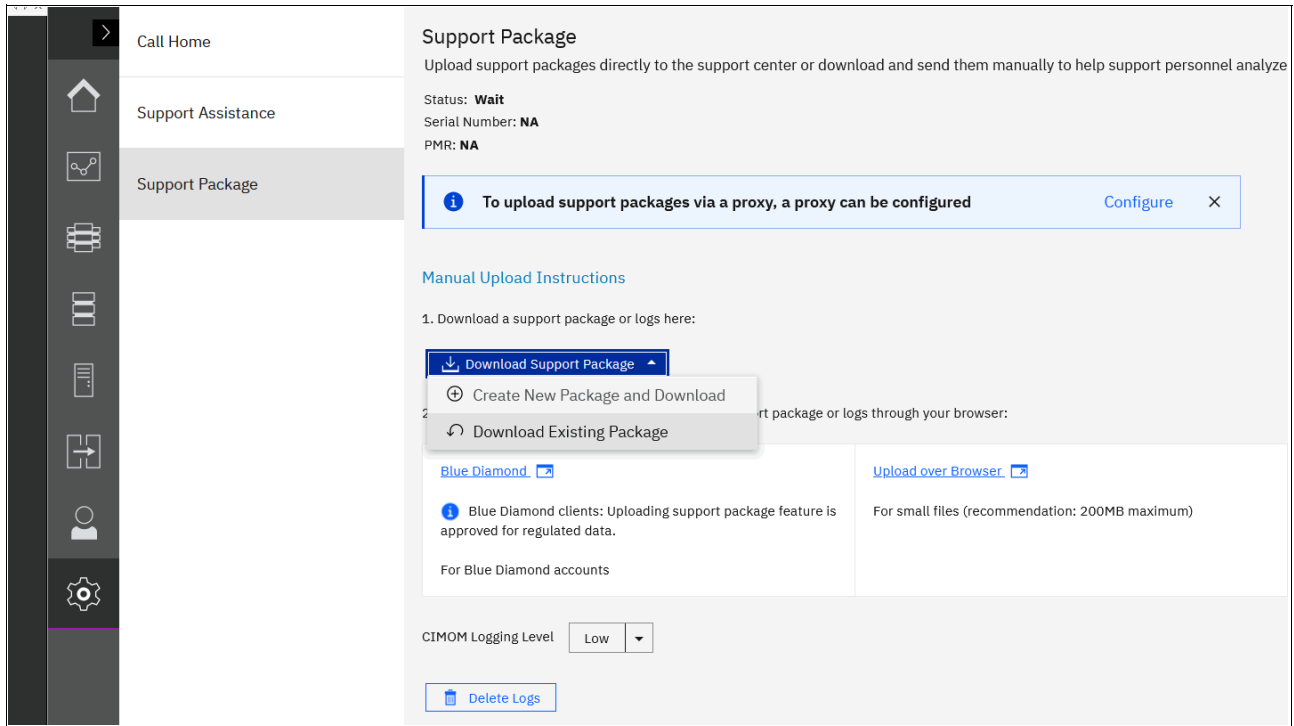


Figure 13-12 Download Existing Package

The Support Package selection window opens, as shown in Figure 13-13.

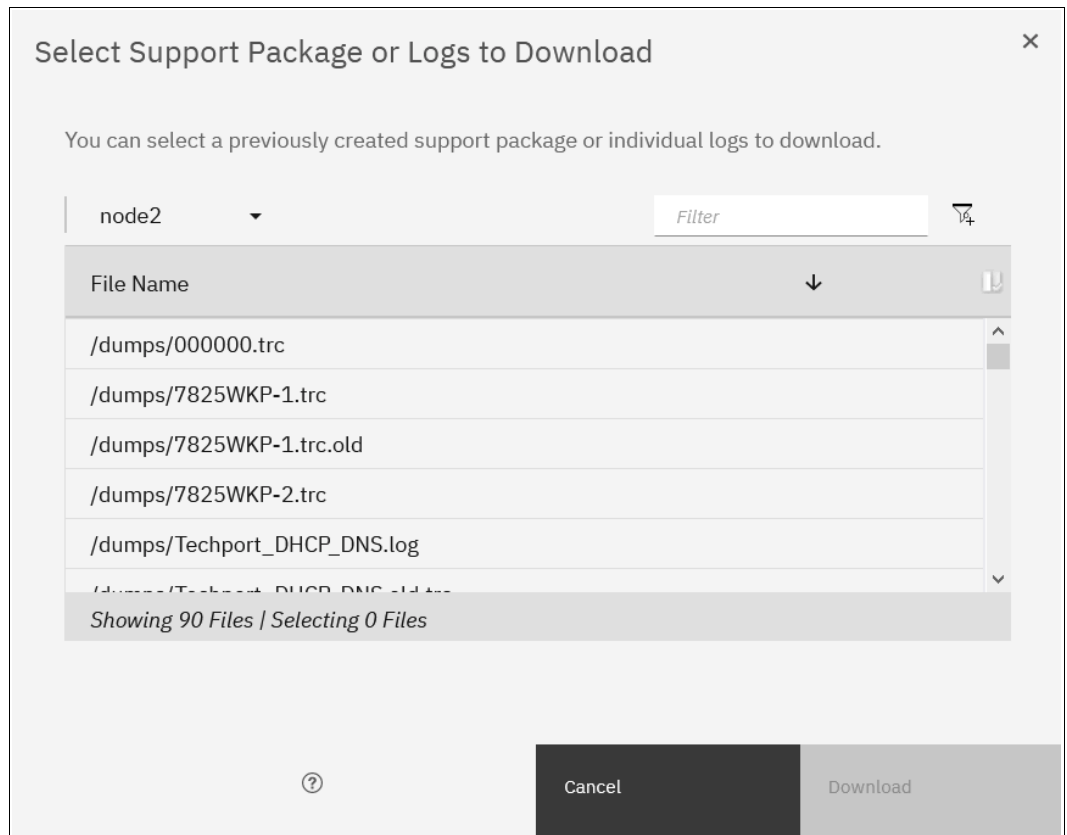


Figure 13-13 Support Package Selection

4. Filter the view by clicking in the **Filter** box, entering backup, and pressing Enter, as shown in Figure 13-14.

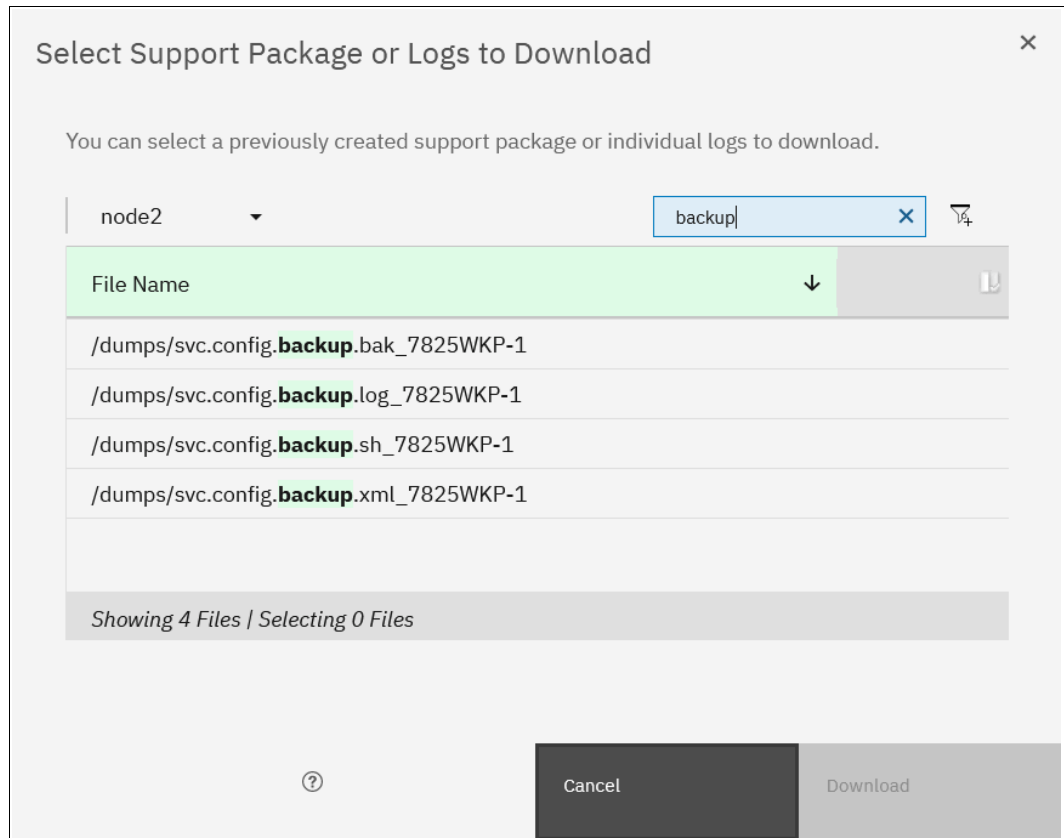


Figure 13-14 Filtering specific files for download

Note: You must select the configuration node in the upper left drop-down menu because the backup files are stored there.

5. Select all the files to include in the compressed file, and then click **Download**. Depending on your browser preferences, you might be prompted about where to save the file, otherwise it downloads to your defined download directory.

13.5 Software update

This section describes the operations to update your system software to Version 8.4.0.

The format for the software update package name ends in four positive integers that are separated by dots. For example, a software update package might have the following name:

IBM_2076_INSTALL_8.4.0

13.5.1 Precautions before the update

This section describes the precautions that you should take before you attempt an update.

Important: Before you attempt any code update, read and understand the concurrent compatibility and code cross-reference matrix for your system. For more information, see [Concurrent Compatibility and Code Cross Reference for IBM Spectrum Virtualize](#) and click **Latest system code**.

During the update, each node in the IBM FlashSystem clustered system is automatically shut down and restarted by the update process. Because each node in an I/O group provides an alternative path to volumes, use the Subsystem Device Driver (SDD) to make sure that all I/O paths between all hosts and SANs work.

If you do not perform this check, certain hosts might lose connectivity to their volumes and experience I/O errors when the IBM FlashSystem node that provides that access is shut down during the update process. You can check the I/O paths by running **datapath query SDD** commands.

13.5.2 IBM FlashSystem update test utility

The software update test utility is an IBM FlashSystem Software utility that checks for known issues that can cause problems during a software update. For more information about the utility, see [Software Upgrade Test Utility](#). Download the software update utility from this page, where you can also download the firmware. This procedure ensures that you receive the current version of this utility. You can use the **svcupgradetest** utility to check for known issues that might cause problems during a software update.

The software update test utility can be downloaded in advance of the update process. Alternately, it can be downloaded and run directly during the software update, as guided by the update wizard.

You can run the utility multiple times on the same system to perform a readiness check-in as preparation for a software update. Run this utility a final time immediately before you apply the software update, but make sure that you always use the latest version of the utility.

The installation and use of this utility is nondisruptive, and it does not require a restart of any IBM FlashSystem nodes. Therefore, there is no interruption to host I/O. The utility is installed only in the current configuration node.

System administrators must continue to check whether the version of code that they plan to install is the latest version. For more information, see [Concurrent Compatibility and Code Cross Reference for IBM Spectrum Virtualize](#).

This utility is intended to supplement rather than duplicate the tests that are performed by the IBM Spectrum Virtualize update procedure (for example, checking for unfixed errors in the error log).

A concurrent software update of all components is supported through the standard Ethernet management interfaces. However, most of the configuration tasks are restricted during the update process.

13.5.3 Updating your IBM FlashSystem to Version 8.4.0.

To update the IBM Spectrum Virtualize Software to Version 8.4.0, complete the following steps:

1. Log in by using superuser credentials. The management home window opens. Hover the cursor over **Settings** and click **System** (see Figure 13-15).

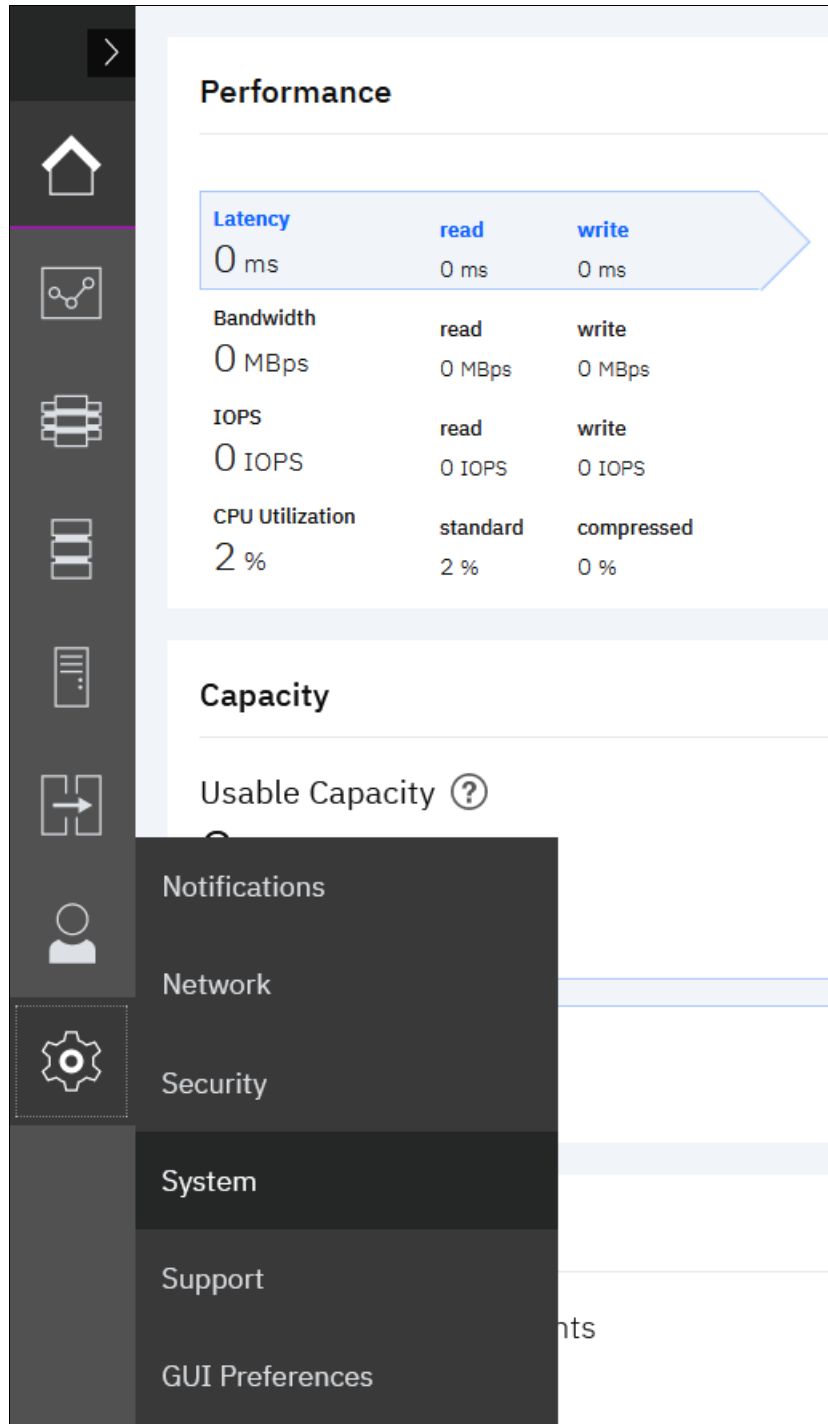


Figure 13-15 Settings menu

2. In the **System** menu, click **Update System**. The Update System window opens (see Figure 13-16).

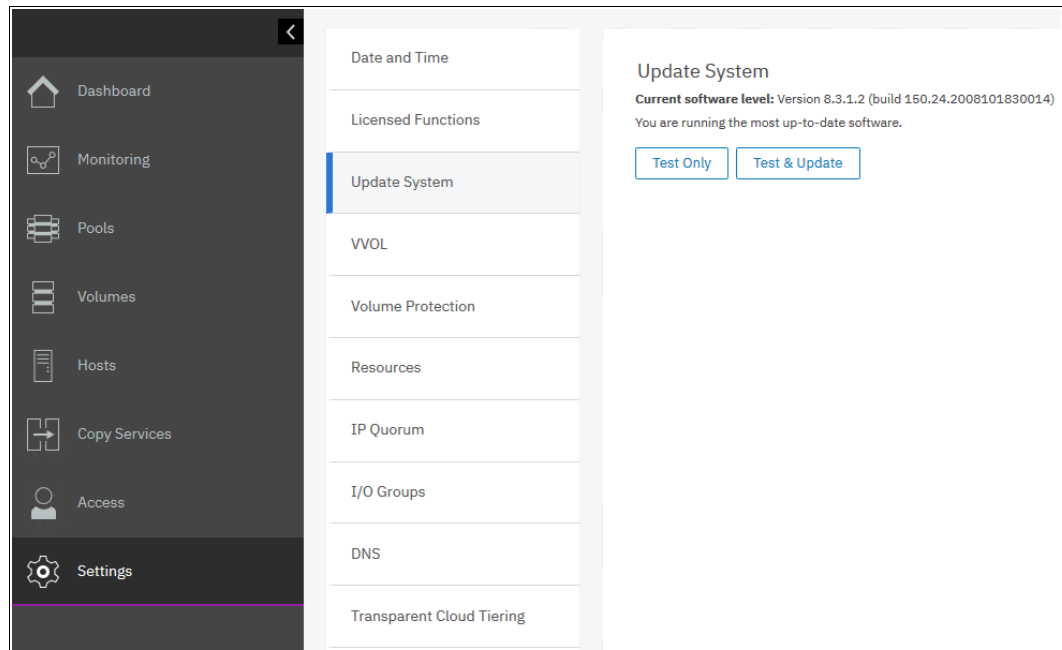


Figure 13-16 Update System window

3. From this window, you can select to run the update test utility and continue with the code update or run the test utility. For this example, we click **Test and Update**.

My Notifications: Use the My Notifications tool to receive notifications of new and updated support information to better maintain your system environment, especially in an environment where a direct internet connection is not possible.

See [My Notifications](#) (an IBM account is required) to add your system to the notifications list to be advised of support information and to download the current code to your workstation for later upload.

4. Because you downloaded both files from [Concurrent Compatibility and Code Cross Reference for IBM Spectrum Virtualize](#), you can click each folder, browse to the location where you saved the files, and upload them to the system. If the files are correct, the GUI detects and updates the target code level, as shown in Figure 13-17.

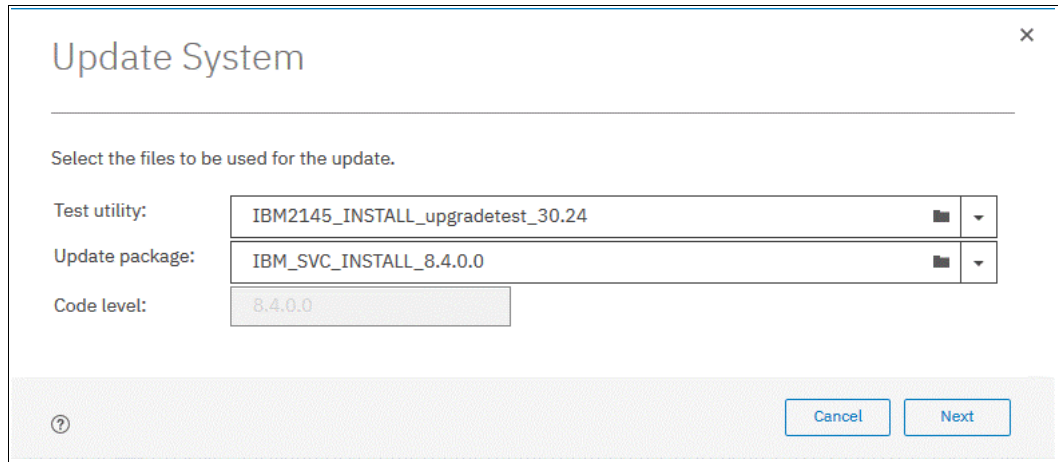


Figure 13-17 Upload option for both the test utility and update package

5. Select the type of update you want to perform, as shown in Figure 13-18. Select **Automatic update** unless IBM Support suggests **Service Assistant Manual update**. The manual update might be preferable in cases where misbehaving host multipathing is known to cause loss of access. Click **Next** to begin the update package upload process.

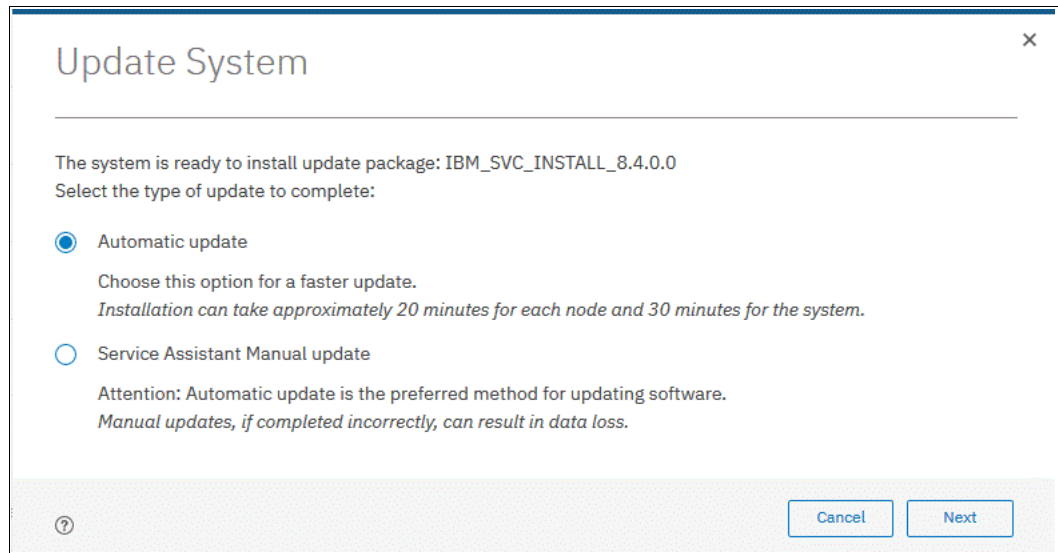


Figure 13-18 The update type selection

When updating from Version 8.1 or later, another window opens, in which you can choose a fully automated update, one that pauses when half the nodes complete the update, or one that pauses after each node update, as shown in Figure 13-19. The pause option requires that you click **Resume** to continue the update after each pause. Click **Finish**.

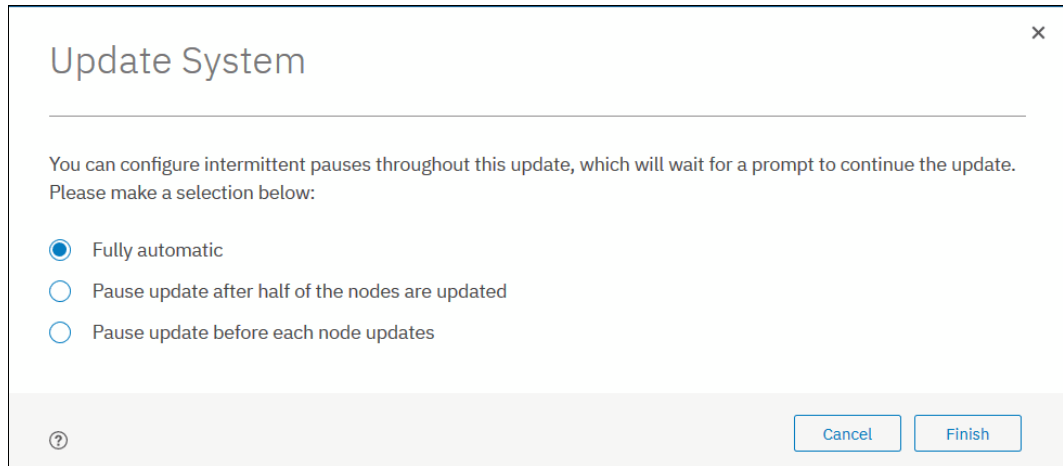


Figure 13-19 New Version 8.1 update pause options

6. After the update packages upload, the update test utility looks for any known issues that might affect a concurrent update of your system. Click **Read more** (see Figure 13-20).

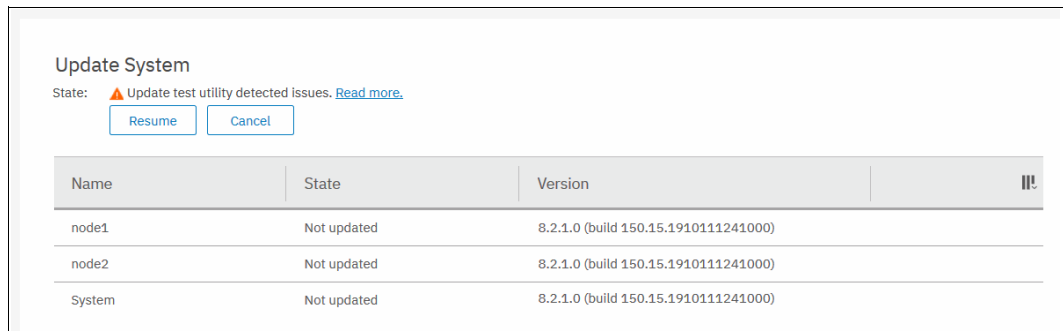


Figure 13-20 Issues that are detected by the update test utility

The results window opens and shows you what issues were detected (see Figure 13-21). In our example, the system identified an error that one or more drives in the system are running microcode with a known issue and a warning that email notification (Call Home) is not enabled. Although this issue is not a recommended condition, it does not prevent the system update from running. Therefore, we click **Close** and proceed with the update. However, you might need to contact IBM Support to help resolve more serious issues before continuing.

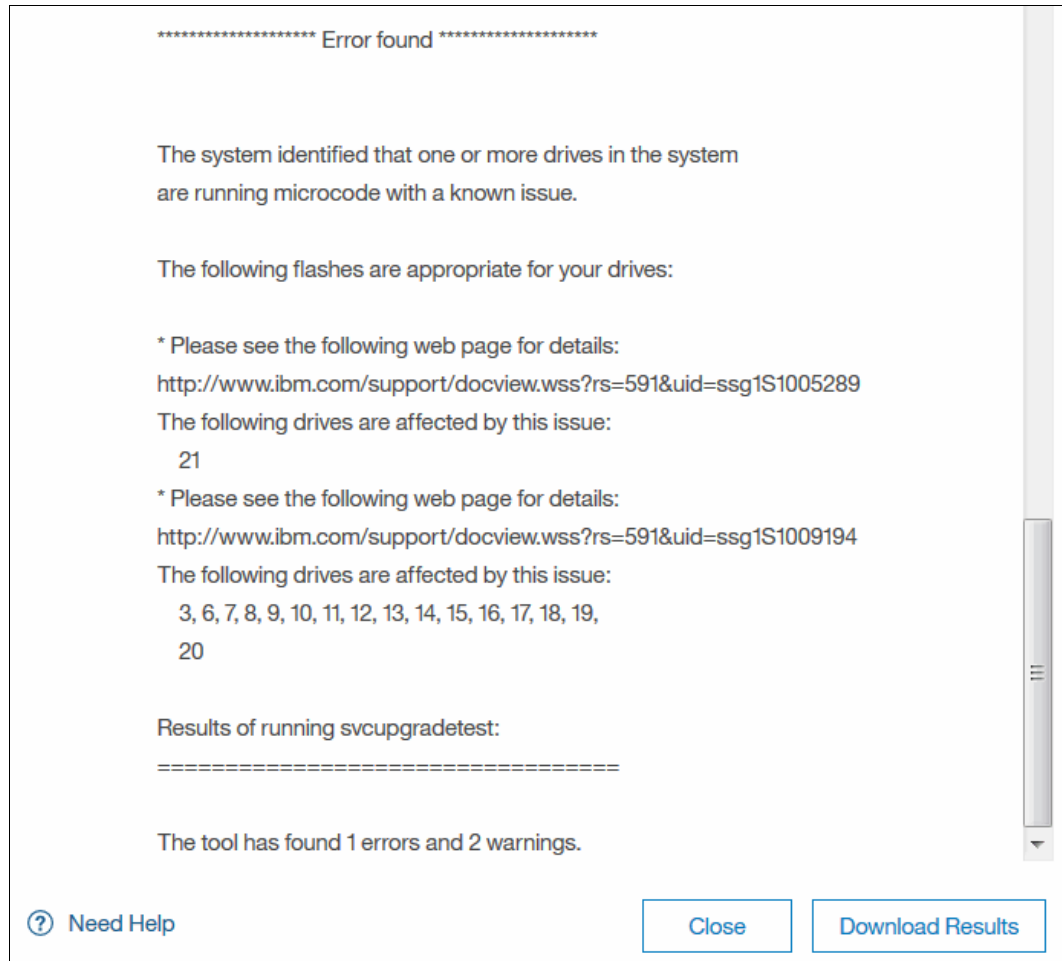


Figure 13-21 Description of the warning from the test utility

7. Click **Resume** in the Update System window and the update proceeds, as shown in Figure 13-22 on page 819.

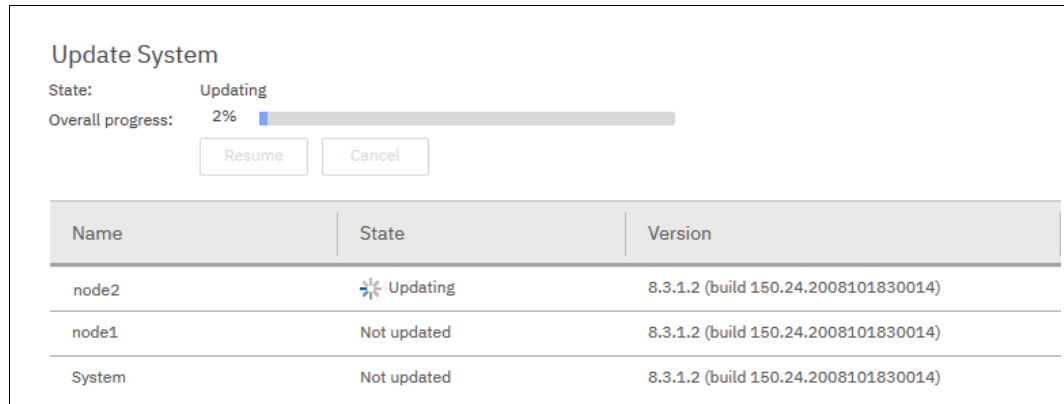


Figure 13-22 Resuming the update

Note: Because the utility detects issues, another warning appears to ensure that you investigated them and are certain that you want to proceed. When you are ready to proceed, click **Yes**.

- The system begins updating the IBM Spectrum Virtualize Software by taking one node offline and installing the new code. This process takes approximately 20 minutes. After the node returns from the update, it is listed as complete, as shown in Figure 13-23.

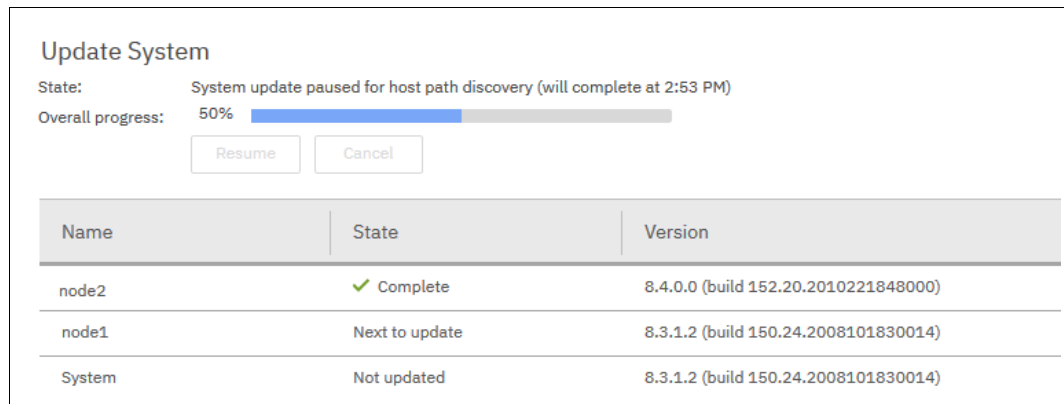


Figure 13-23 Update process paused for host path recovery

9. After a 30-minute pause, a node failover occurs and you temporarily lose connection to the GUI to ensure that multipathing recovered on all attached hosts. A warning window opens and prompts you to refresh the current session, as shown in Figure 13-24.

Tip: If you are updating from Version 7.8 or later, the 30-minute wait period can be adjusted by running `applysoftware -delay (mins)` parameter to begin the update instead of using the GUI.

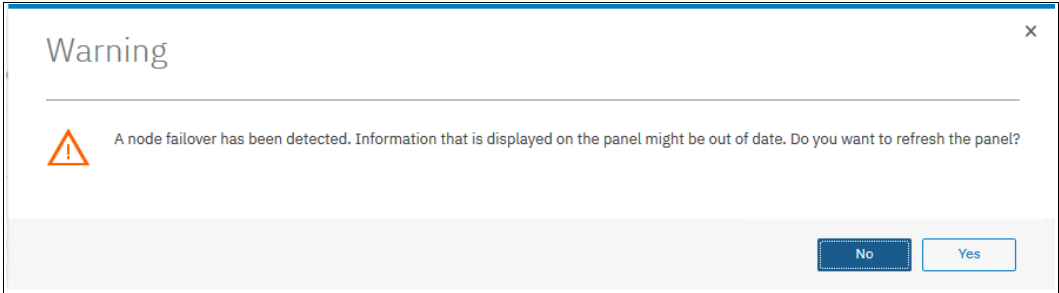


Figure 13-24 Node failover

You now see the new Version 8.4.0 GUI and the status of the second node updating, as shown in Figure 13-25.

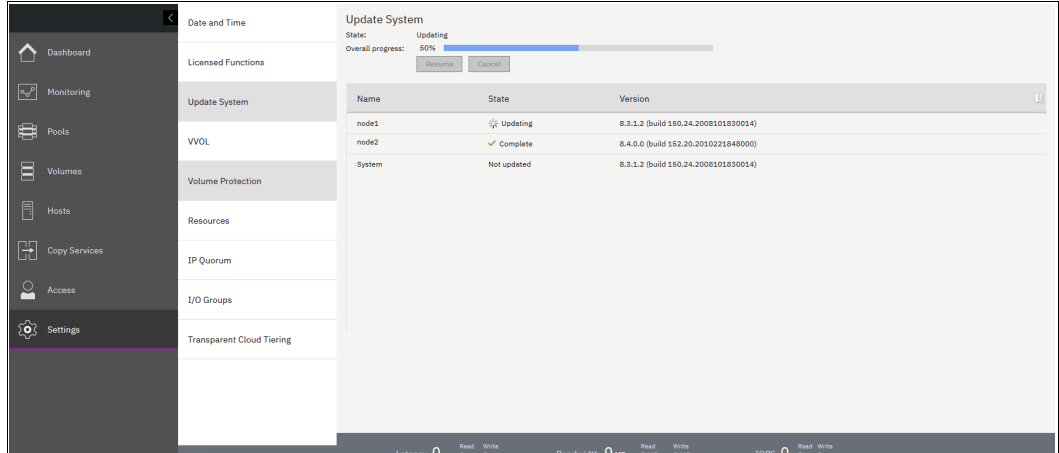


Figure 13-25 New GUI after node failover

After the last node completes, the update is committing to the system, as shown in Figure 13-26 on page 821.

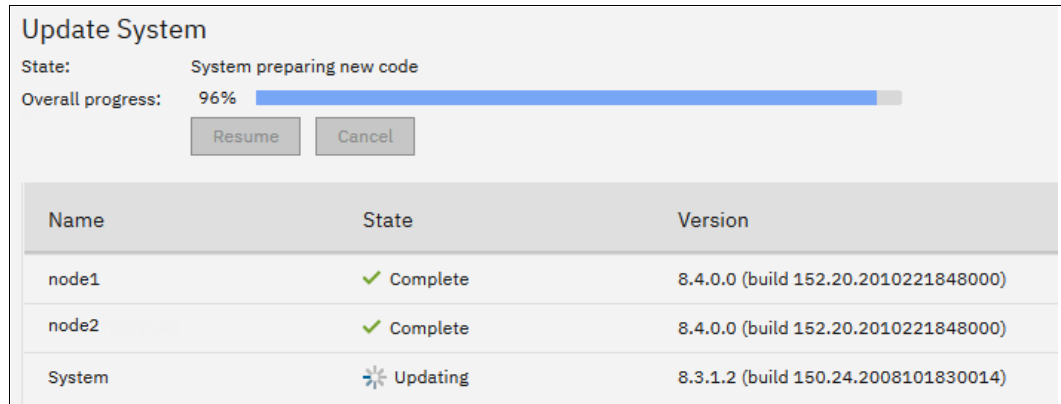


Figure 13-26 Updating the system level

The update process completes when all nodes and the system unit are committed. The final status indicates the new level of code that is installed in the system.

13.5.4 Updating the IBM FlashSystem drive code

After completing the software update as described in 13.5, “Software update” on page 812, the firmware of the disk drives in the system also must be updated. The upgrade test utility identified that earlier drives are in the system, as shown in Figure 13-27. However, this fact does not stop the system software update from being performed.

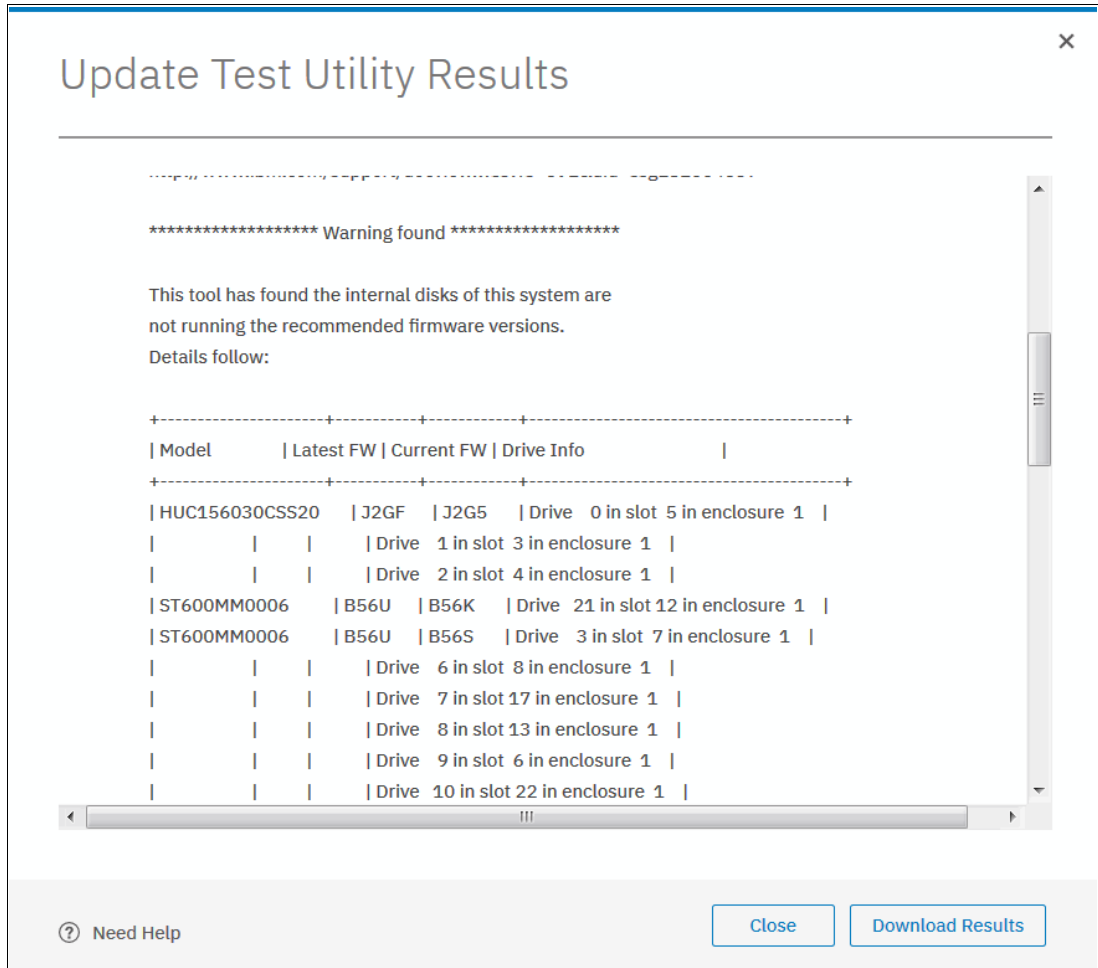


Figure 13-27 Upgrade Test Utility drive firmware warning

To update the drive code, complete the following steps:

1. Download the latest drive firmware package from [IBM Fix Central](#). Make sure that you select the correct product.
2. On the GUI, select **Pools** → **Internal Storage** and select **All Internal Storage**.
3. Click **Actions** and select **Upgrade all**, as shown in Figure 13-28 on page 823.

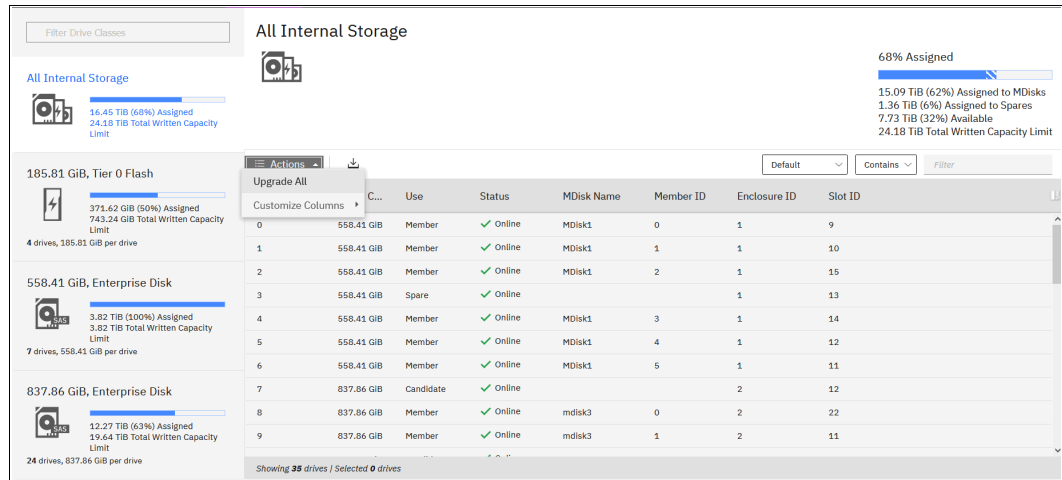


Figure 13-28 Upgrading all internal drives

Tip: The **Upgrade all** action displays only if you did not select any individual drive in the list. If you clicked an individual drive in the list, the action gives you individual drive actions; selecting **Upgrade** upgrades only that drive's firmware. You can clear an individual drive by pressing Ctrl and clicking the drive again.

- The Upgrade All Drives window opens, as shown in Figure 13-29, in which you click the small folder at the right side of the **Upgrade package** drop-down menu to go to where you saved the downloaded file in step 1 on page 822. Click **Upgrade** to upload the firmware package and begin upgrading any drives that are earlier. Do *not* select the option to install firmware, even if the drive is running a newer level. Do that only under guidance from IBM Support.

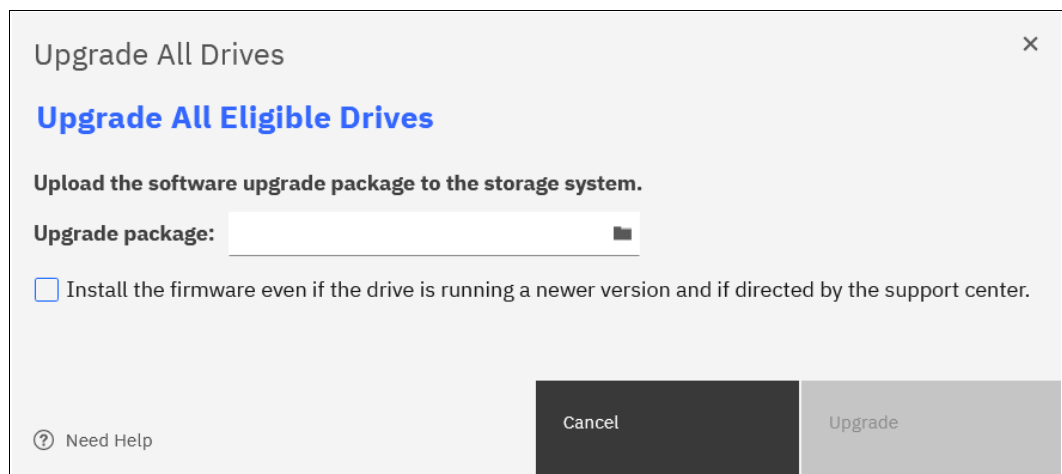


Figure 13-29 Selecting the drive upgrade package

Note: The system upgrades member drives one at a time. Although the firmware upgrades are concurrent, they do cause a brief reset to the drive. However, the redundant array of independent disks (RAID) technology enables the system to continue after this brief interruption. After a drive completes its update, a calculated wait time exists before the next drive updates to ensure that the previous drive is stable after upgrading and can vary on system load.

- With the drive upgrades running, you can view the progress by clicking the **Tasks** icon and clicking **View** for the Drive Upgrade running task, as shown in Figure 13-30.

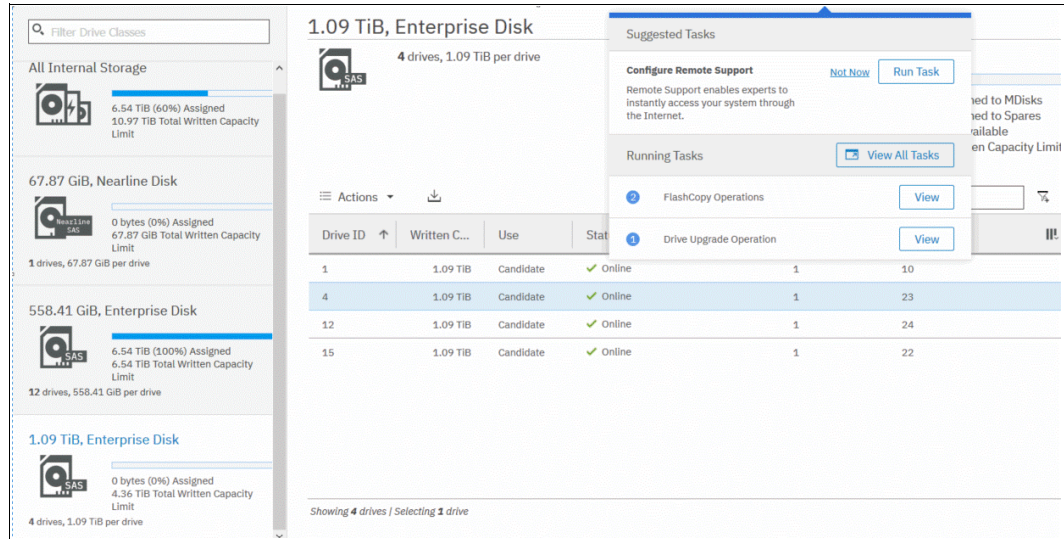


Figure 13-30 Selecting Drive Upgrade running task view

The Drive upgrade running task window opens. The drives that are pending upgrade and an estimated time of completion are visible, as shown in Figure 13-31.

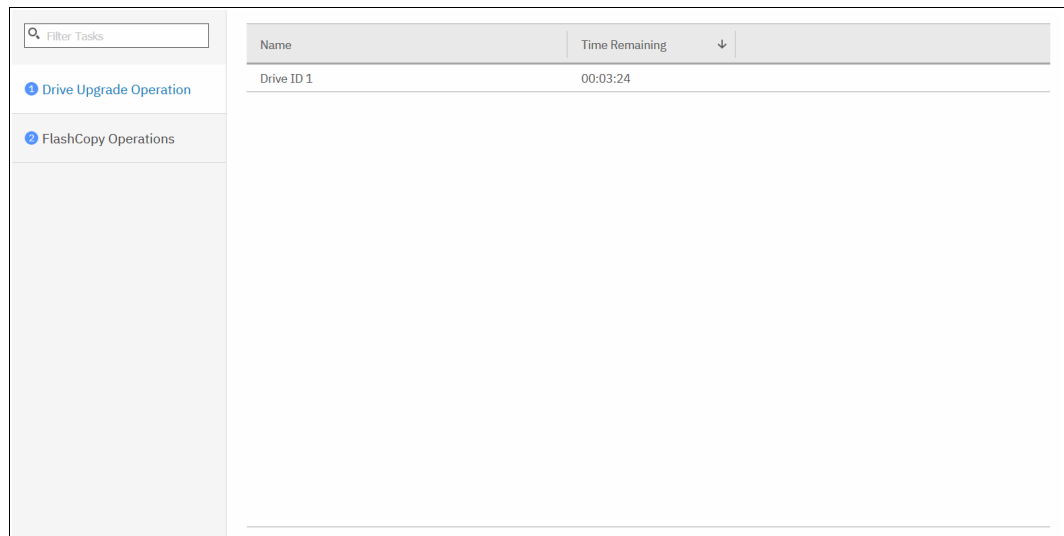


Figure 13-31 Drive upgrade progress for a single drive upgrade

- You can view each drive's firmware level in the Pools Internal Storage All Internal window by enabling the drive firmware option after right-clicking in the column header line, as shown in Figure 13-32.

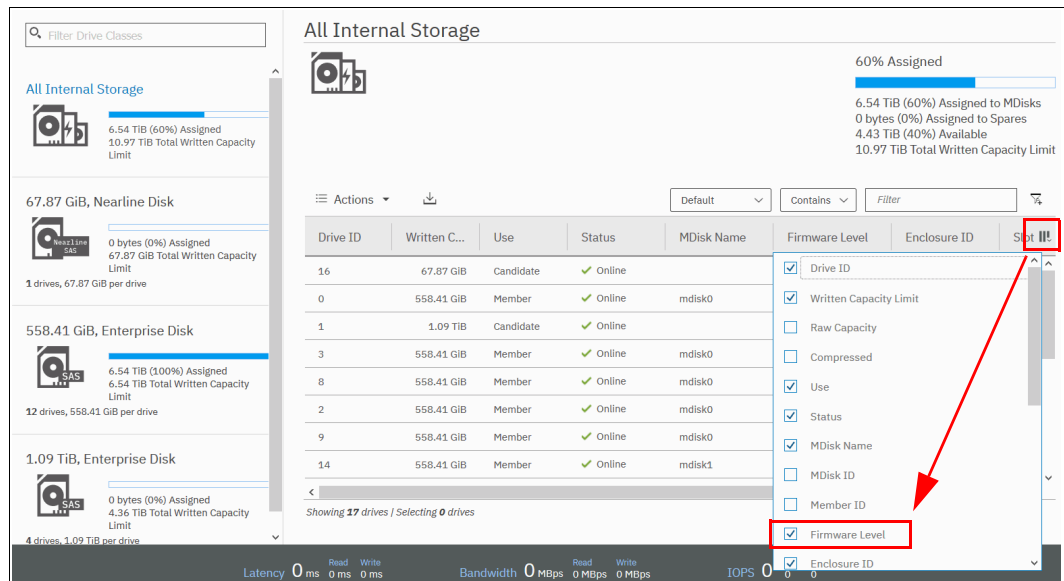


Figure 13-32 Viewing drive firmware levels

With the Firmware Level column enabled, you can see the current level of each drive, as shown in Figure 13-33.

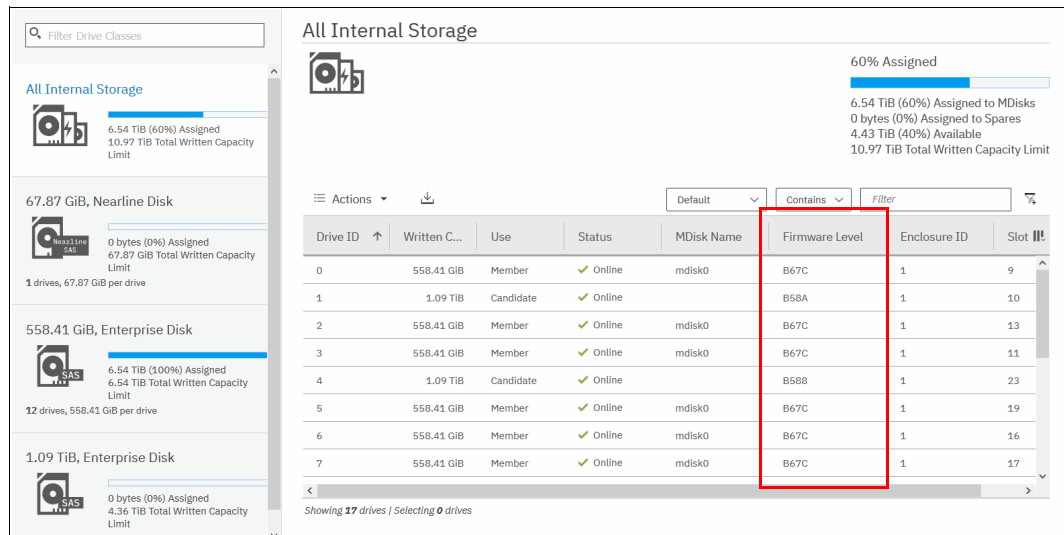


Figure 13-33 Drive firmware display

13.5.5 Manually updating the system

This example assumes that you have an 8-node cluster, as shown in Table 13-12.

Table 13-12 The iogrp setup

iogrp (0)	iogrp (1)	iogrp (2)	iogrp (3)
Node 1 (config node)	Node 3	Node 5	Node 7
Node 2	Node 4	Node 6	Node 8

After uploading the update utility test and software update package to the cluster by using PSCP and running the utility test, complete the following steps:

1. Start by removing node 2, which is the partner node of the configuration node in iogrp 0, by using the cluster GUI or CLI.
2. Log in to the service GUI to verify that the removed node is in the candidate status.
3. Select the candidate node and click **Update Manually** from the left pane.
4. Browse and find the code that you downloaded and saved to your PC.
5. Upload the code and click **Update**.
When the update completes, a message caption indicating software update completion displays. The node then restarts, and appears again in the service GUI (after approximately 20 - 25 minutes) in the candidate status.
6. Select the node and verify that it is updated to the new code.
7. Add the node back by using the cluster GUI or the CLI.
8. Select node 3 from iogrp1.
9. Repeat steps 1 - 7 to remove node 3, update it manually, verify the code, and add it back to the cluster.
10. Proceed to node 5 in iogrp 2.
11. Repeat steps 1 - 7 to remove node 5, update it manually, verify the code, and add it back to the cluster.
12. Move on to node 7 in iogrp 3.
13. Repeat steps 1 - 7 to remove node 5, update it manually, verify the code, and add it back to the cluster.

Note: The update is 50% complete. You now have one node from each iogrp that is updated with the new code manually. Always leave the configuration node for last during a manual software update.

14. Select node 4 from iogrp 1.
15. Repeat steps 1 - 7 to remove node 4, update it manually, verify the code, and add it back to the cluster.
16. Select node 6 from iogrp 2.
17. Repeat steps 1 - 7 to remove node 6, update it manually, verify the code, and add it back to the cluster.
18. Select node 8 in iogrp 3.

19.Repeat steps 1 - 7 to remove node 8, update it manually, verify the code, and add it back to the cluster.

20.Select and remove node 1, which is the configuration node in iogrp 0.

Note: A partner node becomes the configuration node because the original configuration node is removed from the cluster, which keeps the cluster manageable.

The removed configuration node becomes a candidate, and you do not have to apply the code update manually. Add the node back to the cluster. It automatically updates itself and then adds itself back to the cluster with the new code.

21.After all the nodes are updated, you must confirm the update to complete the process. The confirmation restarts each node in order, which takes about 30 minutes to complete.

The update is complete.

13.6 Health checker feature

The IBM Spectrum Control health checker feature runs in IBM Cloud. Based on the weekly Call Home inventory reporting, the health checker proactively creates recommendations. These recommendations are provided at IBM Call Home Web, which is found at [Call Home Web](#) (login required). Select **Support** → **My support** → **Call Home Web** (see Figure 13-34).

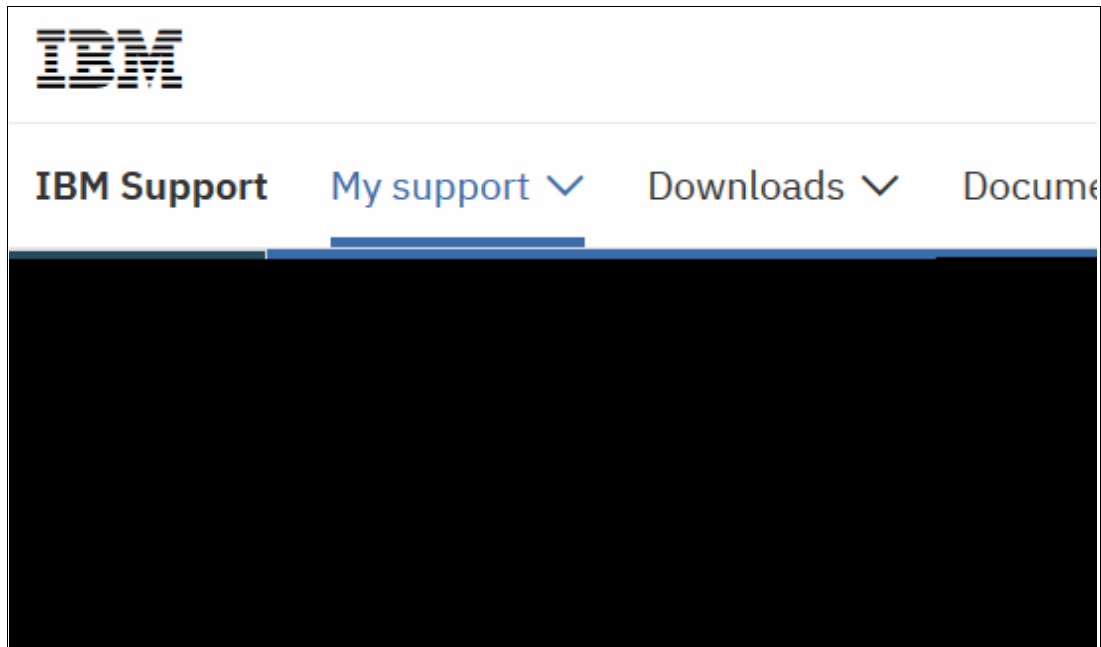


Figure 13-34 Call Home Web on ibm.com

For a video guide about how to set up and use IBM Call Home Web, see [Introducing IBM Call Home Web](#).

Note: You can also go to [Call Home Connect Cloud](#).

Another feature is the *Critical Fix Notification* function, which enables IBM to warn users that a critical issue exists in the level of code that they are using. The system notifies users when they log on to the GUI by using a web browser that is connected to the internet.

Consider the following information about this function:

- ▶ It warns users only about critical fixes, and does not warn them that they are running a previous version of the software.
- ▶ It works only if the browser also has access to the internet. The system itself does not need to be connected to the internet.
- ▶ The function cannot be disabled. Each time that it displays a warning, it must be acknowledged (with the option to not warn the user again for that issue).

The decision about what is a *critical* fix is subjective and requires judgment, which is exercised by the development team. As a result, clients might still encounter bugs in code that were not deemed critical. They continue to review information about new code levels to determine whether they must update, even without a critical fix notification.

Important: Inventory notification must be enabled and operational for these features to work. It is a best practice to enable Call Home and Inventory reporting on your IBM Spectrum Virtualize clusters.

13.7 Troubleshooting and fix procedures

The management GUI of IBM FlashSystem is a browser-based GUI for configuring and managing all aspects of your system. It provides extensive facilities to help troubleshoot and correct problems. This section explains how to effectively use its features to avoid service disruption of your system.

Figure 13-35 on page 829 shows the Monitoring menu icon for System Hardware, Easy Tier Reports, viewing events, or seeing real-time performance statistics.

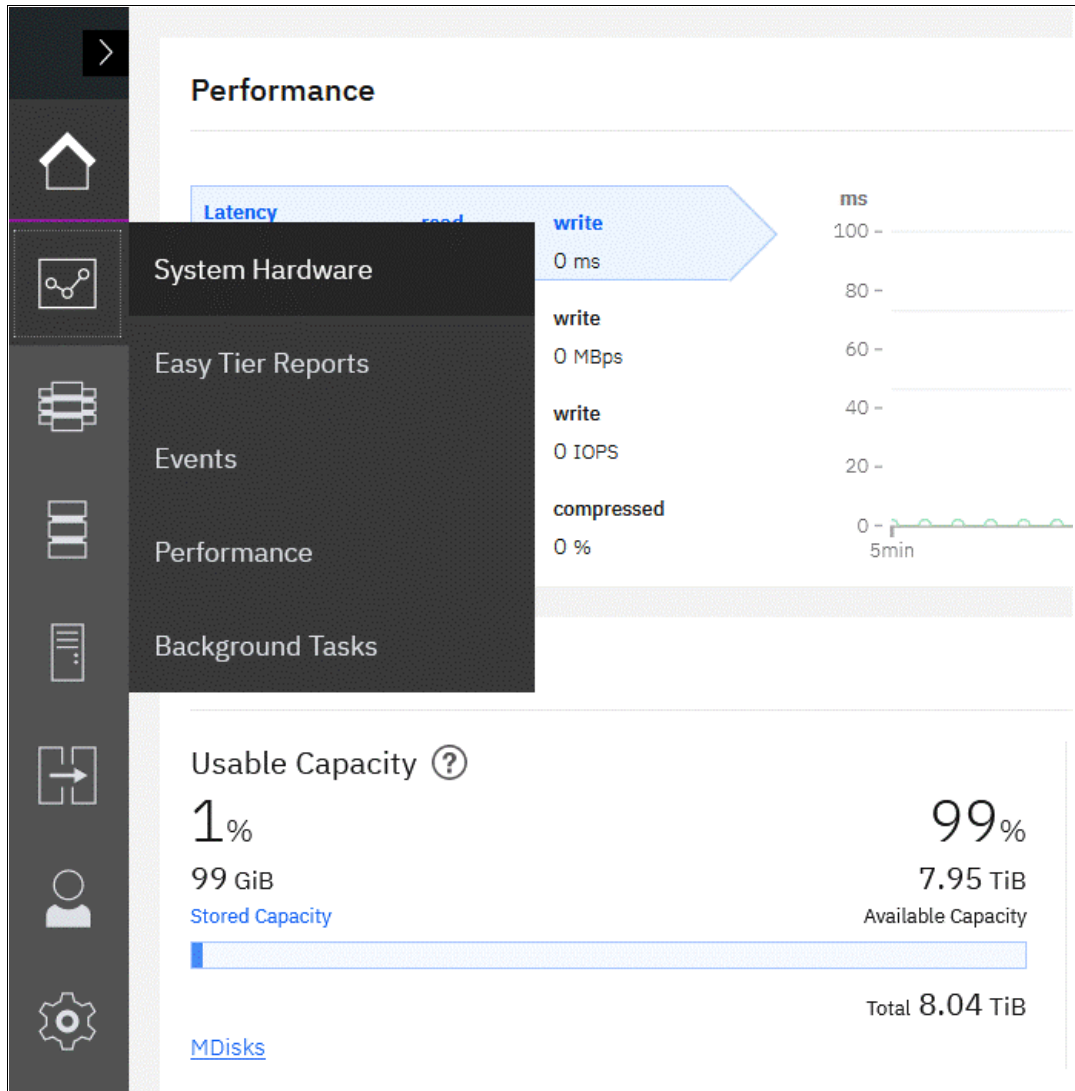


Figure 13-35 Monitoring options

Use the management GUI to manage and service your system. Select **Monitoring** → **Events** to list events that should be addressed and maintenance procedures that walk you through the process of correcting problems. Information in the Events window can be filtered four ways:

► Recommended Actions

Shows only the alerts that require attention. Alerts are listed in priority order and should be resolved sequentially by using the available fix procedures. For each problem that is selected, you can perform the following tasks:

- Run a fix procedure.
- View the properties.

► Unfixed Alerts

Displays only the alerts that are not fixed. For each entry that is selected, you can perform the following tasks:

- Run a fix procedure.
- Mark an event as fixed.

- Filter the entries to show them by specific minutes, hours, or dates.
 - Reset the date filter.
 - View the properties.
- Unfixed Messages and Alerts
- Displays only the alerts and messages that are not fixed. For each entry that is selected, you can perform the following tasks:
- Run a fix procedure.
 - Mark an event as fixed.
 - Filter the entries to show them by specific minutes, hours, or dates.
 - Reset the date filter.
 - View the properties.
- Show All
- Displays all event types whether they are fixed or unfixed. For each entry that is selected, you can perform the following tasks:
- Run a fix procedure.
 - Mark an event as fixed.
 - Filter the entries to show them by specific minutes, hours, or dates.
 - Reset the date filter.
 - View the properties.

Some events require a certain number of occurrences in 25 hours before they are displayed as unfixed. If they do not reach this threshold in 25 hours, they are flagged as *expired*. Monitoring events are below the coalesce threshold, and are transient.

Important: The management GUI is the primary tool that is used to *operate* and *service* your system. Real-time *monitoring* should be established by using SNMP traps, email notifications, or syslog messaging in an automatic manner.

13.7.1 Managing the event log

Regularly check the status of the system by using the management GUI. If you suspect a problem, first use the management GUI to diagnose and resolve the problem.

Use the views that are available in the management GUI to verify the status of the system, the hardware devices, the physical storage, and the available volumes by completing the following steps:

1. Select **Monitoring** → **Events** to see all problems that exist on the system (see Figure 13-36 on page 831).

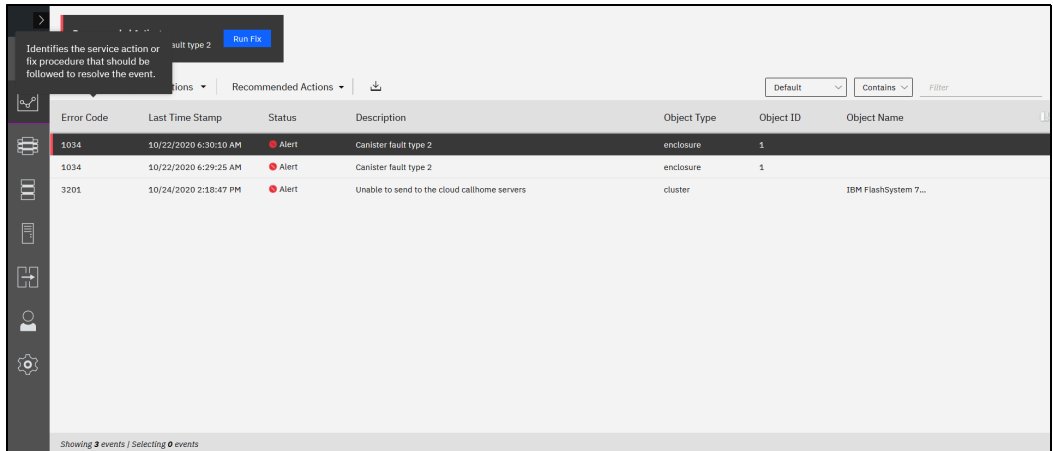


Figure 13-36 Messages in the event log

2. Select **Recommended Actions** from the drop-down list to display the most important events to be resolved (see Figure 13-37). The **Recommended Actions** tab shows the highest priority maintenance procedure that must be run. Use the troubleshooting wizard so that the system can determine the proper order of maintenance procedures.

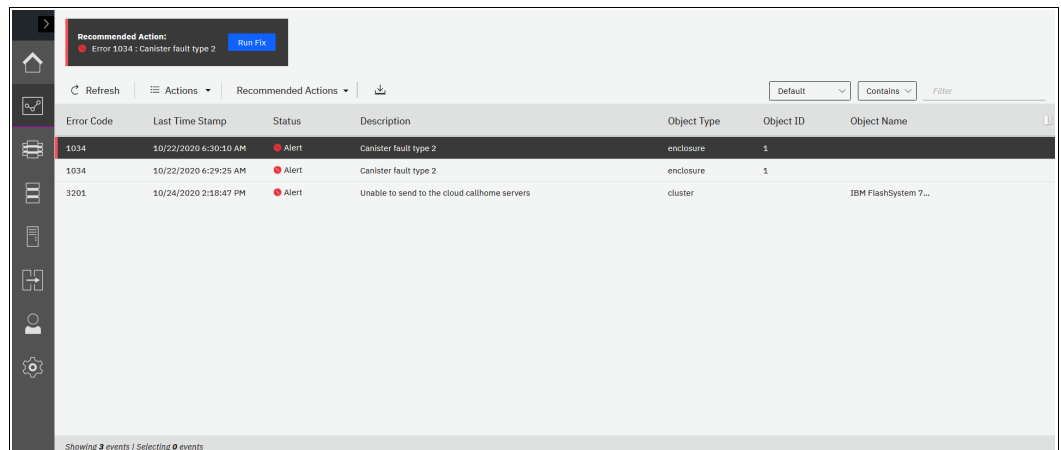


Figure 13-37 Recommended Actions

In this example, there is a canister that has a fault (service error code 1034). At any time and from any GUI window, you can directly go to this menu by clicking the **Status Alerts** icon at the top of the GUI (see Figure 13-38).

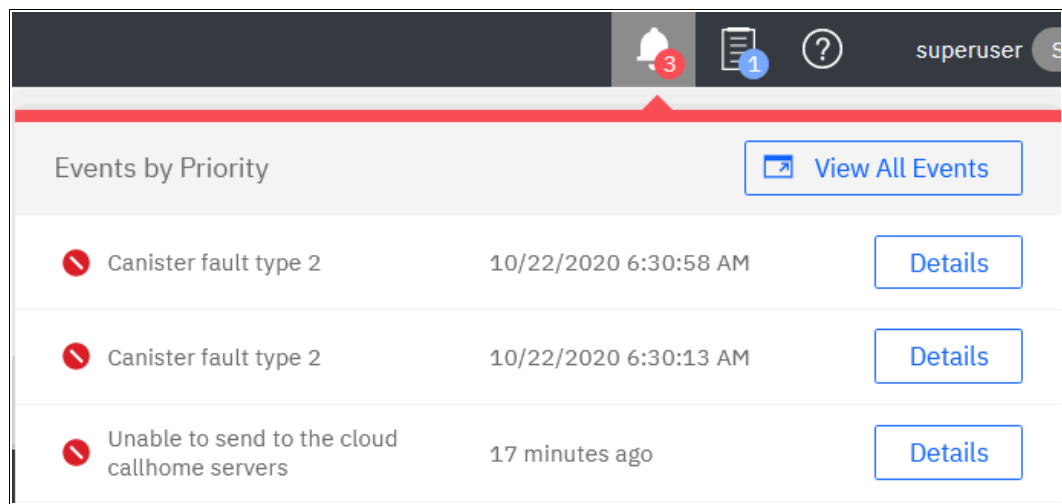


Figure 13-38 Status alerts

13.7.2 Running a fix procedure

If an error code exists for the alert, run the fix procedure to help you resolve the problem. These fix procedures analyze the system and provide more information about the problem. They suggest actions to take and walk you through the actions that automatically manage the system where necessary while ensuring availability. Finally, they verify that the problem is resolved.

If an error is reported, always use the fix procedures from the management GUI to resolve the problem for both software configuration problems and hardware failures. The fix procedures analyze the system to ensure that the required changes do not cause volumes to become inaccessible to the hosts. The fix procedures automatically perform configuration changes that are required to return the system to its optimum state.

The fix procedure displays information that is relevant to the problem, and it provides various options to correct the problem. Where possible, the fix procedure runs the commands that are required to reconfigure the system.

Note: After Version 7.4, you are no longer required to run the fix procedure for a failed drive. Hot plugging a replacement drive automatically triggers the validation processes.

The fix procedure also checks that any other existing problems do not result in volume access being lost. For example, if a PSU in a node enclosure must be replaced, the fix procedure checks and warns you whether the integrated battery in the other PSU is not sufficiently charged to protect the system.

Hint: Always use **Run Fix**, which resolves the most serious issues first. Often, other alerts are corrected automatically because they were the result of a more serious issue.

Resolving alerts in a timely manner

To minimize any impact to your host systems, always perform the recommended actions as quickly as possible after a problem is reported. Your system is resilient to most single hardware failures. However, if it operates for any period with a hardware failure, the possibility increases that a second hardware failure can result in some volume data that is unavailable. If several unfixed alerts exist, fixing any one alert might become more difficult because of the effects of the others.

13.7.3 Event log details

Multiple views of the events and recommended actions are available (see Figure 13-39). When you click the column icon at the right end of the table heading, a menu for the column choices opens.

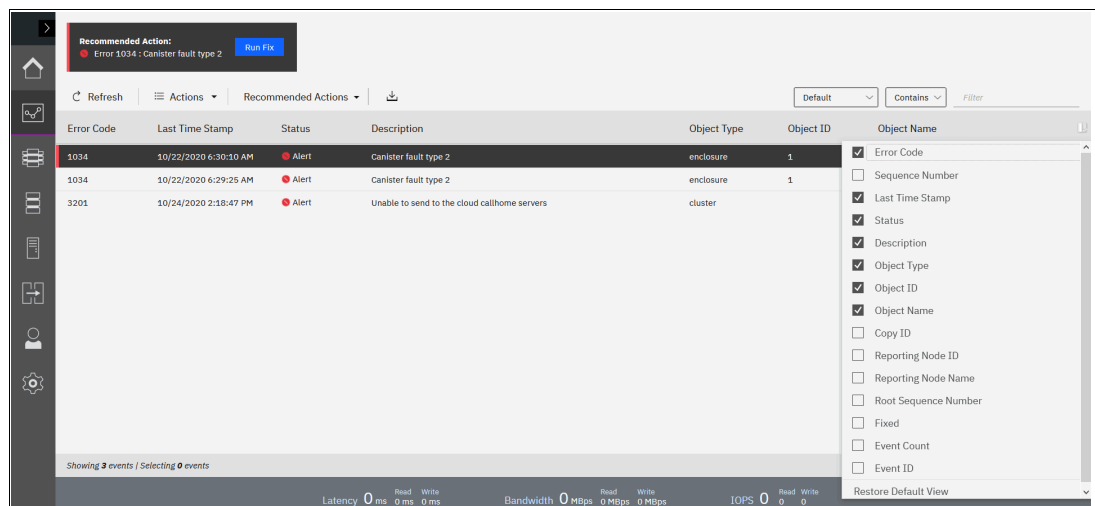


Figure 13-39 Grid options of the event log

Select or remove columns as needed. You can also extend or shrink the width of columns to fit your window resolution and size. This method is relevant for most windows in the management GUI of an IBM FlashSystem system.

Every field of the event log is available as a column in the event log grid. Several fields are useful when you work with IBM Support. The preferred method in this case is to use the Show All filter, with events sorted by timestamp. All fields have the sequence number, event count, and the fixed state. Clicking **Restore Default View** sets the grid back to the defaults.

You might want to see more details about each critical event. Some details are not shown in the main grid. To access the properties and sense data of a specific event, double-click the specific event anywhere in its row.

The properties window opens (see Figure 13-40) with all the relevant sense data. This data includes the first and last time of an event occurrence, number of times the event occurred, worldwide port name (WWPN), worldwide node name (WWNN), enabled or disabled automatic fix, and other information.

First Time Stamp	Last Time Stamp	Fixed Time Stamp	Event Count
10/22/2020 6:30:10 AM	10/22/2020 6:30:10 AM		1

Event ID: 045022
 Event ID Text: Canister has been in a degraded state for too long and cannot be recovered
 Sequence Number: 104
 Object Type: enclosure
 Object ID: 1
 Object Name:
 Secondary Object ID:
 Secondary Object Type:
 Copy ID:
 Reporting Node ID:
 Reporting Node Name:
 Root Sequence Number:
 Error Code: 1034
 Error Code Text: Canister fault type 2
 Dmp Family: IBM

Figure 13-40 Event sense data and properties

For more information about troubleshooting options, search for “Troubleshooting” at [IBM Documentation](#).

13.8 Monitoring

An important step is to correct any issues that are reported by your system as soon as possible. Configure your system to send automatic notifications to a standard Call Home server or to the new Cloud Call Home server when a new event is reported. To avoid having to monitor the management GUI for new events, select the type of event for which you want to be notified. For example, you can restrict notifications to only events that require action.

The following event notification mechanisms are available:

- ▶ Call Home

An event notification can be sent to one or more email addresses. This mechanism notifies individuals of problems. Individuals can receive notifications wherever they have email access, including mobile devices.

- ▶ Cloud Call Home

Cloud services for Call Home is the optimal transmission method for error data because it ensures that notifications are delivered directly to the IBM Support Center.

- ▶ SNMP

An SNMP traps report can be sent to a data center management system, such as IBM Systems Director, which consolidates SNMP reports from multiple systems. With this mechanism, you can monitor your data center from a single workstation.

- ▶ Syslog

A syslog report can be sent to a data center management system that consolidates syslog reports from multiple systems. With this option, you can monitor your data center from a single location.

If your system is within warranty or if you have a hardware maintenance agreement, configure your IBM FlashSystem system to send email events directly to IBM if an issue that requires hardware replacement is detected. This mechanism is known as *Call Home*. When this event is received, IBM automatically opens a problem report and, if appropriate, contacts you to help resolve the reported problem.

Important: If you set up Call Home to IBM, ensure that the contact details that you configure are correct and kept updated. Personnel changes can cause delays in IBM making contact.

Cloud Call Home is designed to work with new service teams and improves connectivity and ultimately should improve customer support.

Note: If the customer does not want to open the firewall, Cloud Call Home does not work and the customer can disable Cloud Call Home. Call Home is used instead.

13.8.1 Email notifications and the Call Home function

The Call Home function of IBM FlashSystem uses the email notification that is sent to the specific IBM Support Center. Therefore, the configuration is like sending emails to the specific person or system owner.

The following procedure summarizes how to configure email notifications and emphasizes what is specific to Call Home:

1. Prepare your contact information that you want to use for the email notification and verify the accuracy of the data. From the GUI menu, select **Settings** → **Support** → **Call Home**.
2. Select **Call Home**, and then click **Enable Notifications** (see Figure 13-41). For more information, see [IBM Documentation](#).

For the correct functioning of email notifications, ask your network administrator if Simple Mail Transfer Protocol (SMTP) is enabled on the management network and is not, for example, blocked by firewalls. Be sure to test the accessibility to the SMTP server by using the `telnet` command (port 25 for a non-secured connection, port 465 for Secure Sockets Layer (SSL)-encrypted communication) by using any server in the same network segment.

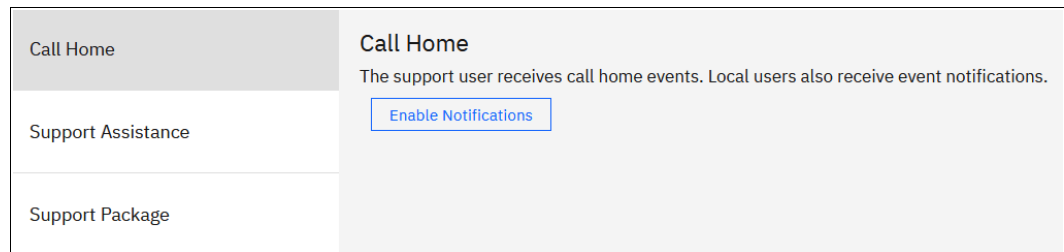


Figure 13-41 Configuring Call Home notifications

Figure 13-42 shows the option to enable Cloud Call Home.

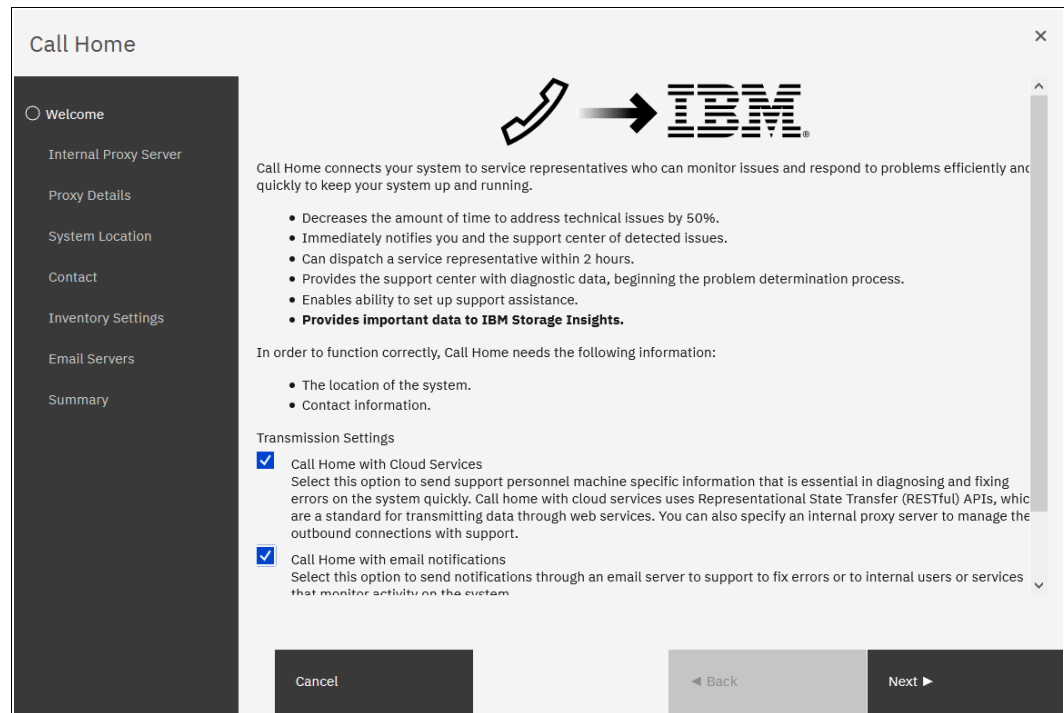


Figure 13-42 Cloud Call Home service

3. After clicking **Next** on the Welcome window, enter the information about the location of the system (see Figure 13-43) and contact information of the system administrator (see Figure 13-44 on page 838) to be contacted by IBM Support. *Always* keep this information current.

System Location

Service parts should be shipped to the same physical location as the system.

Company name:

System address:

City:

State or province:

Postal code:

Country or region: ▼

Machine location:

Figure 13-43 Location of the device

Figure 13-44 shows the contact information of the owner.

The screenshot displays a configuration window titled "Call Home" with a sidebar on the left and a main content area on the right. The sidebar contains a list of menu items: "Welcome" (checked), "Internal Proxy Server" (checked), "System Location" (checked), "Contact" (selected), "Inventory Settings", "Email Servers", and "Summary". The main content area is titled "Contact" and includes a sub-header "The support center contacts this person to resolve issues on the system." Below this is an information box with a blue header "Enter business-to-business contact information" and a note: "To comply with privacy regulations, personal contact information for individuals with your organization is not recommended." The form contains four input fields: "Name" (filled with "Author"), "Email" (filled with "author@ibm.com"), "Phone (primary)" (filled with "012-345-678"), and "Phone (alternate)" (empty). At the bottom, there are three buttons: "Cancel", "Back", and "Apply and Next".

Figure 13-44 Contact information

In the next window, you can enable Inventory Reporting and Configuration Reporting, as shown in Figure 13-45.

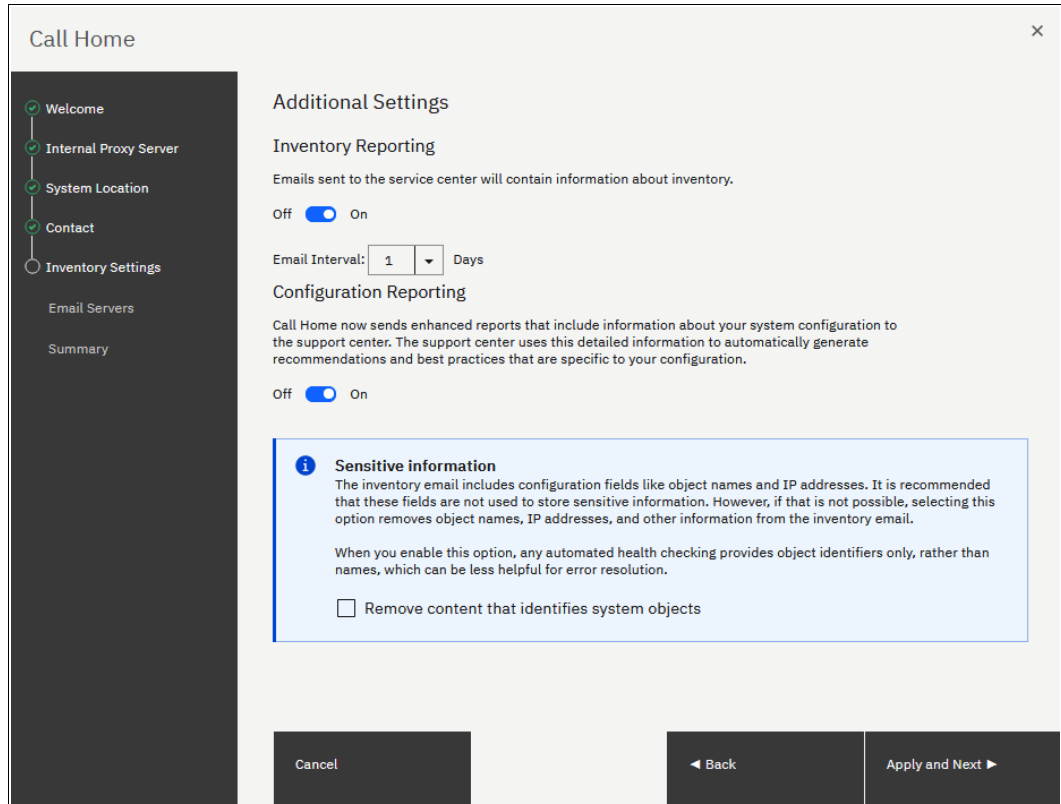


Figure 13-45 Inventory Reporting and Configuration Reporting

4. Configure the SMTP server according to the instructions that are shown in Figure 13-46. When the correct SMTP server is provided, you can test the connectivity by clicking **Ping** to verify that it can be contacted. Then, click **Apply and Next**.

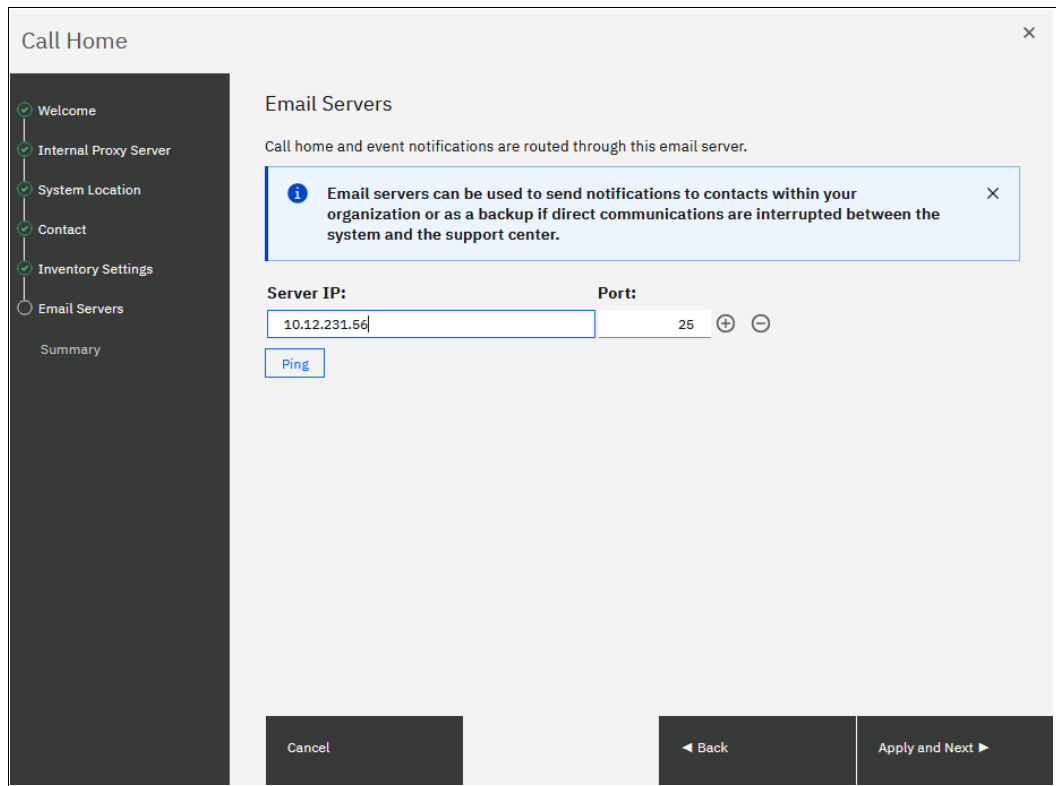


Figure 13-46 Configuring email servers and inventory reporting

5. A summary window opens. Verify all the information, and then click **Finish**. You are returned to the Email Settings window, where you can verify the email addresses of IBM Support (callhome1@de.ibm.com) and optionally add local users who also need to receive notifications (see Figure 13-47).

Call Home

The support user receives call home events. Local users also receive event notifications.

Edit
Disable Notifications

Transmission Settings

Call Home with Cloud Services

Call Home with email notifications

Call Home with cloud services

Connection: Error Test Internet Connection

Last Connection: **Failure** at 10/24/2020 3:36:23 PM

Proxy: Not configured Add Proxy

Call Home with email notifications

Email Servers

IP Address	Server Port	Status
10.12.231.56	25	◆ Untried

Support Center Email

Email Address: callhome1@de.ibm.com Error Events Inventory Test

Email Users

Email Address	Notifications			
	Error	Warning	Info	Inventory
	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 13-47 Setting email recipients and alert types

The default support email address callhome1@de.ibm.com is predefined by the system to receive Error Events and Inventory. Do not change these settings or disable the 7 day reporting interval at the bottom of the Settings window.

You can modify or add local users by using Edit mode after the initial configuration is saved.

The **Inventory Reporting** function is enabled by default for Call Home. Rather than reporting a problem, an email is sent to IBM that describes your system hardware and critical configuration information. Object names and other information, such as IP addresses, are *not* included. By default, the inventory email is sent weekly, which allows an IBM Cloud service to analyze the inventory email and inform you whether the hardware or software that you are using requires an update because of any known issue, as described in 13.6, “Health checker feature” on page 827.

Figure 13-47 on page 841 shows the configured email notification and Call Home settings.

6. After completing the configuration wizard, test the email function. To do so, enter Edit mode, as shown in Figure 13-48. In the same window, you can define more email recipients or alter any contact and location details as needed.

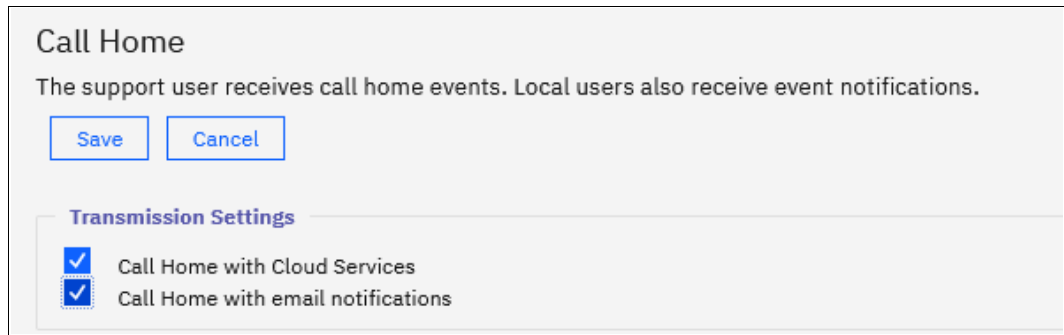


Figure 13-48 Entering Edit mode

- In Edit mode, you can change any of the previously configured settings. After you are finished editing these parameters, adding more recipients, or testing the connection, save the configuration so that the changes take effect (see Figure 13-49).

Call Home

The support user receives call home events. Local users also receive event notifications.

Transmission Settings

- Call Home with Cloud Services
- Call Home with email notifications

Call Home with cloud services

Connection: ● Error

Last Connection: **Failure** at 10/24/2020 3:44:43 PM

Proxy: Not configured

Call Home with email notifications

Email Servers

IP Address	Server Port	Status	
10.12.231.56	25	Untried	⊕ ⊖

Support Center Email

Email Address: Error Events Inventory

Email Users

Email Address	Notifications				
	Error	Warning	Info	Inventory	
redbooks@ibm.com	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	⊕ ⊖

Figure 13-49 Saving a modified configuration

Note: The Test button appears for new email users after first saving and then editing again.

Disabling and enabling notifications

At any time, you can temporarily or permanently disable email notifications, as shown in Figure 13-50. This is best practice when performing activities in your environment that might generate errors on IBM Spectrum Virtualize, such as SAN reconfiguration or replacement activities. After the planned activities, remember to re-enable the email notification function. The same results can be achieved by running the `svctask stopmail` and `svctask startmail` commands.

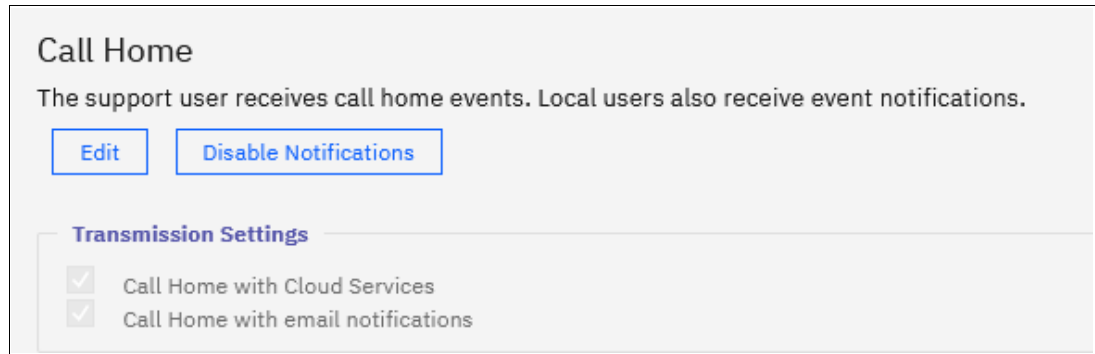


Figure 13-50 Disabling or enabling email notifications

13.8.2 Remote Support Assistance

Introduced with Version 8.1, Remote Support Assistance enables IBM Support to remotely connect to an IBM FlashSystem system through a secure tunnel to perform analysis, log collection, and software updates. The tunnel can be enabled *ad hoc* by the client or as a permanent connection.

Note: Clients who purchased Enterprise Class Support (ECS) are entitled to IBM Support by using Remote Support Assistance to quickly connect and diagnose problems. However, IBM Support might choose to use this feature on non-ECS systems at their discretion. Therefore, configure and test the connection on all systems.

If you are enabling Remote Support Assistance, ensure that the following prerequisites are met:

- ▶ Cloud Call Home or a valid email server are configured (Cloud Call Home is used as the primary method to transfer the token when you initiate a session, with email as backup).
- ▶ A valid service IP address is configured on each node in the system.
- ▶ If your IBM FlashSystem system is behind a firewall or if you want to route traffic from multiple storage systems to the same place, you must configure a Remote Support Proxy server. Before you configure Remote Support Assistance, the proxy server must be installed and configured separately. During the setup for Support Assistance, specify the IP address and the port number for the proxy server on the Remote Support Centers window.
- ▶ If you do not have firewall restrictions and the nodes are directly connected to the internet, request your network administrator to allow connections to 129.33.206.139 and 204.146.30.139 on Port 22.
- ▶ Uploading support packages and downloading software have direct connections to the internet. A DNS must be defined on your system for both of these functions to work.

- ▶ To ensure that support packages are uploaded correctly, configure the firewall to allow connections to the following IP addresses on port 443: 129.42.56.189, 129.42.54.189, and 129.42.60.189.
- ▶ To ensure that software is downloaded correctly, configure the firewall to allow connections to the following IP addresses on port 22: 170.225.15.105, 170.225.15.104, 170.225.15.107, 129.35.224.105, 129.35.224.104, and 129.35.224.107.

Figure 13-51 shows how you can find Setup Remote Support Assistance if you closed the window.

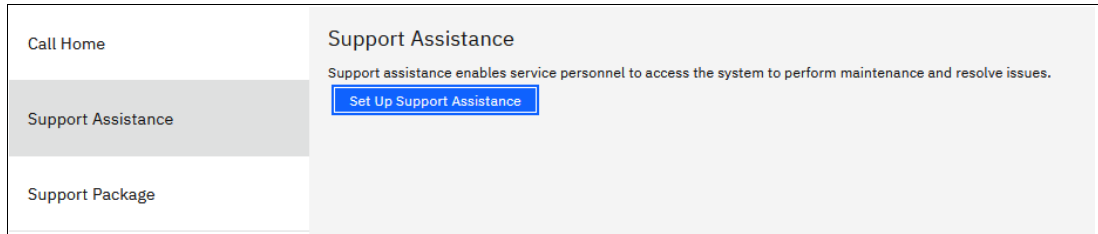


Figure 13-51 Support Assistance menu

Choosing to set up Support Assistance opens a wizard to guide you through the following configuration process:

1. Figure 13-54 on page 847 shows the first wizard window. To keep remote assistance disabled, select **I want support personnel to work on-site only**. To enable remote assistance, select **I want support personnel to access my system both on-site and remotely**. Click **Next**.

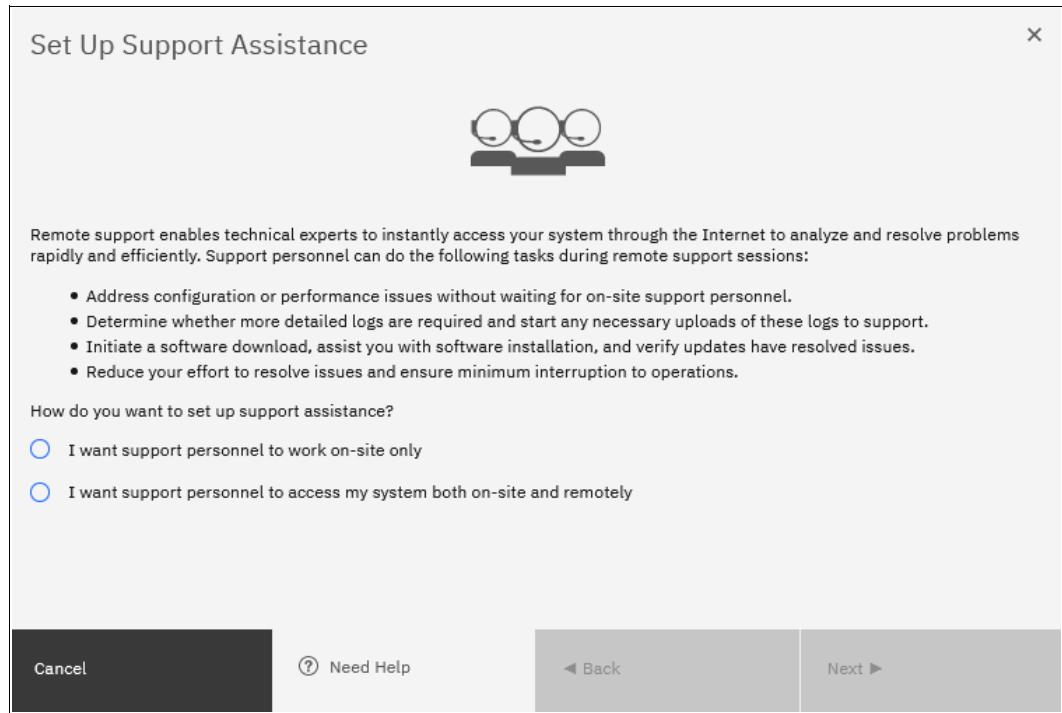


Figure 13-52 Enabling or disabling the support wizard

Note: Selecting **I want support personnel to work on-site only** does not entitle you to expect IBM Support to attend onsite for all issues. Most maintenance contracts are for customer-replaceable unit (CRU) support, where IBM diagnoses your problem and sends a replacement component for you to install, if required.

If you prefer to have IBM perform replacement tasks for you, contact your local sales person to investigate an upgrade to your current maintenance contract.

2. Figure 13-53 lists the IBM Support Center IP addresses and Secure Shell (SSH) port that must be open in your firewall. You can also define a Remote Support Assistance Proxy if you have multiple systems in the data center, which allows for a firewall configuration being required only for the proxy server rather than every storage system. In this example, we do not have a proxy server and leave the field blank. Click **Next**.

Name	IP Address	Port
default_support_center0	129.33.206.139	22
default_support_center1	204.146.30.139	22

Figure 13-53 Support wizard proxy setup

3. The next window prompts you about whether you want to open a tunnel to IBM permanently, which allows IBM to connect to your system **At Any Time**, or **On Permission Only**, as shown in Figure 13-54 on page 847. **On Permission Only** requires a storage administrator to log on to the GUI and enable the tunnel when required. Click **Finish**.

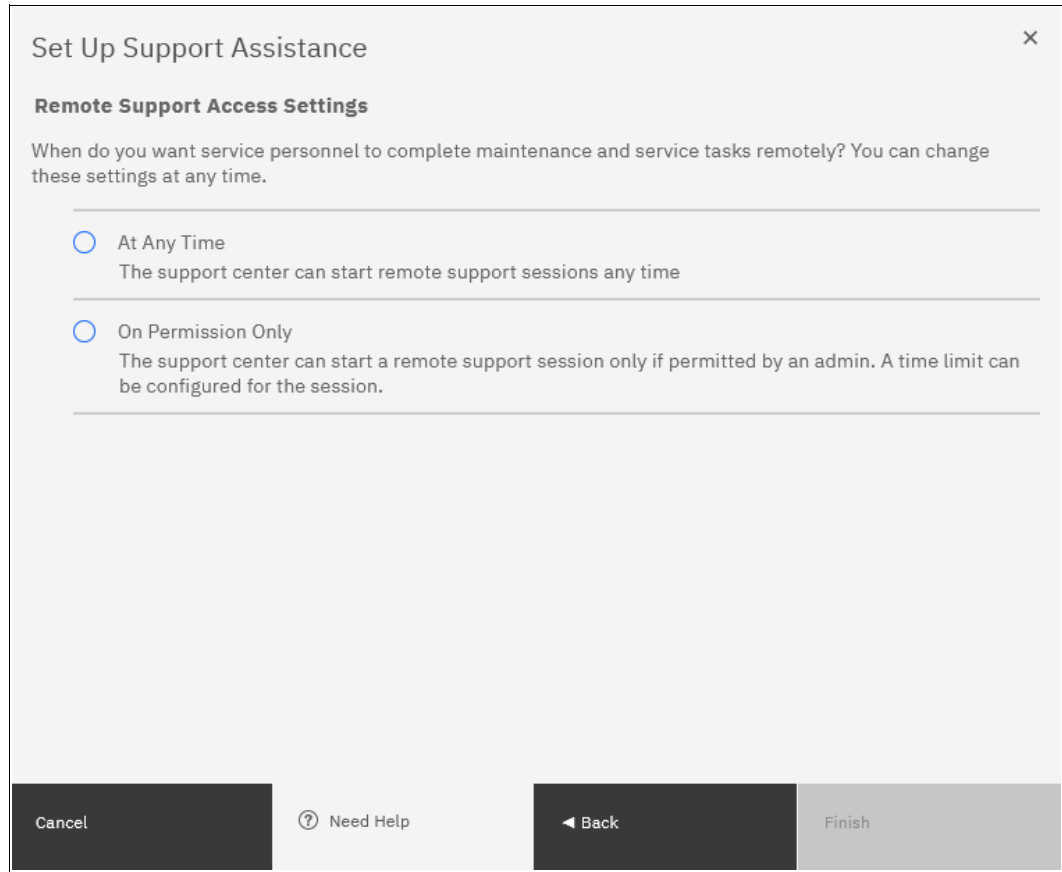


Figure 13-54 Support wizard access choice

4. After completing the remote support setup, you can view the status of any remote connection, start a session, test the connection to IBM, and reconfigure the setup. As shown in Figure 13-55, we successfully tested the connection. Click **Start New Session** to open a tunnel through which IBM Support can connect.

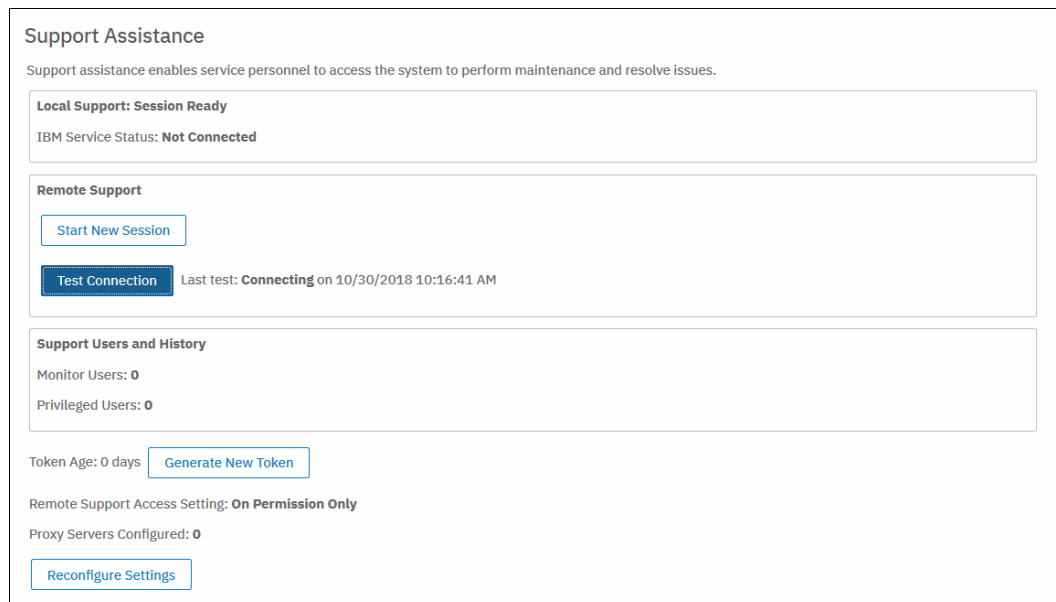


Figure 13-55 Support status and session management

5. A window prompts you for how long you want the tunnel to remain open if no activity occurs by setting a timeout value.

13.8.3 SNMP configuration

SNMP is a standard protocol for managing networks and exchanging messages. The system can send SNMP messages that notify personnel about an event. You can use an SNMP manager to view the SNMP messages that are sent by the system.

You can configure an SNMP server to receive various informational, error, or warning notifications by entering the following information (see Figure 13-56):

- ▶ IP Address

The address for the SNMP server.

- ▶ Server Port

The remote port (RPORT) number for the SNMP server. The RPORT number must be a value of 1 - 65535, where the default is port 162 for SNMP.

- ▶ Community

The SNMP community is the name of the group to which devices and management stations that run SNMP belong. Typically, the default of `public` is used.

- ▶ Event Notifications:

Consider the following points about event notifications:

- Select **Error** if you want the user to receive messages about problems, such as hardware failures, that require prompt action.

Important: Browse to **Recommended Actions** to run the fix procedures on these notifications.

- Select **Warning** if you want the user to receive messages about problems and unexpected conditions. Investigate the cause immediately to determine any corrective action such as a space efficient volume running out of space.

Important: Go to **Recommended Actions** to run the fix procedures on these notifications.

- Select **Info** if you want the user to receive messages about expected events. No action is required for these events.

SNMP	SNMP				
	Add SNMP Server	☰ Actions	↓		
Syslog	Server IP	Error	Warning	Info	Version
	10.10.10.1	✓	✓		3
					Authentication

Figure 13-56 SNMP configuration

To add an SNMP server, select **Actions** → **Add** and complete the Add SNMP Server window, as shown in Figure 13-57. To remove an SNMP server, click the line with the server that you want to remove, and select **Actions** → **Remove**.

The screenshot shows a configuration window titled "Add SNMP Server" with a close button (X) in the top right corner. The window is divided into several sections for configuration:

- Server IP*:** A text input field containing "9.78.43.56".
- Community*:** A text input field containing "public". A "*Required" label is positioned to the right of the field.
- Port*:** A text input field containing "162".
- Events*:** Three checkboxes are present: "Error" (checked), "Warning" (unchecked), and "Info" (unchecked).
- Engine ID:** A text input field with the placeholder text "Enter ID".
- Security Name:** A text input field with the placeholder text "Enter Username".
- Authentication Protocol:** A dropdown menu showing "Select an option".
- Authentication Passphrase (8 characters min.):** A text input field with the placeholder text "Enter Passphrase".
- Privacy Protocol:** A dropdown menu showing "Select an option".
- Privacy Passphrase (8 characters min.):** A text input field with the placeholder text "Enter Passphrase".

At the bottom of the window, there is a "Need Help" link with a question mark icon on the left, and two buttons, "Cancel" and "Add", on the right.

Figure 13-57 Add SNMP Server

Note: The following properties are optional:

- ▶ Engine ID
Indicates the unique identifier (UID) in hexadecimal that identifies the SNMP server.
- ▶ Security Name
Indicates which security controls are configured for the SNMP server. Supported security controls are none, authentication, or authentication and privacy.
- ▶ Authentication Protocol
Indicates the authentication protocol that is used to verify the system to the SNMP server.
- ▶ Privacy Protocol
Indicates the encryption protocol that is used to encrypt data between the system and the SNMP server.
- ▶ Privacy Passphrase
Indicates the user-defined passphrase that is used to verify encryption between the system and SNMP server.

13.8.4 Syslog notifications

The syslog protocol is a standard protocol for forwarding log messages from a sender to a receiver on an IP network. The IP network can be IPv4 or IPv6. The system can send syslog messages that notify personnel about an event.

You can configure a syslog server to receive log messages from various systems and store them in a central repository by selecting **Settings** → **Notifications** → **Syslog**, as shown in Figure 13-58.

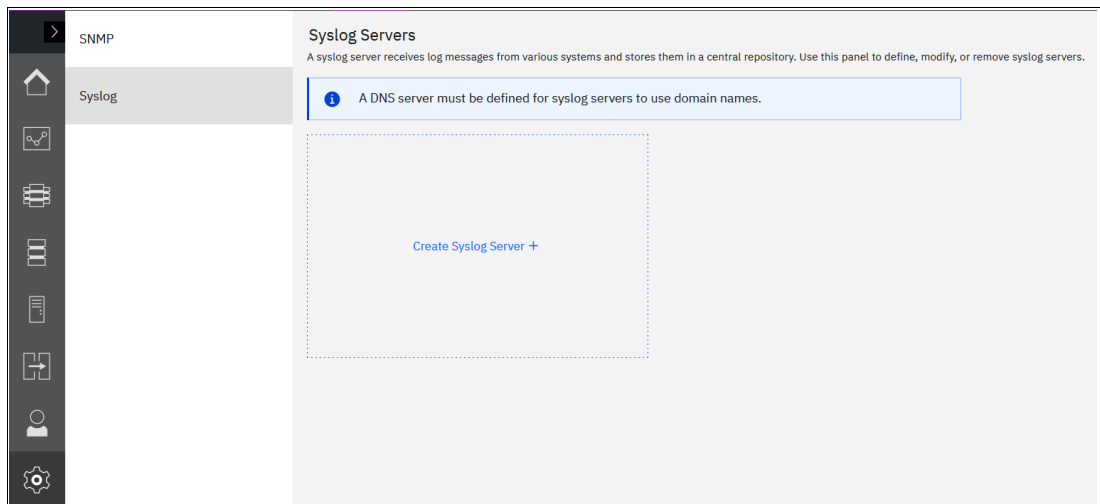


Figure 13-58 Syslog Servers menu

Enter the following information, as shown in Figure 13-59.

Create Syslog Server

IP Address or Domain: 10.10.10.23

Facility: Level 0

Protocol: UDP

Server Port: 541

Notifications:

- Error
- Warning
- Info

Messages:

- CLI
- Login

Buttons: Cancel, Create

Figure 13-59 Syslog configuration

- ▶ IP Address
The IP address for the syslog server.
- ▶ Facility
The facility determines the format for the syslog messages. The facility can be used to determine the source of the message.
- ▶ Protocol
The protocol to be used (UDP or TCP).
- ▶ Server Port
The port to communicate with the syslog server.
- ▶ Notifications
Choose one of the following items for event notifications:
 - Select **Error** if you want the user to receive messages about problems, such as hardware failures, that must be resolved immediately.

Important: Go to **Recommended Actions** to run the fix procedures on these notifications.

- Select **Warning** if you want the user to receive messages about problems and unexpected conditions. Investigate the cause immediately to determine whether any corrective action is necessary.

Important: Go to **Recommended Actions** to run the fix procedures on these notifications.

- Select **Info** if you want the user to receive messages about expected events. No action is required for these events.
- ▶ Messages
Choose one of the following items for messages:
 - **CLI**
Select this option to include any CLI or management GUI operations on the specified syslog servers.
 - **Login**
Select this option to send successful and failed authentication attempts to the specified syslog servers.

13.9 Audit log

The audit log is useful when analyzing past configuration events, especially when trying to determine, for example, how a volume ended up being shared by two hosts, or why the volume was overwritten. The audit log is also included in the `svc_snap` support data to aid in problem determination.

The audit log tracks action commands that are issued through an SSH session, management GUI, or Remote Support Assistance. It provides the following entries:

- ▶ Identity of the user who ran the action command.
- ▶ Name of the actionable command.
- ▶ Timestamp of when the actionable command ran on the configuration node.
- ▶ Parameters that ran with the actionable command.

The following items are not documented in the audit log:

- ▶ Commands that fail are not logged.
- ▶ A result code of 0 (success) or 1 (success in progress) is not logged.
- ▶ Result object ID of node type (for the **addnode** command) is not logged.
- ▶ Views are not logged.

Several specific service commands are not included in the audit log:

- ▶ **dumpconfig**
- ▶ **cpdumps**
- ▶ **cleardumps**
- ▶ **finderr**
- ▶ **dumperrlog**
- ▶ **dumpintervallog**
- ▶ **svcservicetak dumperrlog**
- ▶ **svcservicetask finderr**

Figure 13-60 on page 853 shows the access to the audit log. Click **Audit Log** in the left menu to see which configuration CLI commands were run on the system.

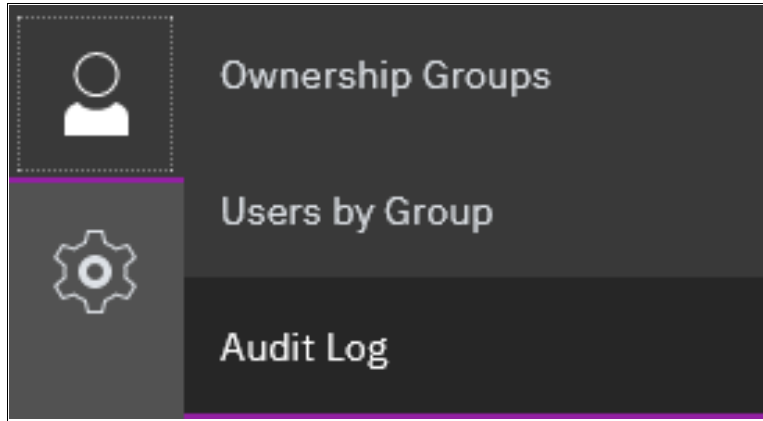


Figure 13-60 Audit Log from the Access menu

Figure 13-61 shows an example of the audit log after a volume is created and mapped to a host.

Date and Time	User Name	Command	Object ID
25/10/2019 17:54:43	JackUser	svctask mkvdiskhostmap -force -gui -host 0 -scsi 1 35	
25/10/2019 17:54:25	JackUser	svctask mkvdisk -gui -iogrp io_grp1 -mdiskgrp 0 -name Redbook...	35
25/10/2019 17:53:46	superuser	svctask startfcmap -gui -prep 1	
25/10/2019 17:52:00	superuser	svctask chfcmap -cleanrate 50 -copyrate 100 -gui 1	
25/10/2019 17:51:53	superuser	svctask chenclosure -gui -managed yes 2	
25/10/2019 17:44:51	superuser	svctask addcontrolenclosure -gui -iogrp 1 -sernum 7822PFG	
25/10/2019 02:00:20	superuser	satask cpfiles -prefix /dumps/svc.config.cron.*_7822PBR-2 -sour...	
25/10/2019 02:00:02	superuser	svctask detectmdisk	
25/10/2019 00:23:49	JackUser	svctask mksnmpserver -community public -error on -gui -info off ...	1
25/10/2019 00:11:04	JackUser	svctask chsra -gui -idletimeout 60 -remotesupport enable	
25/10/2019 00:09:25	JackUser	svctask chsra -enable -gui	
24/10/2019 23:53:31	JackUser	svctask startemail -gui	
24/10/2019 23:53:17	JackUser	svctask mkemailserver -gui -ip 9.71.47.10 -port 25	0
24/10/2019 23:53:14	JackUser	svctask mkemailuser -address callhome1@de.ibm.com -error on...	

Showing 343 entries | Selecting 0 entries

Figure 13-61 Audit log

Changing the view of the Audit Log grid is possible by right-clicking column headings or clicking the sign in the upper right (see Figure 13-62). The grid layout and sorting is under the user's control, so you can view everything in the audit log, sort different columns, and reset the default grid preferences.

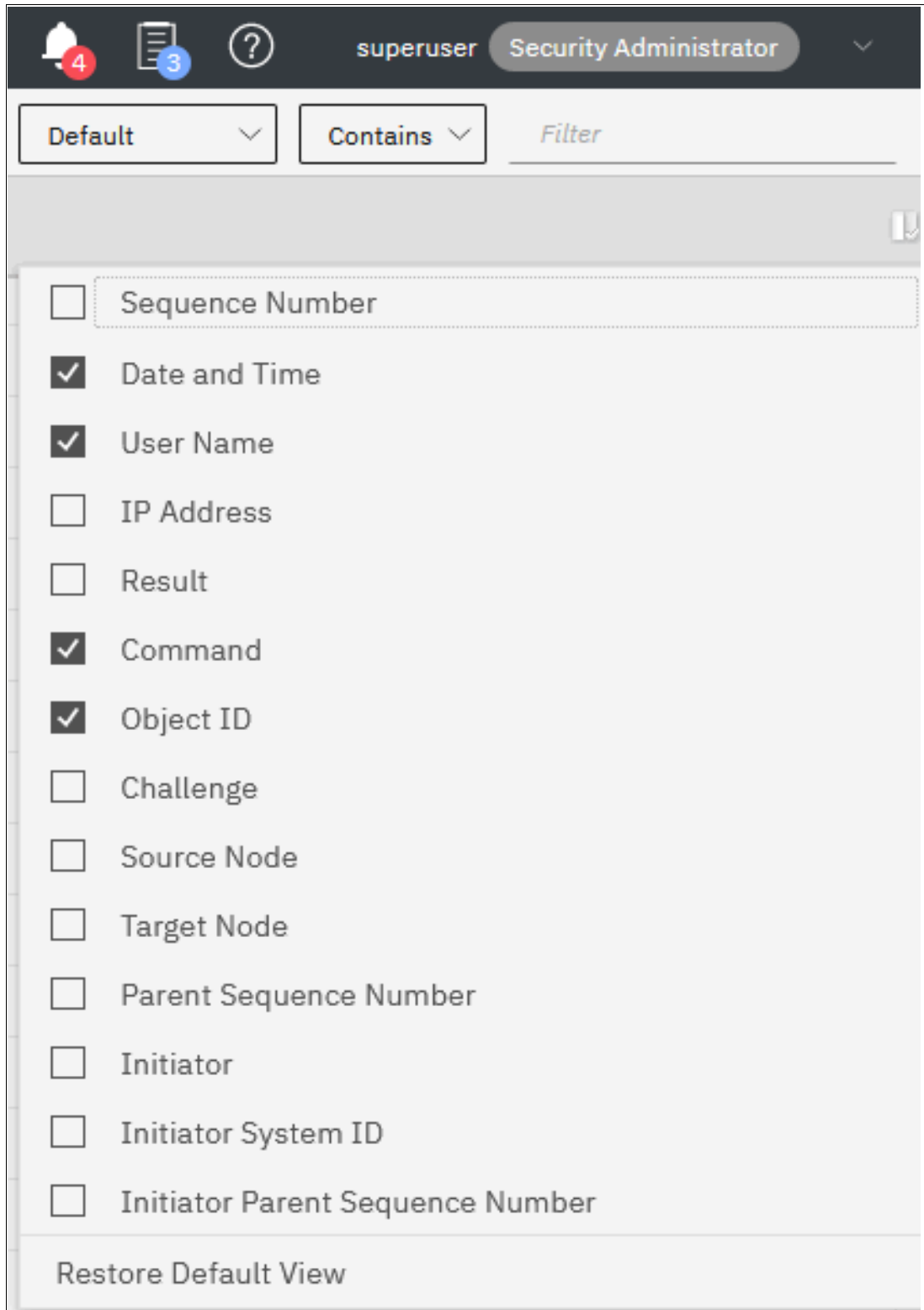


Figure 13-62 How to change audit log column headings

13.10 Collecting support information by using the GUI, CLI, and USB

If you encounter a problem and contact the IBM Support Center, you will be asked to provide a support package. You can collect and upload this package by selecting **Settings** → **Support** menu.

13.10.1 Collecting information by using the GUI

To collect information by using the GUI, complete the following steps:

1. Select **Settings** → **Support** and then the **Support Package** tab (see Figure 13-63).

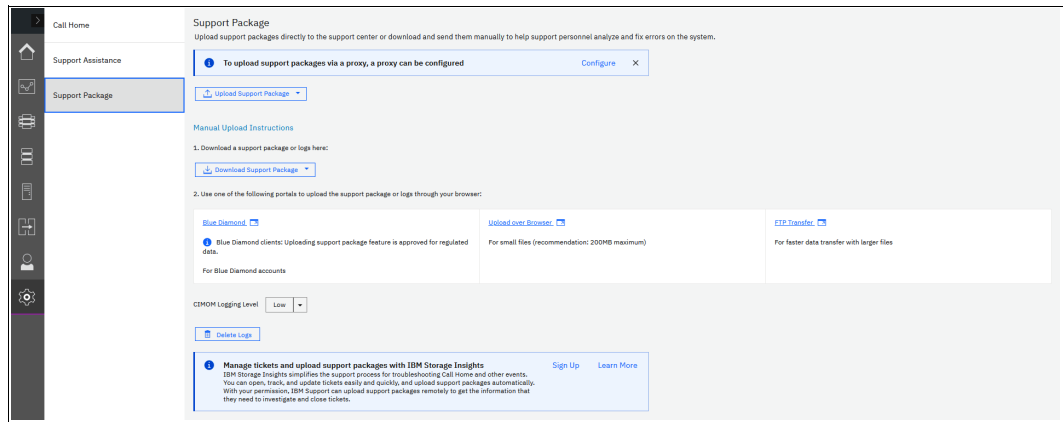


Figure 13-63 Support Package window

2. Click **Upload Support Package** and then **Create New Package and Upload**.

Assuming that the problem that was encountered was an unexpected node restart that logged a 2030 error, collect the default logs and the most recent statesave from each node to capture the most relevant data for support.

Note: When a node unexpectedly restarts, it first dumps its current statesave information before it restarts to recover from an error condition. This statesave is critical for IBM Support to analyze what occurred. Collecting a snap type 4 creates statesaves at the time of the collection, which is not useful for understanding the restart event.

- The Upload Support Package window provides four options for data collection. If you are contacted by IBM Support because your system called home or you manually opened a call with IBM Support, you receive a *Problem Management Record (PMR)* number. Enter that PMR number into the **PMR** field and select the snap type (often referred to as an *option 1, 2, 3, 4 snap*) as requested by IBM Support (see Figure 13-64). In our example, we entered our PMR number, selected **snap type 3 (option 3)** because this option automatically collects the statesaves that were created at the time that the node restarted, and clicked **Upload**.

Tip: To open a service request online, see the [Service requests and PMRs](#).

Upload Support Package

Your system will generate and upload a new package to the IBM support center.

PMR Number: [Don't have PMR?](#)

ppppp,bbb,ccc

Select the type of new support package to generate and upload to the IBM support center:

- Snap Type 1: Standard logs
Contains the most recent logs for the system, including the event and audit logs.
- Snap Type 2: Standard logs plus one existing statesave
Contains all the standard logs plus one existing statesave from any of the nodes in the system.
- Snap Type 3: Standard logs plus most recent statesave from each node
Contains all the standard logs plus each node's most recent statesave.
- Snap Type 4: Standard logs plus new statesaves
Contains all the standard logs and generate a new statesave on each node in the system.

[Need Help](#) Cancel Upload

Figure 13-64 Upload Support Package window

- The procedure to generate the snap on the system, including the most recent statesave from each node canister, starts. This process might take a few minutes (see Figure 13-65).

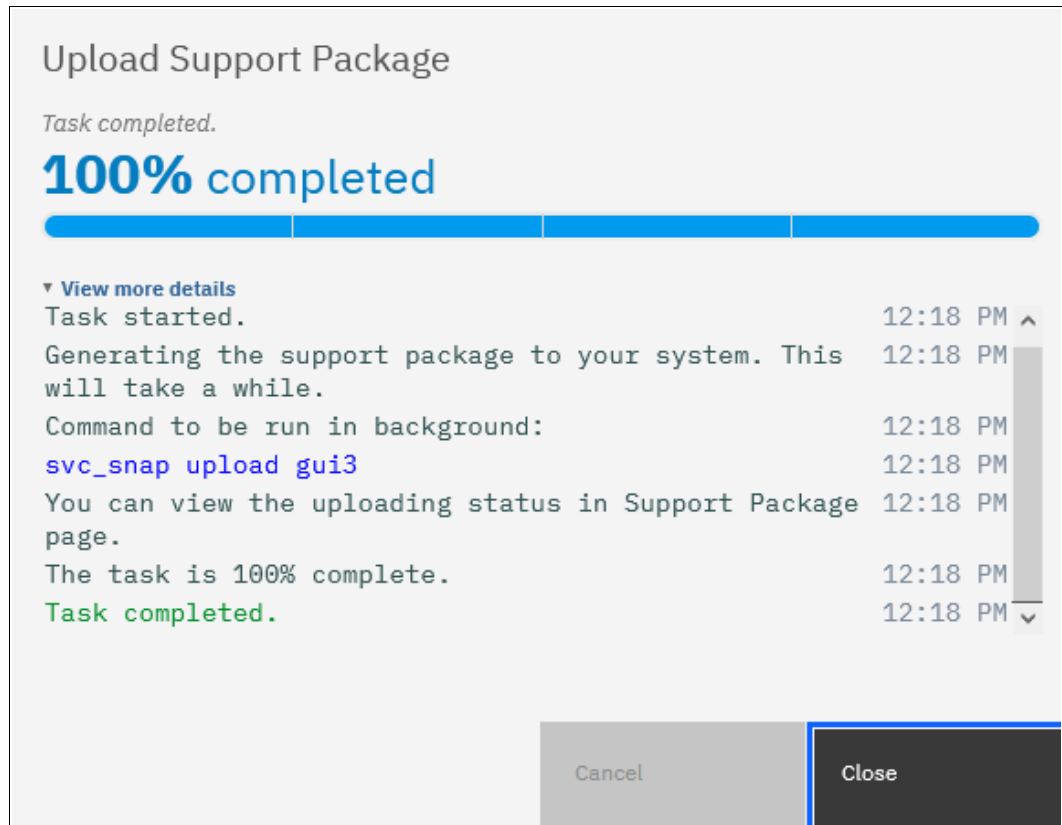


Figure 13-65 Task detail window

The time that it takes to generate the snap and the size of the file that is generated depends mainly on two things: the snap option that you selected, and the size of your system. An option 1 snap takes much less time than an option 4 snap because nothing new must be gathered for an option 1 snap, but an option 4 snap requires the system to collect new statesaves from each node. In an 8-node cluster, this task can take quite some time, so you should always collect the snap option that IBM Support recommends.

Table 13-13 shows the approximate file sizes for each SNAP option.

Table 13-13 Types of snaps

Option	Description	Approximate size (One I/O group, 30 volumes)	Approximate size (Four I/O groups, 250 Volumes)
1	Standard logs	10 MB	340 MB
2	Standard logs plus one existing statesave	50 MB	520 MB
3	Standard logs plus the most recent statesave from each node	90 MB	790 MB
4	Standard logs plus new statesaves	90 MB	790 MB

13.10.2 Collecting logs by using the CLI

The CLI can be used to collect and upload a support package as requested by IBM Support by performing the following steps:

1. Log in to the CLI and run the **svc_snap** command that matches the type of snap that is requested by IBM Support:

- Standard logs (type 1):

```
svc_snap upload pmr=ppppp,bbb,ccc gui1
```

- Standard logs plus one existing statesave (type 2):

```
svc_snap upload pmr=ppppp,bbb,ccc gui2
```

- Standard logs plus most recent statesave from each node (type 3):

```
svc_snap upload pmr=ppppp,bbb,ccc gui3
```

- Standard logs plus new statesaves:

```
svc_livedump -nodes all -yes  
svc_snap upload pmr=ppppp,bbb,ccc gui3
```

In this example, we collect the type 3 (option 3) and have it automatically uploaded to the PMR number that is provided by IBM Support, as shown in Example 13-9.

Example 13-9 The svc_snap command

```
IBM FlashSystem 7200:superuser>svc_snap upload pmr=12345,000,866 gui3  
IBM FlashSystem 7200:superuser>
```

If you do not want to automatically upload the snap to IBM, do not specify the **upload pmr=ppppp,bbb,ccc** part of the commands. When the snap creation completes, it creates a file name that uses the following format:

```
/dumps/snap.<panel_id>.YYMMDD.hhmmss.tgz
```

It takes a few minutes for the snap file to complete (longer if statesaves are included).

The generated file can then be retrieved from the GUI by selecting **Settings** → **Support** → **Manual Upload Instructions** → **Download Support Package**, and then clicking **Download Existing Package**, as shown in Figure 13-66 on page 859.

Manual Upload Instructions

1. Download a support package or logs here:

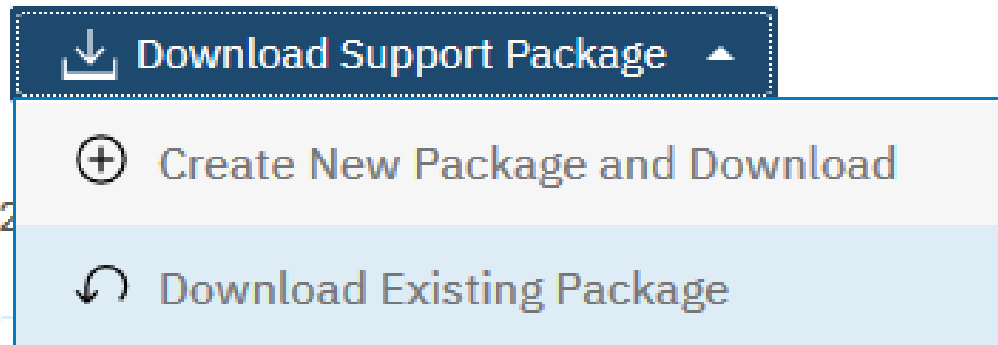


Figure 13-66 Downloaded Existing Package

2. Click in the **Filter** box and enter snap to see a list of snap files, as shown in Figure 13-67. Find the exact name of the snap that was generated by running the `svc_snap` command that was run earlier. Select that file, and click **Download**.

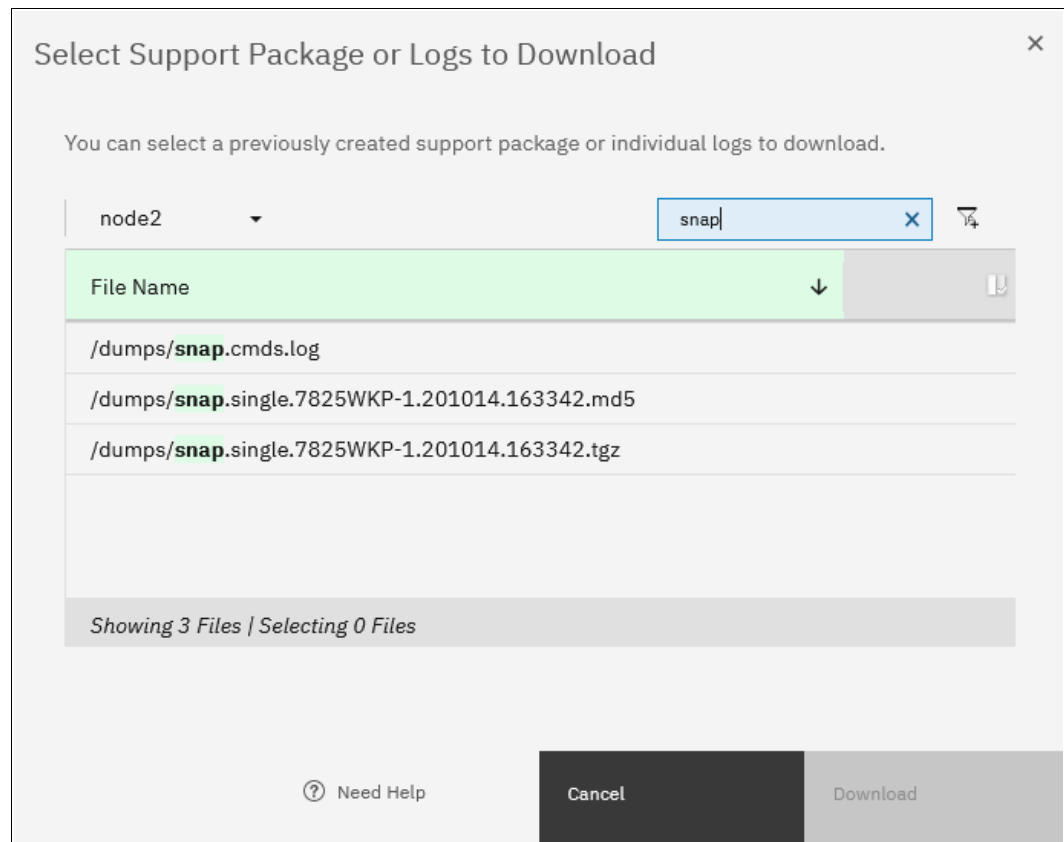


Figure 13-67 Filtering on snap to download

3. Save the file to a folder of your choice on your workstation.

13.10.3 Collecting logs by using a USB flash drive

As a backup, in case there is no connectivity to CLI and GUI (rare), it is possible to get a snap from a node from the USB ports on the rear.

Note: This procedure collects a single snap from the node canister, not a cluster snap. It is useful for determining the state of the node canister.

When a USB flash drive is plugged into a node canister, the canister code searches for a text file that is named `satask.txt` in the root directory. If the code finds the file, it attempts to run a command that is specified in the file. When the command completes, a file that is called `satask_result.html` is written to the root directory of the USB flash drive. If this file does not exist, it is created. If it exists, the data is inserted at the start of the file. The file contains the details and results of the command that was run and the status and the configuration information from the node canister. The status and configuration information matches the detail that is shown on the service assistant home page windows.

To collect a snap, complete the following steps:

1. Ensure that your USB drive is formatted with an FAT32 file system on its first partition.
2. Create a text (`.txt`) file on the USB flash drive in the root directory called `satask.txt` (case-sensitive).
3. In the `satask.txt` file, write **satask snap** and save the file.
4. Put the USB into one of the USB ports on the rear of the canister and wait for a short time. The fault LED on the node canister flashes when the USB service action is being completed. When the fault LED stops flashing, it is safe to remove the USB flash drive.
5. Unplug the USB drive from the system and plug it into your workstation. If the procedure was successful, the `satask.txt` file was deleted and you have a `satask_result.html` file and a single snap from the node canister. This snap can be uploaded to the IBM Support Center, as shown in 13.10.4, “Uploading files to the IBM Support Center” on page 860.

Note: If there was a problem with the procedure, the html file still is generated, and reasons why the procedure did not work are listed in it.

13.10.4 Uploading files to the IBM Support Center

If you chose to not have the system upload the support package automatically, it can still be uploaded for analysis from the Enhanced Customer Data Repository (ECuRep). Any uploads should be associated with a specific PMR. The PMR is also known as a *service request* and is a mandatory requirement when uploading.

To upload the information, complete the following steps:

1. Using a web browser, go to [Enhanced Customer Data Repository \(ECuRep\)](#) (see Figure 13-68).

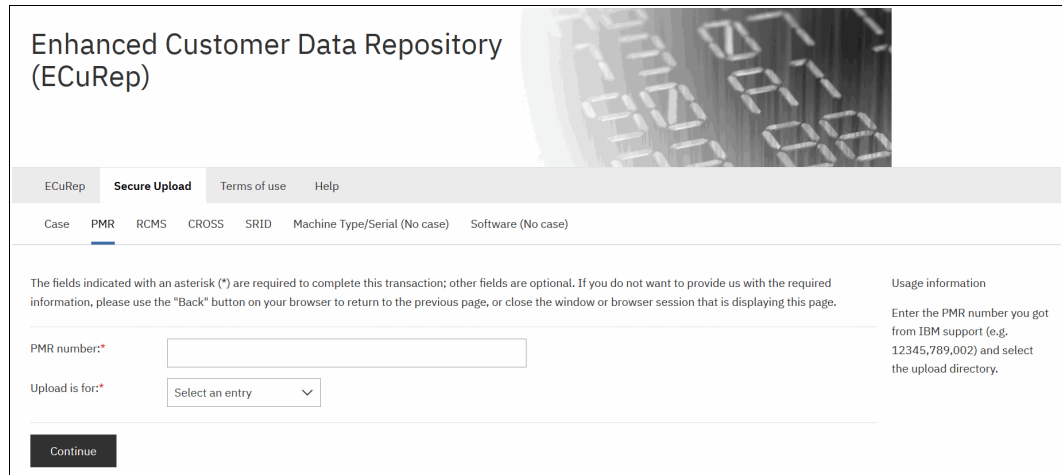


Figure 13-68 ECuRep details

2. Complete the required fields:

- The **PMR number** that is provided by IBM Support for your specific case. This number uses the format of ppppp,bbb,ccc; for example, 12345,000,866, which uses a comma (,) as a separator.
- **Upload is for:** Select **Hardware** from the drop-down menu.

Although the **Email address** field is not mandatory, it is a best practice to enter your email address so that you are automatically notified of a successful or unsuccessful upload.

3. When the form is completed, click **Continue** to open the input window (see Figure 13-69).

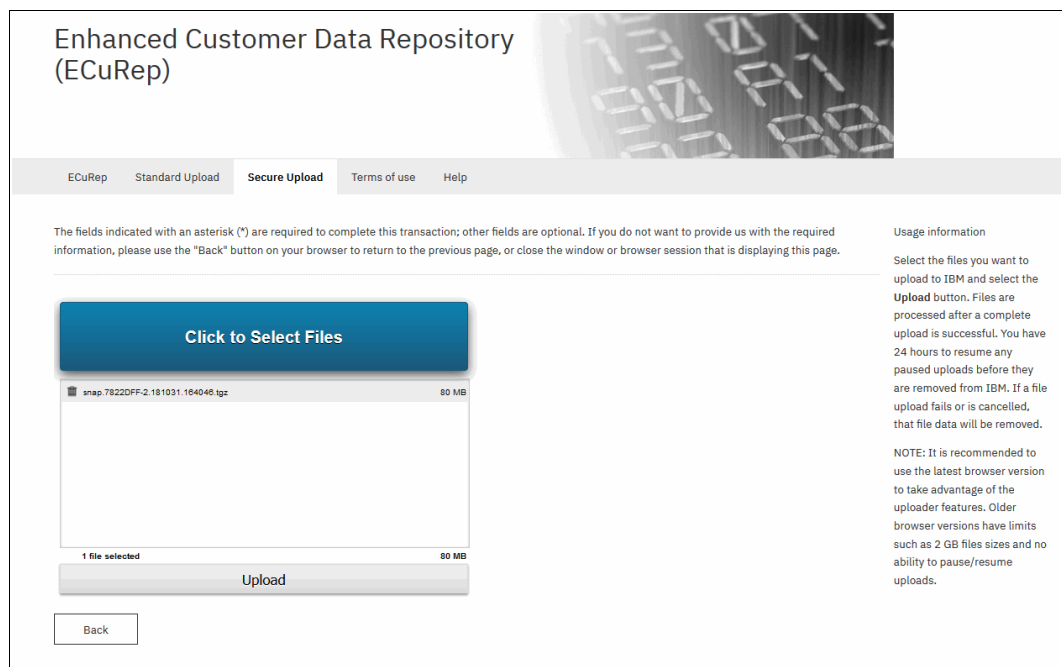


Figure 13-69 ECuRep File upload

4. Select one or more files, click **Upload** to continue, and follow the directions.

13.11 Service Assistant Tool

The SAT is a web-based GUI that is used to service individual node canisters, primarily when a node has a fault and is in a service state. A node is not an active part of a clustered system while it is in service state.

Typically, the system is configured with the following IP addresses:

- ▶ One service IP address for each of control canisters.
- ▶ One cluster management IP address, which is set when the cluster is created.

The SAT is available even when the management GUI is not accessible. The following information and tasks can be accomplished with the SAT:

- ▶ Status information about the connections and the node canister
- ▶ Basic configuration information, such as configuring IP addresses
- ▶ Service tasks, such as restarting the Common Information Model Object Manager (CIMOM) and updating the WWNN
- ▶ Details about node error codes
- ▶ Details about the hardware, such as IP addresses and Media Access Control (MAC) addresses

The SAT GUI is available by using a service assistant IP address that is configured on each IBM FlashSystem node. It can also be accessed through the cluster IP addresses by appending `/service` to the cluster management IP.

It is also possible to access the SAT GUI of the config node if you enter the Uniform Resource Locator (URL) of the service IP address of the config node into any web browser and click **Service Assistant Tool** (see Figure 13-70 on page 863).

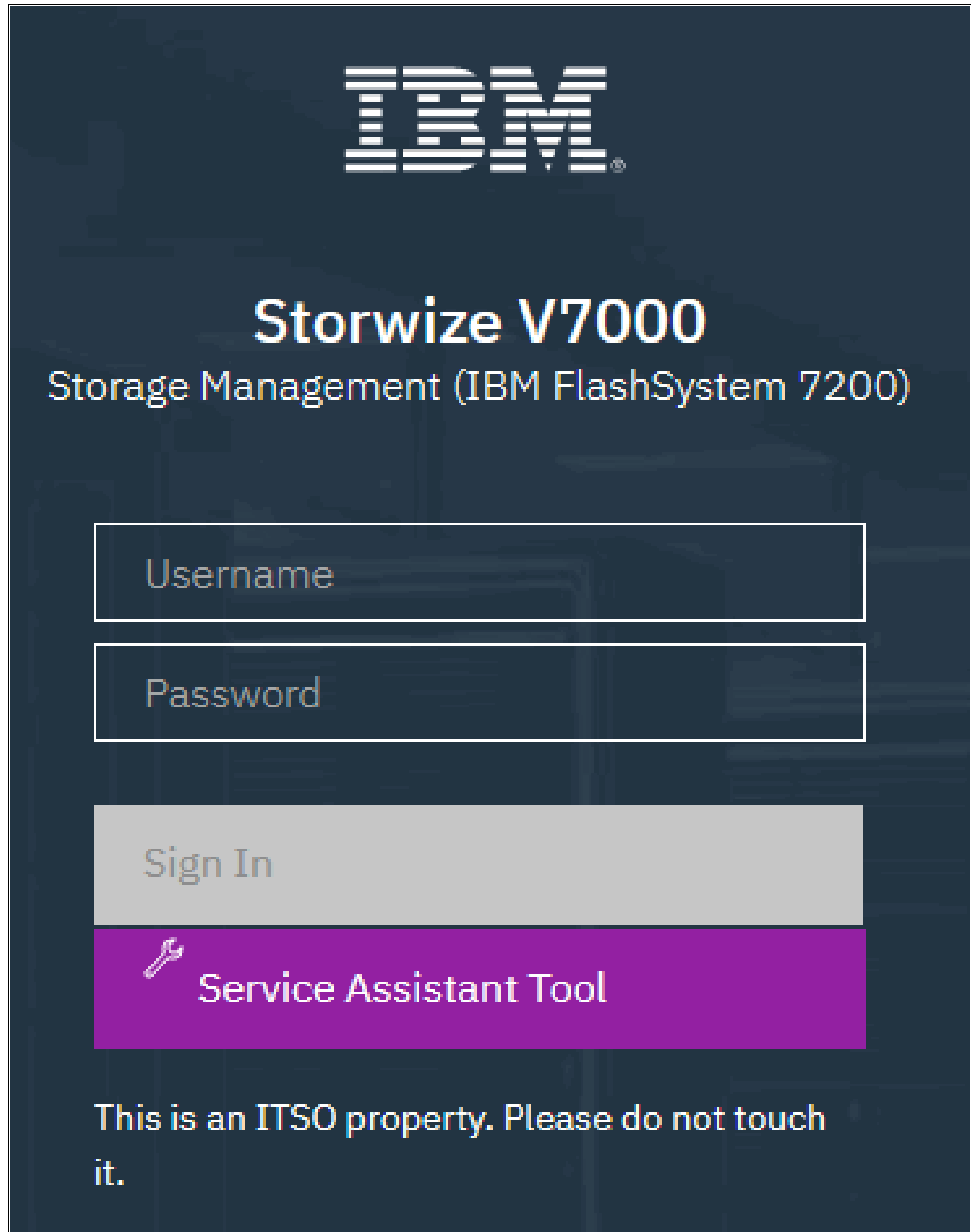


Figure 13-70 Service Assistant Tool login

If the clustered system is down, the only method of communicating with the node canisters is through the SAT IP address directly. Each node can have a single service IP address on Ethernet port 1, which should be configured on all nodes of the cluster.

To open the SAT GUI, enter one of the following URLs into a web browser:

- ▶ Enter `http(s)://<cluster IP address of your cluster>/service`.
- ▶ Enter `http(s)://<service IP address of a node>/service`.
- ▶ Enter `http(s)://<service IP address of config node>` and click **Service Assistant Tool**.

To access the SAT, complete the following steps:

1. If you are accessing SAT by using *cluster IP address/service*, the configuration node canister SAT GUI login window opens. Enter the Superuser Password, as shown in Figure 13-71.

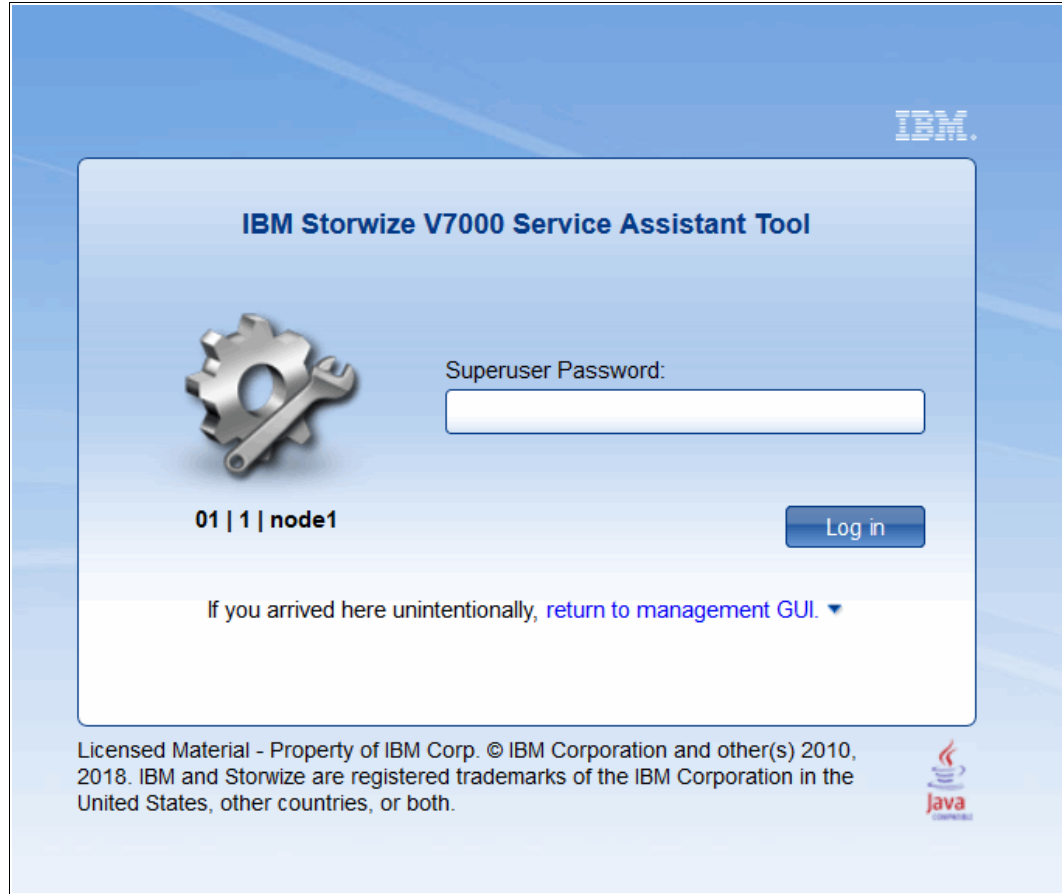


Figure 13-71 Service Assistant Tool Login GUI

2. After you are logged in, you see the Service Assistant Home window, as shown in Figure 13-72. The SAT can view the status and run service actions on other nodes in addition to the node to which the user is logged in.

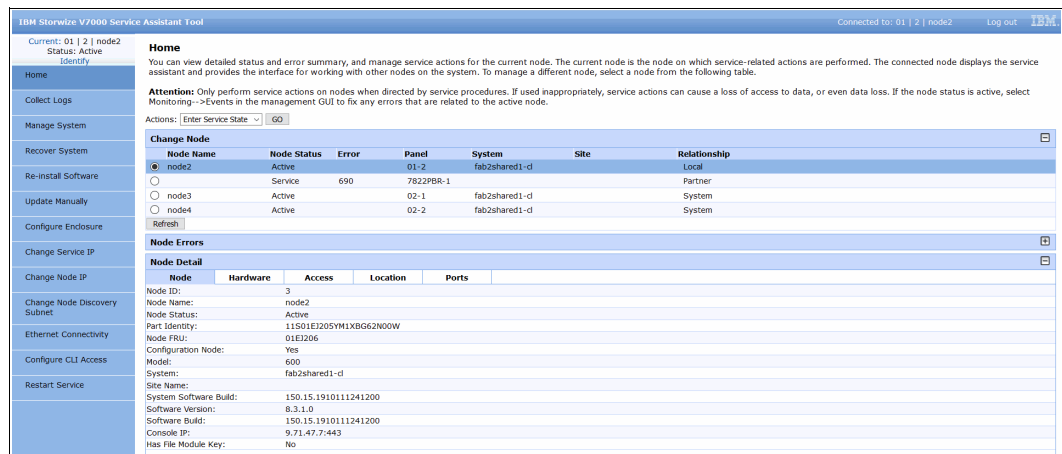


Figure 13-72 Service Assistant Tool GUI

- The current node canister is displayed in the upper left corner of the GUI. As shown in Figure 13-72 on page 864, this is node2. Select the node that you want in the **Change Node** section of the window. You see the details in the upper left change to reflect the selected node canister.

Note: The SAT GUI provides access to service procedures and shows the status of the node canisters. These procedures should be carried out only if you are directed to do so by IBM Support.

For more information about how to use the SAT, see [IBM Documentation](#).

13.12 IBM Storage Insights monitoring

With IBM Storage Insights, you can optimize your storage infrastructure by using this cloud-based storage management and support platform with predictive analytics.

The monitoring capabilities that IBM Storage Insights provides are useful for things like capacity planning, workload optimization, and managing support tickets for ongoing issues.

After you add your systems to IBM Storage Insights, you see the Dashboard, where you can select a system that you want to see the overview for, as shown in Figure 13-73.

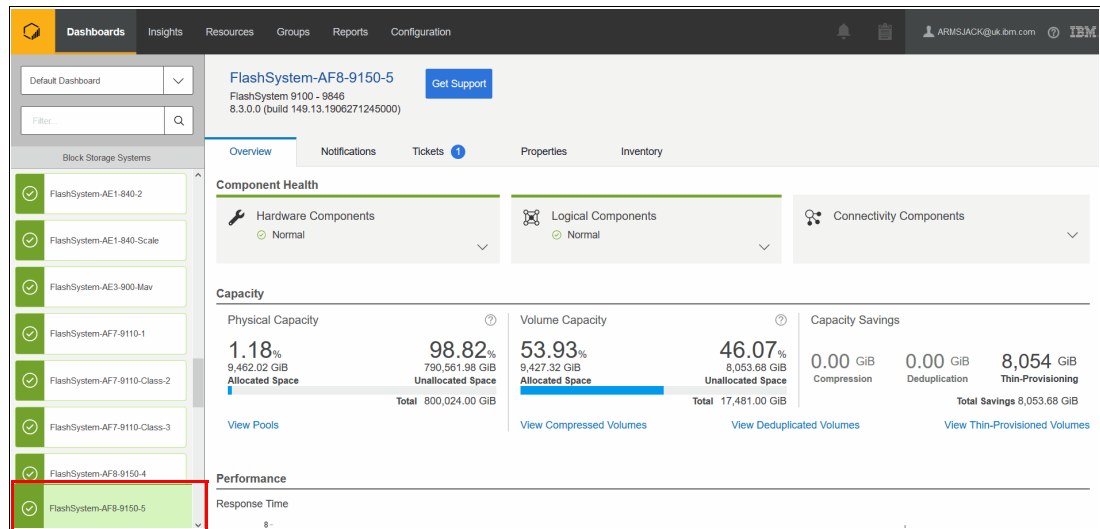


Figure 13-73 IBM Storage Insights System overview

Component health is shown at the upper center of the window. If there is a problem with one of the Hardware, Logical or Connectivity components, errors are shown here, as shown in Figure 13-74.

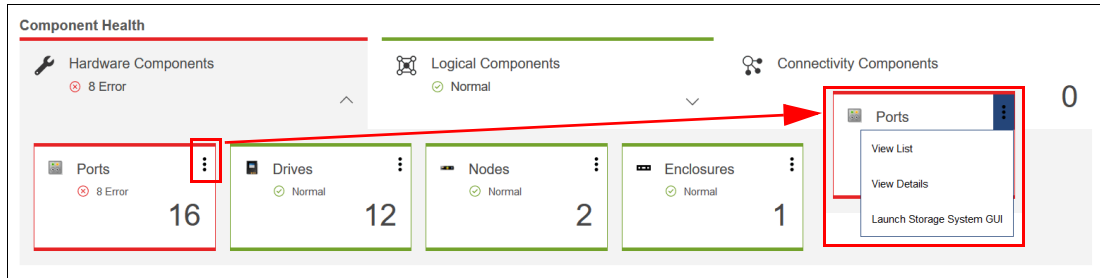


Figure 13-74 Component Health overview

The error entries can be expanded to obtain more details by selecting the three dots at the upper right corner of the component that has an error and then selecting **View Details**. The relevant part of the more detailed System View opens, and what you see depends on which component has the error, as shown in Figure 13-75.

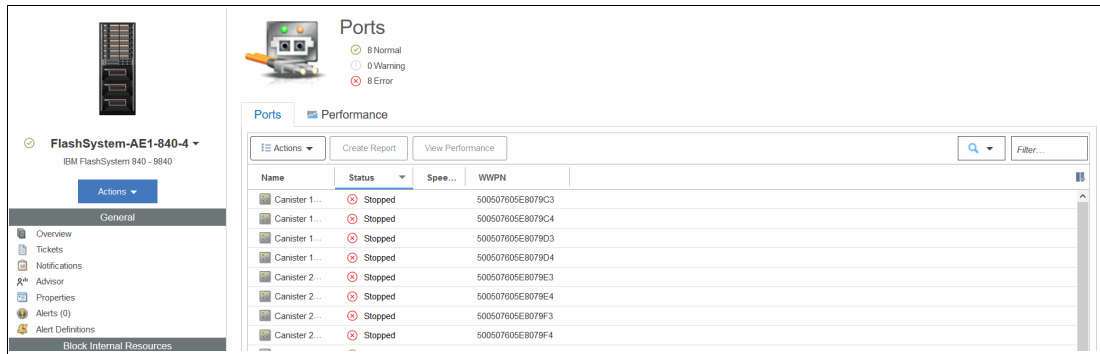


Figure 13-75 Ports in error

From here, it is now obvious which components have the problem and exactly what is wrong with them, so now you can log a support ticket with IBM if necessary.

13.12.1 Capacity monitoring

You can see key statistics such as Physical Capacity, Volume Capacity, and Capacity Savings, depending on what is configured after you select a system, as shown in Figure 13-76.

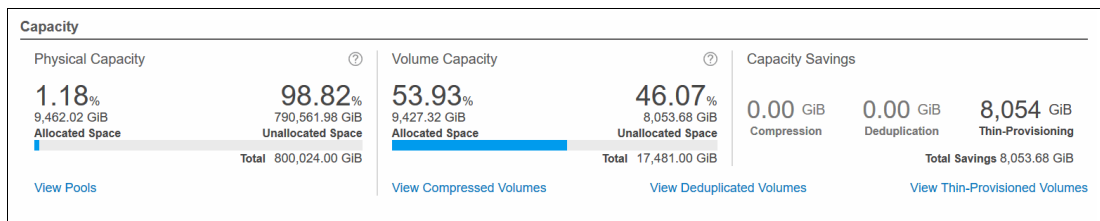


Figure 13-76 Capacity area of the IBM Storage Insights system overview

In the Capacity view, the user can click **View Pools**, **View Compress Volumes**, **View Deduplicated Volumes**, and **View Thin-Provisioned Volumes**. Clicking any of these items takes the user to the detailed system view for the selection option. From there, you can click **Capacity** to get a historical view of how the system capacity changed over time, as shown in Figure 13-77. At any time, the user can select the timescale, resources, and metrics to be displayed on the graph by clicking any options around the graph.

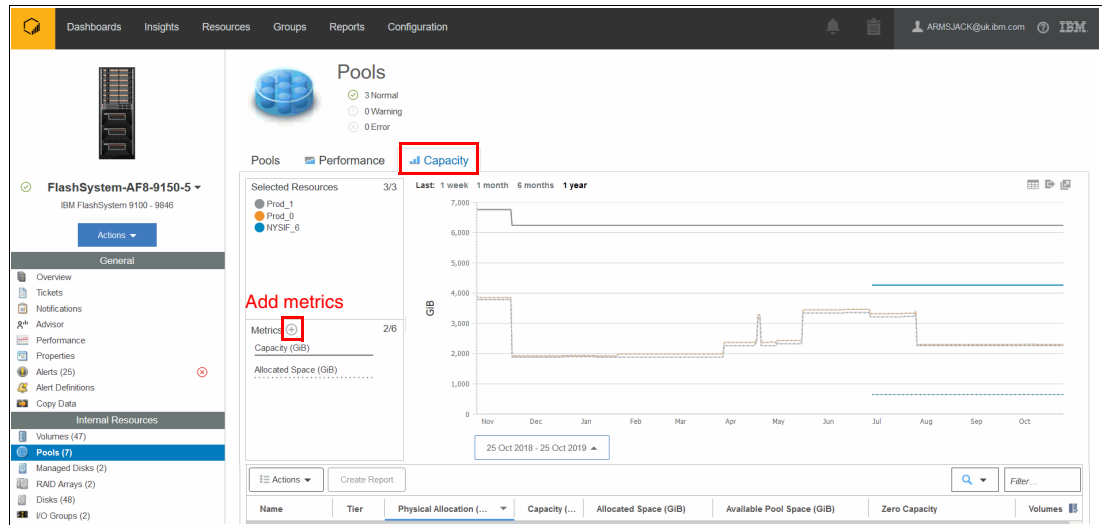


Figure 13-77 IBM Storage Insights Capacity view

If you scroll down below the graph, you find a list view of the selected option. In this example, we selected **View Pools** so the configured pools are shown with the relevant key capacity metrics, as shown in Figure 13-78. Double-clicking a pool in the table display the properties for it.

Name	Tier	Physical Allocation (...)	Capacity (...)	Allocated Space (GiB)	Available Pool Space (GiB)	Zero Capacity	Volumes
Prod_0	Tier 1	41 %	6,239.00	2,304.79	3,889.00	None	232
Prod_1	Tier 0	40 %	6,239.00	2,262.50	3,731.00	None	232
NYSIF_0		15 %	4,264.00	650.00	3,610.00	None	6

Showing 3 items | Selected 0 items | 25 Sep 2019 00:00:00 – 25 Oct 2019 21:51:48 | Refreshed a few moments ago

Figure 13-78 Pools list view

13.12.2 Performance monitoring

From the system overview, you can scroll down and see the three key performance statistics for your system, as shown in Figure 13-79. For the Performance overview, these statistics are aggregated across the whole system, and you cannot drill down by Pool, Volume, or other items.

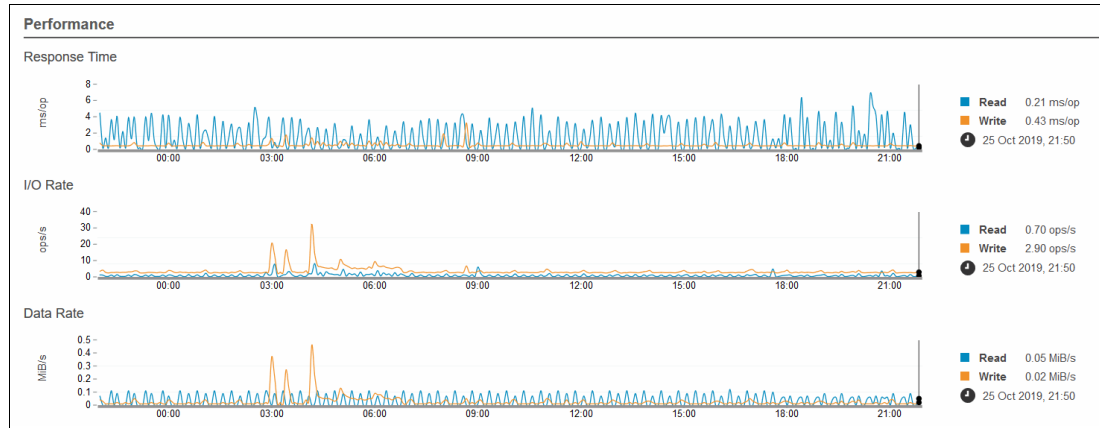


Figure 13-79 System overview: Performance

To view more detailed performance statistics, enter the system view again, as described in 13.12.1, “Capacity monitoring” on page 866.

For this performance example, we select **View Pools**, and then select **Performance** from the System View pane, as shown in Figure 13-80.

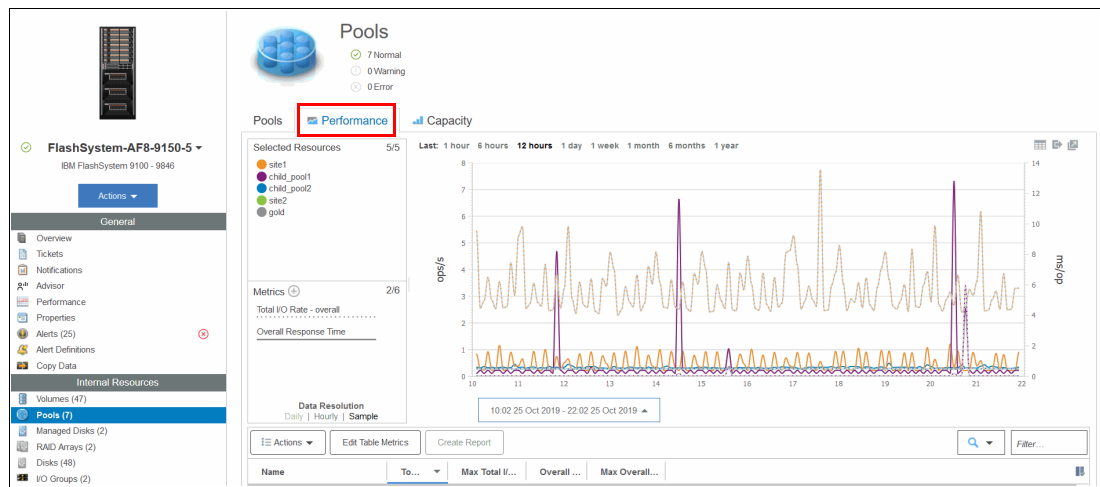


Figure 13-80 IBM Storage Insights: Performance view

It is possible to customize what can be seen on the graph by selecting the metrics and resources. In Figure 13-81, the Overall Response Time for one pool over a 12-hour period is displayed.

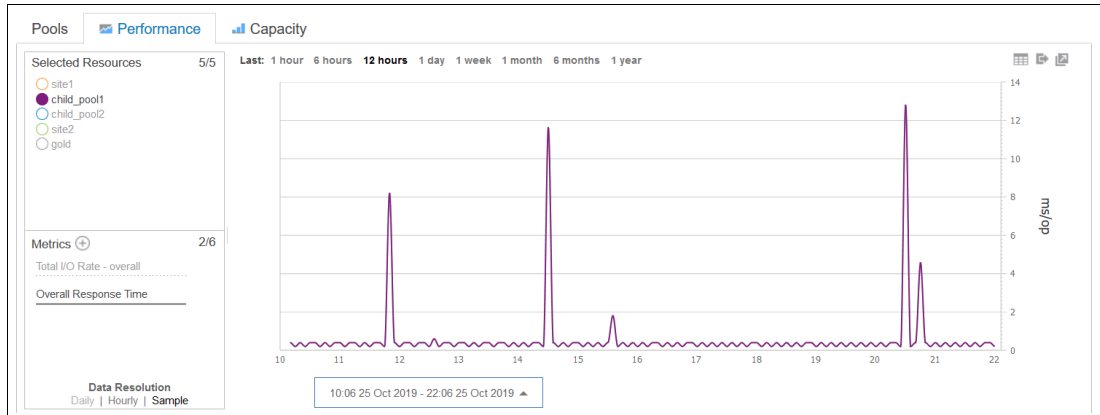


Figure 13-81 Filtered performance graph

Scrolling down the graph, the Performance List view is visible, as shown in Figure 13-82. Metrics can be selected by clicking the filter button at the right of the column headers. If you select a row, the graph is filtered for that selection only. Multiple rows can be selected by holding down the Shift or Ctrl keys.

Name	Total I/O Rate -...	Max Total I/O Rate - overall (o...	Overall Response Time (ms/op)	Max Overall Response T
site1	3.28	7.74	0.69	
child_pool2	0.28	0.31	0.59	
child_pool1	0.04	3.40	3.13	
gold	0.00	0.00	0.00	
site2	0.00	0.00	0.00	0.00

Figure 13-82 Performance List View

13.12.3 Logging support tickets by using IBM Storage Insights

With IBM Storage Insights, you can log existing support tickets that greatly complement the enhanced monitoring opportunities that the software provides. When an issue is detected and you want to engage IBM Support, complete the following steps:

1. Select the system to open the System Overview and click **Get Support**, as shown in Figure 13-83.

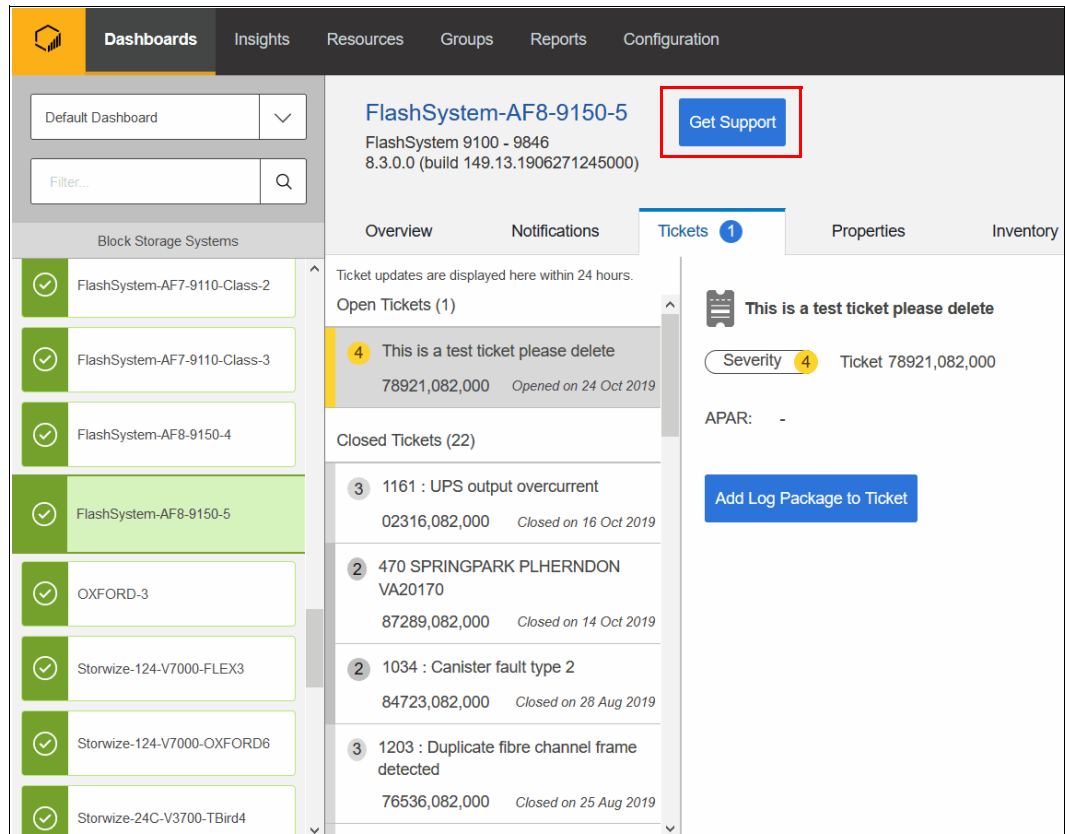


Figure 13-83 Get Support

A window opens where you can create a ticket or update an existing ticket, as shown in Figure 13-84.

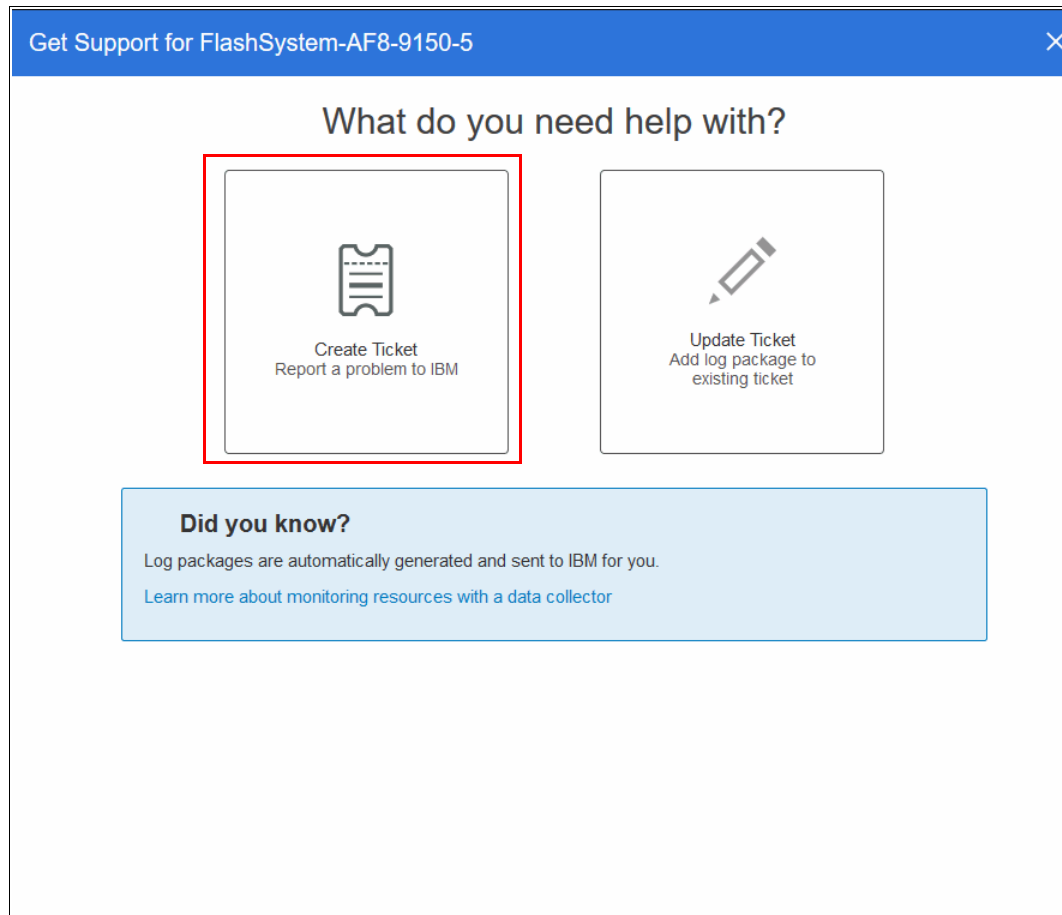


Figure 13-84 Get Support window

2. Select **Create Ticket**, and the ticket creation wizard opens. Details of the system are automatically populated, including the customer number, as shown in Figure 13-85. Select **Next**.

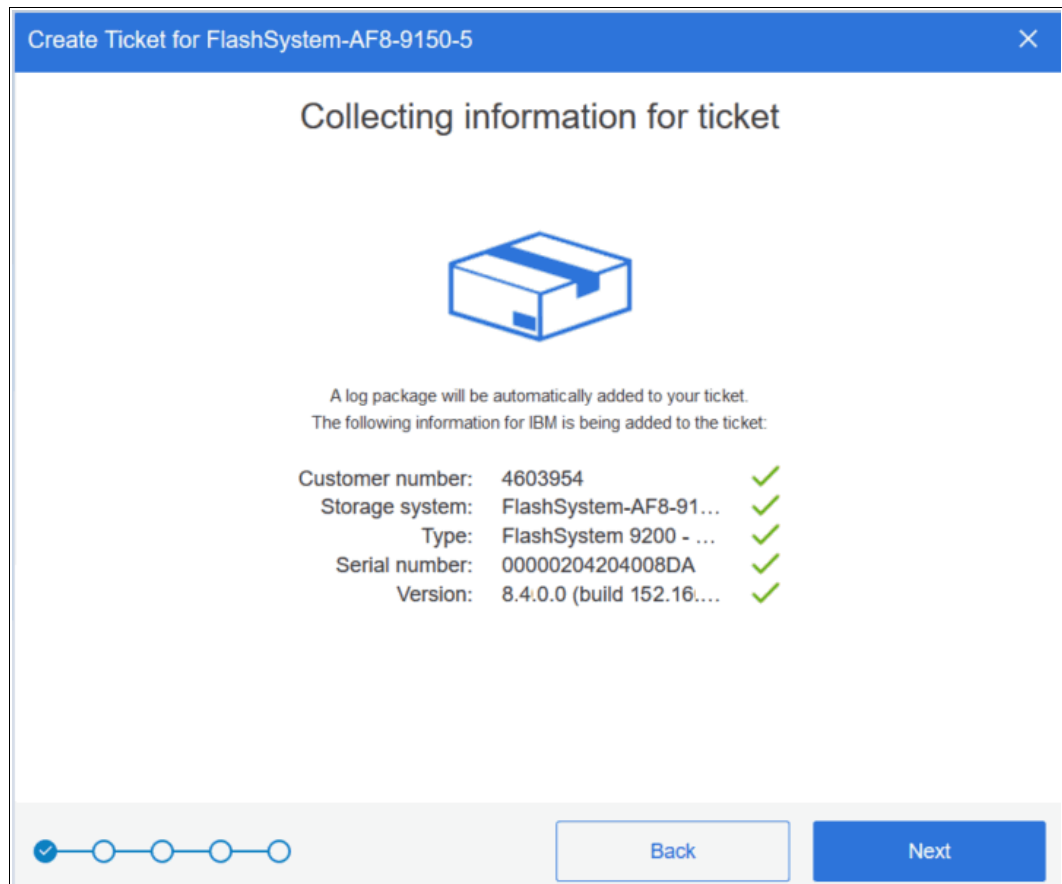


Figure 13-85 Create Ticket wizard

3. You can enter relevant details about your problem to the ticket, as shown in Figure 13-86. It is also possible to attach images or files to the ticket, such as PuTTY logs and screen captures. Once done, select **Next**.

The screenshot shows a window titled "Create Ticket for FlashSystem-AF8-9150-5" with a close button (X) in the top right corner. The main heading is "Add a note or attachment".

There are two text input fields:

- The first field contains the text "Ports are offline" and has a character count of "55" in the top right corner. Below it is a hint: "Hint: Include what happened and the error code, if any."
- The second field contains the text "I have seen on storage insights that I have storage ports offline. Please assist." and has a character count of "2919" in the top right corner. Below it is a hint: "Hint: Include the time the problem or error occurred, the affected resources, and details of any maintenance or other activities that occurred before the problem."

Below the text fields, there are two options for attaching files:

- A blue button labeled "Browse" under the heading "Attach Image or File:".
- A dashed blue box containing an upward-pointing arrow and the text "Drag file here".

Between these two options is the text "OR".

At the bottom left, there is a progress indicator consisting of five circles in a row. The first two circles are filled with a checkmark, and the last three are empty.

At the bottom right, there are two buttons: a light blue "Back" button and a dark blue "Next" button.

Figure 13-86 Add a note or attachment window

4. You can select a severity for the ticket. Examples of what severity you should select are shown in Figure 13-87. Because in our example there are storage ports offline with no impact, we select severity 2 because we lost redundancy.

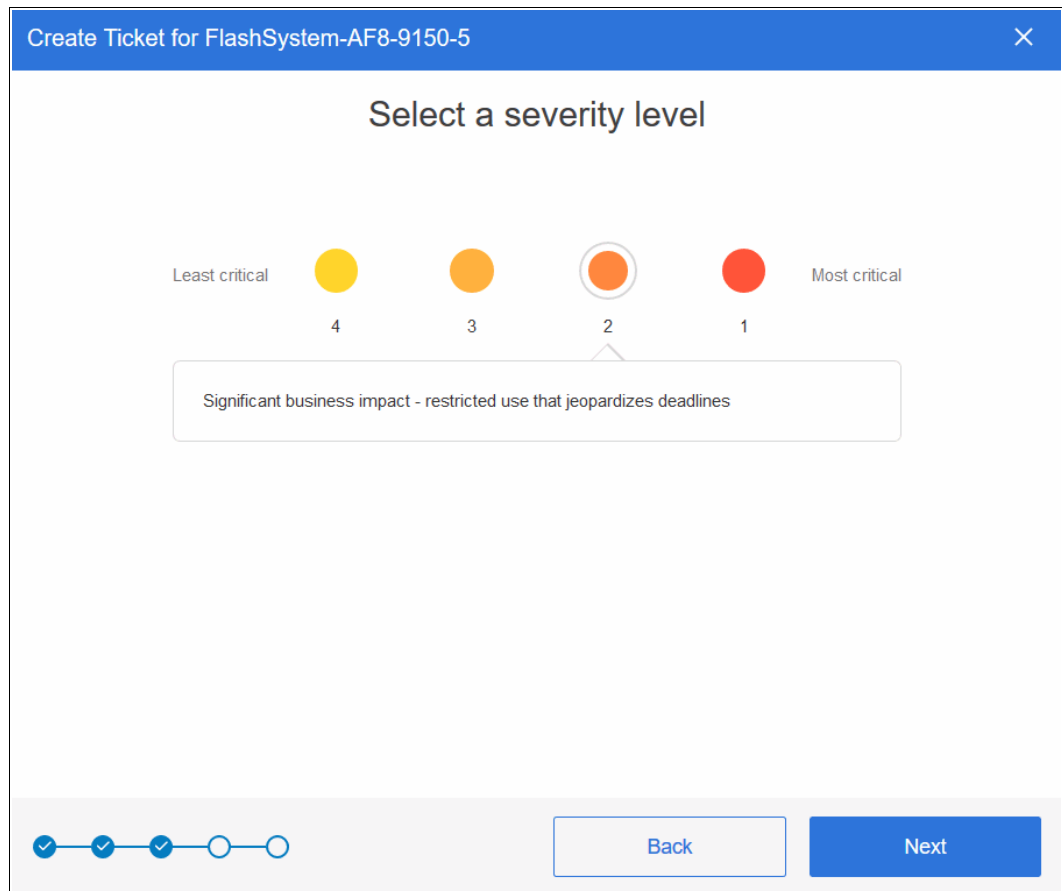


Figure 13-87 Selecting a Severity Level window

5. Choose whether this is a hardware or a software problem. Select the relevant option (for this example, the offline ports are likely caused by a physical layer hardware problem). Once done, click **Next**.

6. Review the details of the ticket that will be logged with IBM, as shown in Figure 13-88. Contact details must be entered so that IBM Support can respond to the correct person. You also must choose which type of logs should be attached to the ticket. For more information about the types of snap, see Table 13-13 on page 857.

Create Ticket for FlashSystem-AF8-9150-5

Review the ticket

Problem summary: Ports are offline

Description: I have seen on storage insights that I have storage ports offline. ...

Severity level: 2 Significant business impact - restricted use that jeopardizes ...

Type of problem: Hardware

Contact name: IBM Redbooks

Contact email: redbooks@ibm.com

Contact phone: 0000000000

Customer number: 4603954 United States

Storage system: FlashSystem-AF8-9250-5

Type: FlashSystem 9200 - 9848

Serial number: 00000204204008DA

Version: 8.4.0.0 (build 152.16.2009271245000)

Log package: Type 3: Standard logs and the most recent state save log from each node

Back Create Ticket

Figure 13-88 Review the ticket window

7. Once done, select **Create Ticket**. A confirmation window opens, as shown in Figure 13-89, and IBM Storage Insights automatically uploads the snap to the ticket when it is collected.

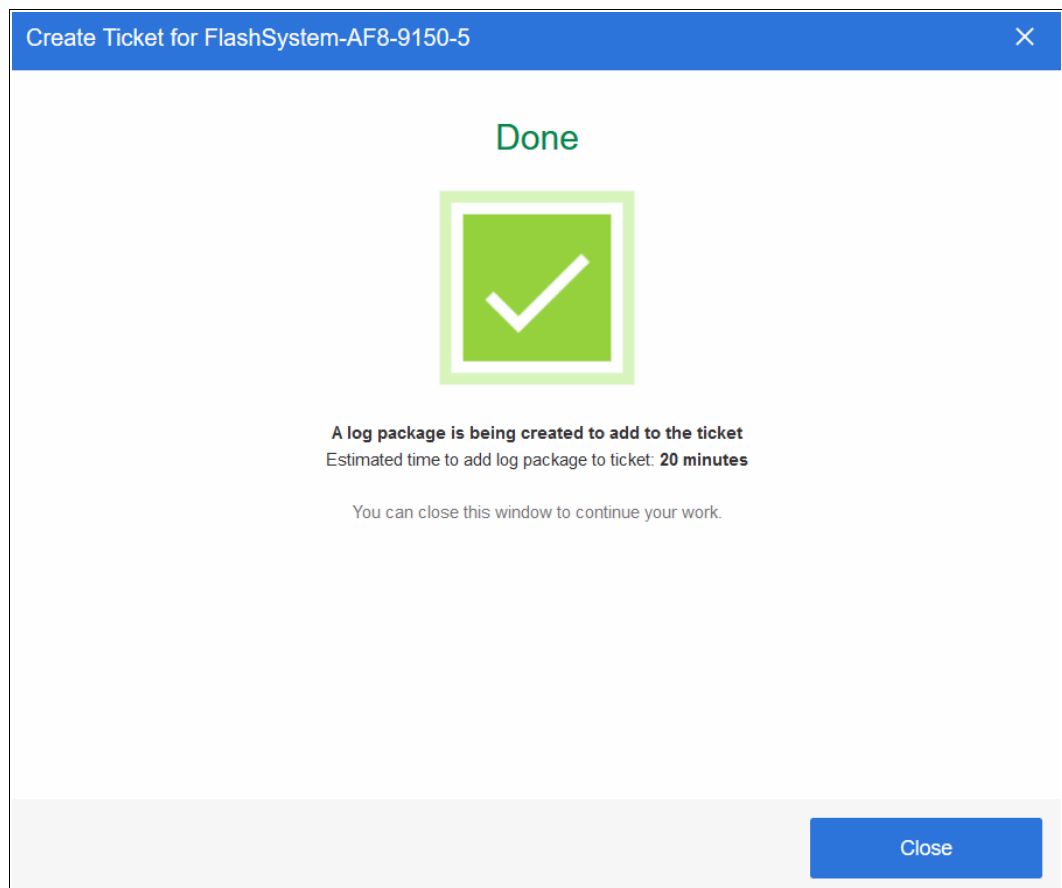


Figure 13-89 Ticket Creation confirmation window

13.12.4 Managing existing support tickets by using IBM Storage Insights and uploading logs

With IBM Storage Insights, you can track existing support tickets and upload logs to them. To do so, complete the following steps:

1. From the System Overview window, select **Tickets**, as shown in Figure 13-90.

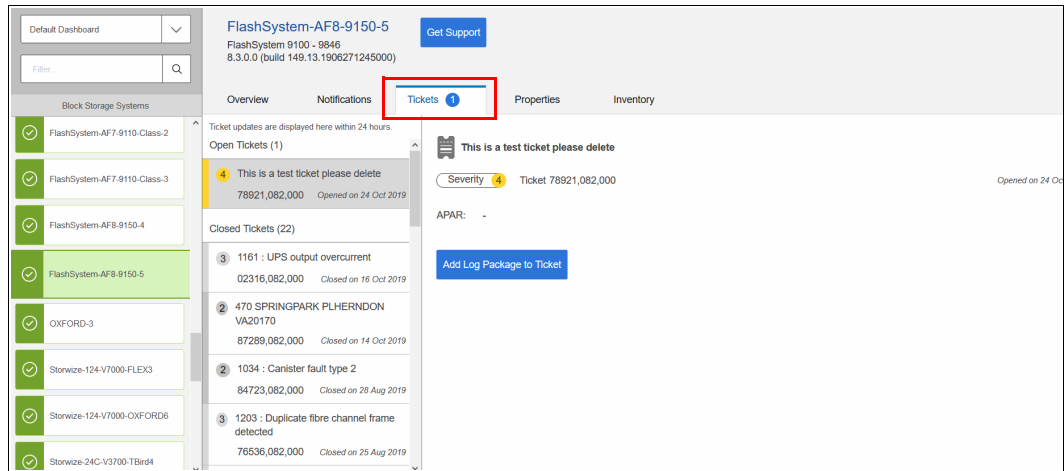


Figure 13-90 View Tickets

In this window, you see a large history of support tickets that were logged through IBM Storage Insights for the system. Tickets that are not currently open are listed under **Closed Tickets**, and currently open tickets are listed under **Open Tickets**.

2. To quickly add logs to a ticket without having to browse to the system GUI or use IBM ECuRep, click **Add Log Package to Ticket**. A window opens that guides you through the process, as shown in Figure 13-91. You can select which type of log package you want and add a note to the ticket with the logs.

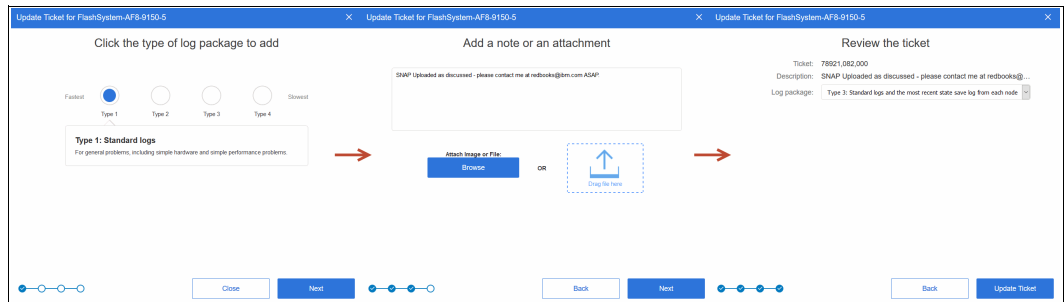


Figure 13-91 Adding a log package to the ticket

3. After clicking **Update Ticket**, a confirmation opens, as shown in Figure 13-92. You can exit the wizard. IBM Storage Insights runs in the background to gather the logs and upload them to the ticket.

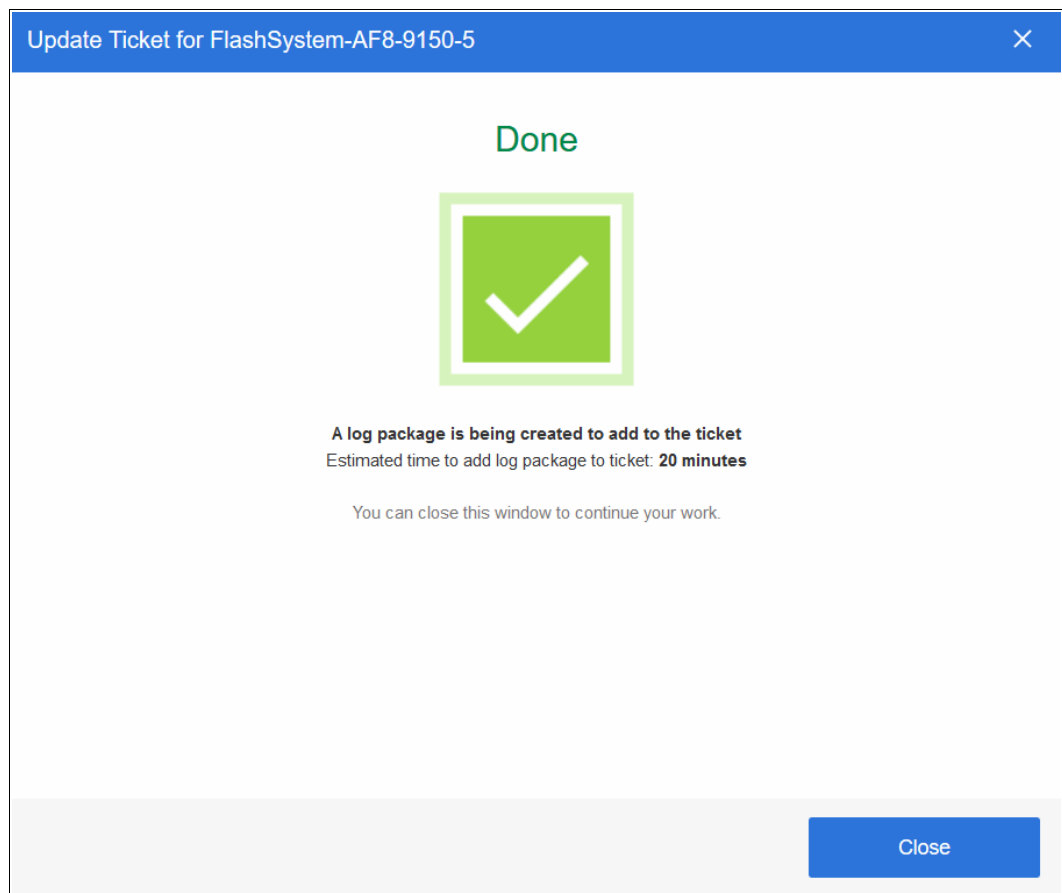


Figure 13-92 Confirming the log upload



Performance data and statistics gathering

This appendix provides a brief overview of the performance analysis capabilities of the IBM Storage System and IBM Spectrum Virtualize V8.4. It also describes a method that you can use to collect and process IBM Spectrum Virtualize performance statistics.

It is beyond the intended scope of this book to provide an in-depth understanding of performance statistics or to explain how to interpret them. For more information about the performance of the IBM Storage Systems, see *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521.

This appendix includes the following topics:

- ▶ “IBM Storage System performance overview” on page 880
- ▶ “Performance monitoring” on page 882

IBM Storage System performance overview

Storage virtualization with IBM Spectrum Virtualize provides many administrative benefits. In addition, it can provide an increase in performance for some workloads. The caching capability of IBM Spectrum Virtualize and its ability to stripe volumes across multiple disk arrays can provide a performance improvement over what can otherwise be achieved when midrange disk subsystems are used.

To ensure that the performance levels of your system are maintained, monitor performance periodically to provide visibility into potential problems that exist or are developing so that they can be addressed in a timely manner.

Performance considerations

When you are designing the IBM Spectrum Virtualize infrastructure or maintaining an existing infrastructure, you must consider many factors in terms of their potential effect on performance. These factors include, but are not limited to, dissimilar workloads that are competing for the same resources, overloaded resources, insufficient available resources, poor performing resources, and similar performance constraints.

Remember the following high-level rules when you are designing your storage area network (SAN) and IBM Spectrum Virtualize layout:

- ▶ Host-to-system inter-switch link (ISL) oversubscription.

This area is the most significant input/output (I/O) load across ISLs. A best practice is to maintain a maximum of 7-to-1 oversubscription. A higher ratio is possible, but it tends to lead to I/O bottlenecks. This best practice also assumes a core-edge design, where the hosts are on the edges and the IBM FlashSystem is the core.

- ▶ Storage-to-system ISL oversubscription.

This area is the second most significant I/O load across ISLs. The maximum oversubscription is 7-to-1. A higher ratio is not supported. Again, this best practice assumes a multiple-switch SAN fabric design.

- ▶ Node-to-node ISL oversubscription.

This area does not apply to IBM FlashSystem clusters composed of a unique control enclosure. This area is the least significant load of the three possible oversubscription bottlenecks. In standard setups, this load can be ignored. Although this area is not entirely negligible, it does not contribute significantly to the ISL load. However, node-to-node ISL oversubscription is mentioned here in relation to the stretched cluster capability.

When the system is running in this manner, the number of ISL links becomes more important. As with the storage-to-system ISL oversubscription, this load also has a maximum of 7-to-1 oversubscription. Exercise caution and careful planning when you determine the number of ISLs to implement. If you need assistance, contact your IBM representative and request technical assistance.

- ▶ ISL trunking or port channeling.

For the best performance and availability, use ISL trunking or port channeling. Independent ISL links can easily become overloaded and turn into performance bottlenecks. Bonded or trunked ISLs automatically share load and provide better redundancy in a failure.

- ▶ Number of paths per host multipath device.

The maximum supported number of paths per multipath device that is visible on the host is eight (With HyperSwap, you may have up to 16 active paths). Although most vendor multipathing software can support more paths, the IBM Storage System expects a maximum of eight paths. In general, you see only an effect on performance from more paths than eight. Although IBM Spectrum Virtualize can work with more than eight paths, that configuration is unsupported.

- ▶ Do not intermix dissimilar array types or sizes.

Although IBM Spectrum Virtualize supports an intermix of different types of storage within storage pools, it is a best practice to always use the same array model, redundant array of independent disks (RAID) mode. RAID size (RAID 5 6+P+S does not mix well with RAID 6 14+2), and drive speeds.

Rules and guidelines are no substitution for monitoring performance. Monitoring performance can provide validation that design expectations are met, and identify opportunities for improvement.

IBM Spectrum Virtualize performance perspectives

IBM Spectrum Virtualize Software was developed by the IBM Research Group. It runs on commodity hardware (mass-produced Intel -based processors (CPUs) with mass-produced expansion cards), and provides distributed cache and a scalable cluster architecture. One of the main goals of its design is to use refreshes in hardware. Currently, the IBM Storage System cluster is scalable up to eight nodes (four control enclosures).

Note: For IBM FlashSystem 5030 and IBM FlashSystem 5100, only four nodes (two control enclosures) are supported.

The performance is near linear when nodes are added into the cluster until performance eventually becomes limited by the attached components. Although virtualization provides significant flexibility in terms of the components that are used, it does not diminish the necessity of designing the system around the components so that it can deliver the level of performance that you want.

The key item for planning is your SAN layout. Switch vendors have slightly different planning requirements, but the goal is that you always want to maximize the bandwidth that is available to the IBM Storage System ports. An IBM FlashSystem system is one of the few devices that can drive ports to their limits on average, so it is imperative that you put significant thought into planning the SAN layout.

Essentially, performance improvements are gained by selecting the most appropriate internal disk drive types, spreading the workload across a greater number of back-end resources when using external storage, and adding more caching. These capabilities are provided by the IBM Storage System cluster. However, the performance of individual resources eventually becomes the limiting factor.

Performance monitoring

This section highlights several performance monitoring techniques.

Collecting performance statistics

IBM Spectrum Virtualize is constantly collecting performance statistics. The default frequency by which files are created is 5-minute intervals. The collection interval can be changed by running the **startstats** command.

The statistics files for volumes, managed disks (MDisks), nodes, and drives are saved at the end of the sampling interval. A maximum of 16 files (each) are stored before they are overlaid in a rotating log fashion. This design provides statistics for the most recent 240-minute period if the default 15-minute sampling interval is used. IBM Spectrum Virtualize supports user-defined sampling intervals of 1 - 60 minutes. IBM Storage Insights requires and recommends interval of 5 minutes.

For each type of object (volumes, MDisks, nodes, and drives), a separate file with statistic data is created at the end of each sampling period and stored in `/dumps/iostats`.

Run the **startstats** command to start the collection of statistics, as shown in Example A-1.

Example: A-1 The startstats command

```
IBM FlashSystem 7200:superuser>startstats -interval 5
```

This command starts statistics collection and gathers data at 5-minute intervals.

To verify the statistics collection interval, display the system properties again, as shown in Example A-2.

Example: A-2 Statistics collection status and frequency

```
IBM FlashSystem 7200:superuser>>lssystem
statistics_status on
statistics_frequency 5
-- The output has been shortened for easier reading. --
```

It is not possible to stop statistics collection with the command **stopstats** starting with Version 8.1.

Collection intervals: Although more frequent collection intervals provide a more detailed view of what happens within IBM Spectrum Virtualize and IBM FlashSystem, they shorten the amount of time that the historical data is available on IBM Spectrum Virtualize. For example, rather than a 240-minute period of data with the default 15-minute interval, if you adjust to 2-minute intervals, you have a 32-minute period instead.

Statistics are collected per node. The sampling of the internal performance counters is coordinated across the cluster so that when a sample is taken, all nodes sample their internal counters concurrently. Collect all files from all nodes for a complete analysis. Tools such as IBM Spectrum Control and IBM Spectrum Insight® Pro perform this intensive data collection for you.

Statistics file naming

The statistics files that are generated are written to the `/dumps/iostats/` directory. The file name has the following formats:

- ▶ `Nm_stats_<node_id>_<date>_<time>` for MDisks statistics
- ▶ `Nv_stats_<node_id>_<date>_<time>` for Volumes statistics
- ▶ `Nn_stats_<node_id>_<date>_<time>` for node statistics
- ▶ `Nd_stats_<node_id>_<date>_<time>` for drives statistics

The `node_id` is the name of the node on which the statistics were collected. The date is in the form `<yymmdd>`, and the time is in the form `<hhmmss>`. The following example shows an MDisk statistics file name:

```
Nm_stats_113986_161019_151832
```

Example A-3 shows typical MDisk, volume, node, and disk drive statistics file names.

Example: A-3 File names of per node statistics

```
IBM FlashSystem 7200:superuser>lsdumps -prefix /dumps/iostats
id  filename
0   Nd_stats_7825WKP-2_201029_171058
1   Nv_stats_7825WKP-2_201029_171058
2   Nn_stats_7825WKP-2_201029_171058
3   Nm_stats_7825WKP-2_201029_171058
4   Nn_stats_7825WKP-1_201029_172558
5   Nd_stats_7825WKP-2_201029_172558
6   Nd_stats_7825WKP-1_201029_172558
...
129 Nn_stats_7825WKP-1_201029_211058
130 Nv_stats_7825WKP-1_201029_211058
131 Nd_stats_7825WKP-1_201029_211058
IBM FlashSystem 7200:superuser>
```

Note: For more information about the statistics files name convention, see [IBM Documentation](#).

Tip: The performance statistics files can be copied from the IBM FlashSystem nodes to a local drive on your workstation by using `pscp.exe` (included with PuTTY) from an MS-DOS command prompt, as shown in this example:

```
C:\Program Files\PuTTY>pscp -unsafe -load IBM FlashSystem 7200
superuser@9.71.42.30:/dumps/iostats/* c:\statsfiles
```

Use the `-load` parameter to specify the session that is defined in PuTTY.

Specify the `-unsafe` parameter when you use wildcards.

You can obtain PuTTY from [Download PuTTY: latest release \(0.74\)](#) [Download PuTTY: latest release \(0.74\)](#).

Real-time performance monitoring

IBM Storage System supports real-time performance monitoring. Real-time performance statistics provide short-term status information for the IBM FlashSystem system. The statistics are shown as graphs in the management GUI, or can be viewed from the command-line interface (CLI). With system-level statistics, you can quickly view the CPU usage and the bandwidth of volumes, interfaces, and MDisks. Each graph displays the current bandwidth in megabytes per second (MBps) or input/output operations per second (IOPS), and a view of bandwidth over time.

Each node collects various performance statistics (mostly at 5-second intervals) and the statistics that are available from the config node in a clustered environment. This information can help you determine the performance effect of a specific node.

As with system statistics, node statistics help you to evaluate whether the node is operating within normal performance metrics.

Real-time performance monitoring gathers the following system-level performance statistics:

- ▶ CPU utilization
- ▶ Port utilization and I/O rates
- ▶ Volume and MDisk I/O rates
- ▶ Bandwidth
- ▶ Latency

Real-time statistics are not a configurable option and cannot be disabled.

Real-time performance monitoring with the CLI

The `lsnodecanisterstats` and `lssystemstats` commands are available for monitoring the statistics by using the CLI.

The `lsnodecanisterstats` command provides performance statistics for the nodes that are part of a clustered system, as shown in Example A-4. The output is truncated and shows only part of the available statistics. You can also specify a node name in the command to limit the output for a specific node.

Example: A-4 The `lsnodecanisterstats` command output

```
IBM FlashSystem 7200:superuser>lsnodecanisterstats
node_id node_name stat_name          stat_current stat_peak stat_peak_time
1       node1     compression_cpu_pc 0             0         201029211843
1       node1     cpu_pc             2             2         201029211843
1       node1     fc_mb              0             14        201029211603
1       node1     fc_io              10            116       201029211603
1       node1     sas_mb             11            124       201029211803
1       node1     sas_io             45            1214      201029211658
1       node1     iscsi_mb           0             0         201029211843
1       node1     iscsi_io           0             0         201029211843
1       node1     write_cache_pc    10            10        201029211843
1       node1     total_cache_pc    79            79        201029211843
1       node1     vdisk_mb           0             14        201029211603
1       node1     vdisk_io           0             105       201029211603
1       node1     vdisk_ms           0             0         201029211843
1       node1     mdisk_mb           0             16        201029211603
1       node1     mdisk_io           0             84        201029211603
1       node1     mdisk_ms           0             0         201029211843
1       node1     drive_mb           11            124       201029211803
```

```

1      node1      drive_io      45          492         201029211803
1      node1      drive_ms      13          31          201029211643
1      node1      vdisk_r_mb    0           14          201029211603
1      node1      vdisk_r_io    0           105         201029211603
...
3      node2      drive_w_ms    6           10          201029211713
3      node2      iplink_mb     0           0           201029211843
3      node2      iplink_io     0           0           201029211843
3      node2      iplink_comp_mb 0           0           201029211843
3      node2      cloud_up_mb   0           0           201029211843
3      node2      cloud_up_ms   0           0           201029211843
3      node2      cloud_down_mb 0           0           201029211843
3      node2      cloud_down_ms 0           0           201029211843
3      node2      iser_mb      0           0           201029211843
3      node2      iser_io      0           0           201029211843
IBM FlashSystem 7200:superuser>

```

Example A-4 on page 884 shows statistics for the two node members of system ITS0. For each node, the following columns are displayed:

- ▶ `stat_name`: The name of the statistic field
- ▶ `stat_current`: The current value of the statistic field
- ▶ `stat_peak`: The peak value of the statistic field in the last 5 minutes
- ▶ `stat_peak_time`: The time that the peak occurred

The `l1nodecanisterstats` command can also be used with a node canister name or ID as an argument. For example, you can enter the command `l1nodecanisterstats node1` to display the statistics of node name node1 only.

The `lssystemstats` command lists the same set of statistics that is listed with the `l1nodecanisterstats` command, but represents all nodes in the cluster. The values for these statistics are calculated from the node statistics values in the following way:

- ▶ **Bandwidth**: Sum of bandwidth of all nodes
- ▶ **Latency**: Average latency for the cluster, which is calculated by using data from the whole cluster, not an average of the single node values
- ▶ **IOPS**: Total IOPS of all nodes
- ▶ **CPU percentage**: Average CPU percentage of all nodes

Example A-5 shows the resulting output of the `lssystemstats` command.

Example: A-5 The lssystemstats command output

```

IBM FlashSystem 7200:superuser>lssystemstats
stat_name      stat_current  stat_peak  stat_peak_time
compression_cpu_pc 0           0          201029212153
cpu_pc         2           2          201029212153
fc_mb         0           0          201029212153
fc_io         20          26         201029212123
sas_mb        38          145        201029212028
sas_io        197         1226       201029211658
iscsi_mb      0           0          201029212153
iscsi_io      0           1          201029212008
write_cache_pc 10          10         201029212153
total_cache_pc 79          80         201029211743
vdisk_mb      0           0          201029212153

```

```

vdisk_io          0          5          201029212123
vdisk_ms          0          0          201029212153
mdisk_mb          2          2          201029212153
mdisk_io          2          2          201029212153
mdisk_ms          9          13         201029212018
drive_mb          38         145        201029212028
drive_io          152         585        201029212028
drive_ms          10          20         201029211813
vdisk_r_mb        0          0          201029212153
vdisk_r_io        0          0          201029212153
vdisk_r_ms        0          0          201029212153
vdisk_w_mb        0          0          201029212153
vdisk_w_io        0          5          201029212123
vdisk_w_ms        0          0          201029212153
mdisk_r_mb        0          0          201029212153
mdisk_r_io        0          0          201029212153
mdisk_r_ms        0          0          201029212153
mdisk_w_mb        2          2          201029212153
mdisk_w_io        2          2          201029212153
mdisk_w_ms        9          13         201029212018
drive_r_mb        35         143        201029212028
drive_r_io        142         574        201029212028
drive_r_ms        11          22         201029211813
drive_w_mb        2          3          201029212023
drive_w_io        10          15         201029212113
drive_w_ms        6          10         201029211913
power_w           529         541        201029211733
temp_c            22          22         201029212153
temp_f            71          71         201029212153
iplink_mb         0          0          201029212153
iplink_io         0          0          201029212153
iplink_comp_mb    0          0          201029212153
cloud_up_mb       0          0          201029212153
cloud_up_ms       0          0          201029212153
cloud_down_mb     0          0          201029212153
cloud_down_ms     0          0          201029212153
iser_mb           0          0          201029212153
iser_io           0          0          201029212153
IBM FlashSystem 7200:superuser>

```

Table A-1 gives the descriptions of the different counters that are presented by the **Issystemstats** and **Isnodecanisterstats** commands.

Table A-1 List of counters for the Issystemstats and Isnodecanisterstats commands

Value	Description
compression_cpu_pc	Displays the percentage of allocated CPU capacity that is used for compression.
cpu_pc	Displays the percentage of allocated CPU capacity that is used for the system.
fc_mb	Displays the total number of megabytes transferred per second for Fibre Channel (FC) traffic on the system. This value includes host I/O and any bandwidth that is used for communication within the system.

Value	Description
fc_io	Displays the total I/O operations that are transferred per second for FC traffic on the system. This value includes host I/O and any bandwidth that is used for communication within the system.
sas_mb	Displays the total number of megabytes transferred per second for serial-attached Small Computer System Interface (SCSI) (SAS) traffic on the system. This value includes host I/O and bandwidth that is used for background RAID activity.
sas_io	Displays the total I/O operations that are transferred per second for SAS traffic on the system. This value includes host I/O and bandwidth that is used for background RAID activity.
iscsi_mb	Displays the total number of megabytes transferred per second for internet Small Computer Systems Interface (iSCSI) traffic on the system.
iscsi_io	Displays the total I/O operations that are transferred per second for iSCSI traffic on the system.
write_cache_pc	Displays the percentage of the write cache usage for the node.
total_cache_pc	Displays the total percentage for both the write and read cache usage for the node.
vdisk_mb	Displays the average number of megabytes transferred per second for read and write operations to volumes during the sample period.
vdisk_io	Displays the average number of I/O operations that are transferred per second for read and write operations to volumes during the sample period.
vdisk_ms	Displays the average amount of time in milliseconds (ms) that the system takes to respond to read and write requests to volumes over the sample period.
mdisk_mb	Displays the average number of megabytes transferred per second for read and write operations to MDisks during the sample period.
mdisk_io	Displays the average number of I/O operations that are transferred per second for read and write operations to MDisks during the sample period.
mdisk_ms	Displays the average amount of time in milliseconds that the system takes to respond to read and write requests to MDisks over the sample period.
drive_mb	Displays the average number of megabytes transferred per second for read and write operations to drives during the sample period.
drive_io	Displays the average number of I/O operations that are transferred per second for read and write operations to drives during the sample period.
drive_ms	Displays the average amount of time in milliseconds that the system takes to respond to read and write requests to drives over the sample period.
vdisk_w_mb	Displays the average number of megabytes transferred per second for read and write operations to volumes during the sample period.
vdisk_w_io	Displays the average number of I/O operations that are transferred per second for write operations to volumes during the sample period.
vdisk_w_ms	Displays the average amount of time in milliseconds that the system takes to respond to write requests to volumes over the sample period.
mdisk_w_mb	Displays the average number of megabytes transferred per second for write operations to MDisks during the sample period.

Value	Description
mdisk_w_io	Displays the average number of I/O operations that are transferred per second for write operations to MDisks during the sample period.
mdisk_w_ms	Displays the average amount of time in milliseconds that the system takes to respond to write requests to MDisks over the sample period.
drive_w_mb	Displays the average number of megabytes transferred per second for write operations to drives during the sample period.
drive_w_io	Displays the average number of I/O operations that are transferred per second for write operations to drives during the sample period.
drive_w_ms	Displays the average amount of time in milliseconds that the system takes to respond write requests to drives over the sample period.
vdisk_r_mb	Displays the average number of megabytes transferred per second for read operations to volumes during the sample period.
vdisk_r_io	Displays the average number of I/O operations that are transferred per second for read operations to volumes during the sample period.
vdisk_r_ms	Displays the average amount of time in milliseconds that the system takes to respond to read requests to volumes over the sample period.
mdisk_r_mb	Displays the average number of megabytes transferred per second for read operations to MDisks during the sample period.
mdisk_r_io	Displays the average number of I/O operations that are transferred per second for read operations to MDisks during the sample period.
mdisk_r_ms	Displays the average amount of time in milliseconds that the system takes to respond to read requests to MDisks over the sample period.
drive_r_mb	Displays the average number of megabytes transferred per second for read operations to drives during the sample period.
drive_r_io	Displays the average number of I/O operations that are transferred per second for read operations to drives during the sample period.
drive_r_ms	Displays the average amount of time in milliseconds that the system takes to respond to read requests to drives over the sample period.
iplink_mb	The total number of megabytes transferred per second for IP replication traffic on the system. This value does not include iSCSI host I/O operations.
iplink_comp_mb	Displays the average number of compressed MBps over the IP replication link during the sample period.
iplink_io	The total I/O operations that are transferred per second for IP partnership traffic on the system. This value does not include iSCSI host I/O operations.
cloud_up_mb	Displays the average number of megabits per second (Mbps) for upload operations to a cloud account during the sample period.
cloud_up_ms	Displays the average amount of time (in milliseconds) it takes for the system to respond to upload requests to a cloud account during the sample period.
cloud_down_mb	Displays the average number of Mbps for download operations to a cloud account during the sample period.

Value	Description
cloud_down_ms	Displays the average amount of time (in milliseconds) that it takes for the system to respond to download requests to a cloud account during the sample period.
iser_mb	Displays the total number of megabytes transferred per second for iSCSI Extensions for Remote Direct Memory Access (RDMA) (iSER) traffic on the system.
iser_io	Displays the total I/O operations that are transferred per second for iSER traffic on the system.

Real-time performance statistics monitoring with the GUI

The IBM Spectrum Virtualize Dashboard provides performance at a glance by displaying important information about the system. You can see the entire cluster (the system) performance by selecting information such as bandwidth, response time, IOPS, or CPU utilization. You can also display a Node Comparison by selecting the same information as for the cluster and then switching the button, as shown in Figure A-1.

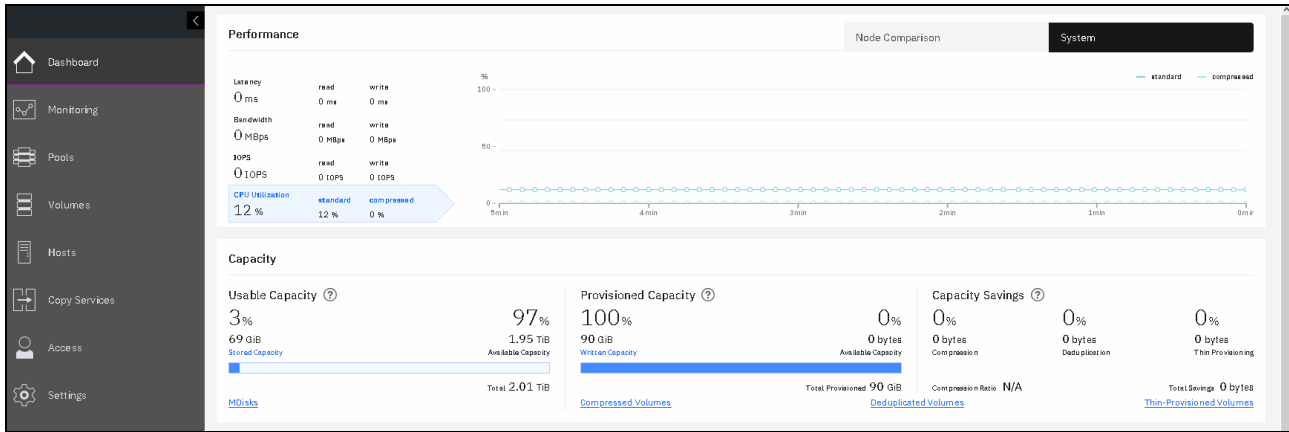


Figure A-1 IBM Spectrum Virtualize Dashboard displaying the System Performance overview

Figure A-2 shows the display after switching the button.

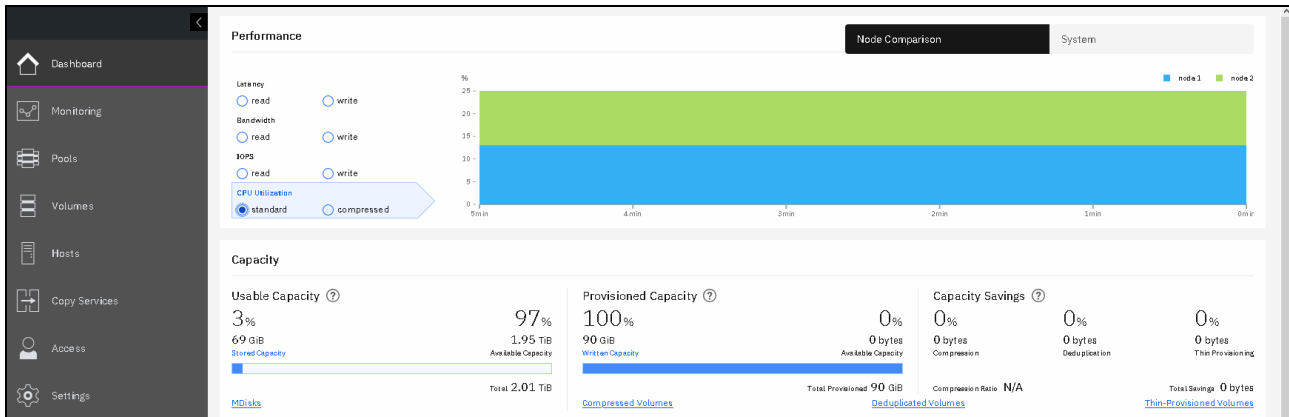


Figure A-2 IBM Spectrum Virtualize Dashboard displaying the Nodes Performance overview

You can also use real-time statistics to monitor CPU utilization, volume, interface, and the MDisk bandwidth of your system and nodes. Each graph represents 5 minutes of collected statistics and provides a means of assessing the overall performance of your system.

The real-time statistics are available from the IBM Spectrum Virtualize GUI. To open the Performance Monitoring window, select **Monitoring** → **Performance** (as shown in Figure A-3).

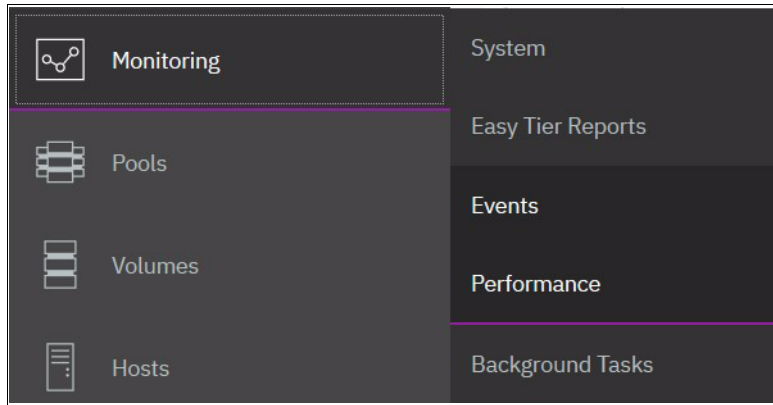


Figure A-3 Selecting the Performance menu in the Monitoring menu

As shown in Figure A-4, the Performance monitoring window is divided into the following sections that provide utilization views for the following resources.



Figure A-4 IBM Spectrum Virtualize Performance window

- ▶ **CPU Utilization:** The CPU Utilization graph shows the current percentage of CPU utilization and peaks in utilization. It can also display compression CPU utilization for systems with compressed volumes.
- ▶ **Volumes:** Shows four metrics about the overall volume utilization graphics:
 - Read
 - Write
 - Read latency
 - Write latency
- ▶ **Interfaces:** The Interfaces graph displays data points for FC, iSCSI, SAS, and IP Remote Copy (RC) interfaces. You can use this information to help determine connectivity issues that might affect performance:
 - FC
 - iSCSI

- SAS
- IP Remote Copy
- ▶ MDisks: Also shows four metrics on the overall MDisks graphics:
 - Read
 - Write
 - Read latency
 - Write latency

You can use these metrics to help determine the overall performance health of the volumes and MDisks on your system. Consistent unexpected results can indicate errors in configuration, system faults, or connectivity issues.

The system's performance is always visible at the bottom of the IBM Spectrum Virtualize window.

Note: The indicated values in the graphics are averaged on a 5-second-based sample.

You can also select to view performance statistics for each of the available nodes of the system, as shown in Figure A-5.

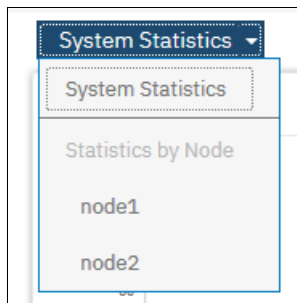


Figure A-5 Viewing statistics per node or for the entire system

You can also change the metric between MBps or IOPS, as shown in Figure A-6.

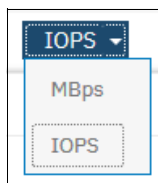


Figure A-6 Viewing performance metrics by MBps or IOPS

On any of these views, you can select any point by using your cursor to see the exact value and when it occurred. When you place your cursor over the timeline, it becomes a dotted line with the various values gathered, as shown in Figure A-7.

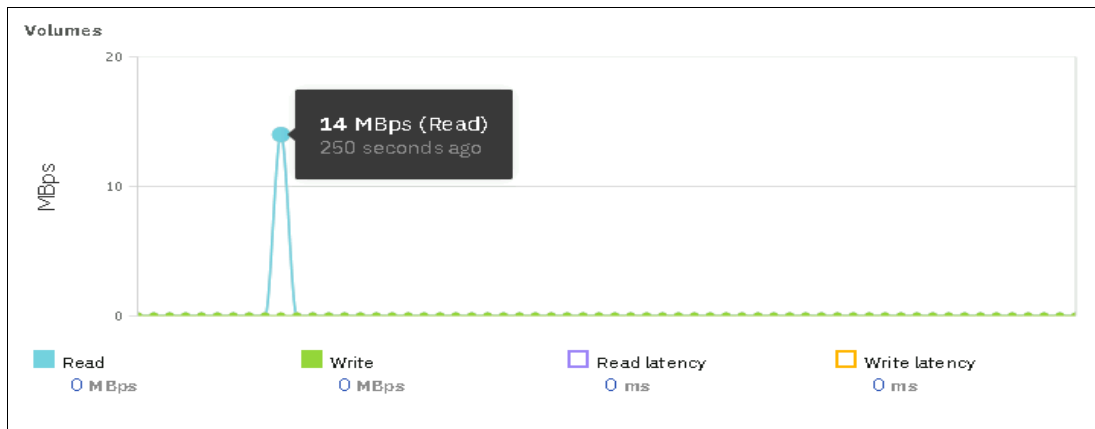


Figure A-7 Viewing performance with details

For each of the resources, various metrics are available, and you can select which ones to be displayed. For example, as shown in Figure A-8, from the four available metrics for the MDisks view (Read, Write, Read latency, and Write latency), only Read and Write IOPS are selected.

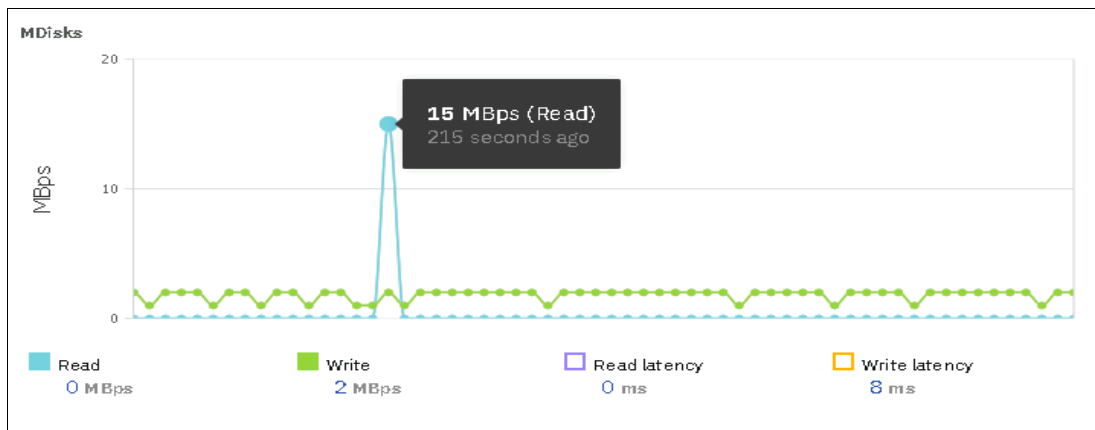


Figure A-8 Displaying performance counters

Performance data collection and IBM Spectrum Control

Although you can obtain performance statistics in standard .xml files, the use of .xml files is a less practical and more complicated method to analyze the IBM Spectrum Virtualize performance statistics. IBM Spectrum Control is the supported IBM tool to collect and analyze IBM Storage Systems performance statistics.

IBM Spectrum Control is installed separately on a dedicated system, and is not part of the IBM Spectrum Virtualize bundle.

For more information about using IBM Spectrum Control to monitor your storage subsystem, see [Harness the full power of your IT infrastructure](#).

As an alternative to IBM Spectrum Control, a cloud-based tool that is called IBM Storage Insights is available that provides a single dashboard that gives you a clear view of all your IBM block storage by showing performance and capacity information. You do not have to install this tool in your environment because it is a cloud-based solution. Only an agent is required to collect data of the storage devices.

For more information about IBM Storage Insights, see [IBM Storage Insights](#).



B

Terminology

This appendix summarizes the IBM Spectrum Virtualize and IBM Storage terms that are commonly used in this book.

For more information about the complete set of terms that relate to IBM FlashSystem systems, see [IBM Documentation](#).

Commonly encountered terms

This book uses the common IBM Spectrum Virtualize and IBM FlashSystem terminology that is listed in this section.

Access mode

One of the modes in which a logical unit (LU) in a disk controller system can operate. The three access modes are image mode, managed space mode, and unconfigured mode. See also “Image mode” on page 908, “Managed mode” on page 911, and “Unconfigured mode” on page 921.

Activation key

See “License key” on page 910

Allocatable extent limit

A maximum total capacity for the system. The allocatable extent limit is calculated from pool extent sizes.

Array

An ordered collection, or group, of physical devices (disk drive modules) that are used to define logical volumes or devices. An array is a group of drives that is designated to be managed with a redundant array of independent disks (RAID).

Asymmetric virtualization

Asymmetric virtualization is a virtualization technique in which the virtualization engine is outside the data path and performs a metadata-style service. The metadata server contains all the mapping and locking tables, and the storage devices contain only data. See also “Symmetric virtualization” on page 920.

Asynchronous replication

Asynchronous replication is a type of replication in which control is given back to the application as soon as the write operation is made to the source volume. Later, the write operation is made to the target volume. See also “Synchronous replication” on page 920.

Audit Log

An unalterable record of all commands or user interactions that are issued to the system.

Automatic data placement mode

Automatic data placement mode is an Easy Tier operating mode in which the host activity on all the volume extents in a pool are “measured,” a migration plan is created, and then automatic extent migration is performed.

Auxiliary volume

The auxiliary volume that contains a mirror of the data on the master volume. See also “Master volume” on page 911, and “Relationship” on page 917.

Available (usable) capacity

See “Capacity” on page 897.

Back end

See “Front end and back end” on page 906.

Caching input/output group

The caching input/output (I/O group) is the I/O group in the system that performs the cache function for a volume.

Call Home

Call Home is a communication link that is established between a product and a service provider. The product can use this link to call IBM or another service provider when the product requires service. With access to the machine, service personnel can perform service tasks, such as viewing error and problem logs or initiating trace and dump retrievals.

Canister

A canister is a single processing unit within a storage system.

Capacity

IBM applies the following definitions to capacity:

- ▶ Available capacity

The amount of usable capacity that is not yet used in a system, pool, array, or managed disk (MDisk).

- ▶ Data reduction

A set of techniques that can be used to reduce the amount of usable capacity that is required to store data. Examples of data reduction include data deduplication and compression.

- ▶ Data reduction savings

The total amount of usable capacity that is saved in a system, pool, or volume through the application of an algorithm, such as compression or deduplication on the written data. This saved capacity is the difference between the written capacity and the used capacity.

- ▶ Effective capacity

The amount of provisioned capacity that can be created in a system or pool without running out of usable capacity given the current data reduction savings being achieved. This capacity equals the usable capacity that is divided by the data reduction savings percentage.

- ▶ Overhead capacity

An amount of usable capacity that is occupied by metadata in a system or pool and other data that is used for system operations.

- ▶ Overprovisioned ratio

The ratio of provisioned capacity to usable capacity in the pool or system.

- ▶ Overprovisioning

The result of creating more provisioned capacity in a storage system or pool than there is usable capacity. Overprovisioning occurs when thin provisioning or data reduction techniques ensure that the used capacity of the provisioned volumes is less than their provisioned capacity.

- ▶ Physical Capacity

Physical capacity indicates the total capacity in all storage on the system. Physical capacity includes all the storage the system can virtualize and assign to pools.

- ▶ **Provisioned capacity**
Total capacity of all volumes and Volume copies in a pool or system.
- ▶ **Provisioning limit - maximum provisioned capacity - overprovisioning limit**
In some storage systems, restrictions in the storage hardware or configured by the user define the limit of the maximum provisioned capacity in a pool or system.
- ▶ **Raw capacity**
The reported capacity of the drives in the system before formatting or RAID is applied.
- ▶ **Standard provisioning**
The ability to completely use a volume's capacity for that specific volume.
- ▶ **Standard provisioned volume**
A volume that uses all the storage at creation.
- ▶ **Thin-provisioning savings**
The total amount of usable capacity that is saved in a pool, system, or volume by using usable capacity when needed as a result of write operations. The capacity that is saved is the difference between the provisioned capacity minus the written capacity.
- ▶ **Total capacity savings**
The total amount of usable capacity that is saved in a pool, system, or volume through thin-provisioning and data reduction techniques. The capacity that is saved is the difference between the used usable capacity and the provisioned capacity.
- ▶ **Usable capacity**
The amount of capacity that is provided for storing data on a system, pool, array, or MDisk after formatting and RAID techniques are applied. Usable capacity is the total of used and available capacity. For example, 50 TiB used, 50 TiB available is a usable capacity of 100 TiB.
- ▶ **Used capacity**
The amount of usable capacity that is taken up by data or capacity in a system, pool, array, or MDisk after data reduction techniques are applied.
- ▶ **Written capacity**
The amount of usable capacity that might be used to store written data in a pool or system before data reduction is applied.
- ▶ **Written capacity limit**
The largest amount of capacity that can be written to a drive, array, or MDisk. The limit can be reached even when usable capacity is still available.

Capacity licensing

Capacity licensing is a licensing model that licenses features with a price-per-terabyte model. Licensed features are IBM FlashCopy, Metro Mirror (MM), Global Mirror (GM), and virtualization. See also “FlashCopy” on page 905, “Metro Mirror” on page 912, and “Virtualized storage” on page 921.

Capacity recycling

Capacity recycling means the amount of provisioned capacity that can be recovered without causing stress or performance degradation. This capacity identifies the amount of resources that can be reclaimed and provisioned to other objects in an environment.

Certificate

A digital document that binds a public key to the identity of the certificate owner, which enables the certificate owner to be authenticated. A certificate is issued by a certificate authority (CA) and is digitally signed by that authority.

Chain

A set of enclosures that is attached to provide redundant access to the drives inside the enclosures. Each control enclosure can have one or more chains.

Challenge Handshake Authentication Protocol

Challenge Handshake Authentication Protocol (CHAP) is an authentication protocol that protects against eavesdropping by encrypting the username and password.

Change volume

A volume that is used in GM that holds earlier consistent revisions of data when changes are made.

Channel extender

A channel extender is a device that is used for long-distance communication that connects other storage area network (SAN) fabric components. Generally, channel extenders can involve protocol conversion to asynchronous transfer mode (ATM), IP, or another long-distance communication protocol.

Child pool

Administrators can use child pools to control capacity allocation for volumes that are used for specific purposes. Rather than being created directly from MDisks, child pools are created from existing capacity that is allocated to a parent pool. As with parent pools, volumes can be created that specifically use the capacity that is allocated to the child pool. Child pools are similar to parent pools with similar properties. Child pools can be used for volume copy operation. See also “Parent pool” on page 913.

Clone

A copy of a volume on a server at a particular point in time (PiT). The contents of the copy can be customized while the contents of the original volume are preserved.

Cloud account

An agreement with a cloud service provider (CSP) to use storage or other services at that service provider. Access to the cloud account is granted by presenting valid credentials.

Cloud container

A cloud container is a virtual object that includes all of the elements, components, or data that is common to a specific application or data.

Cloud service provider

A CSP is the company or organization that provides off- and on-premises cloud services, such as storage, server, and network. IBM Spectrum Virtualize includes built-in software capabilities to interact with CSPs such as IBM Cloud, Amazon S3, and deployments of OpenStack Swift.

Cloud tenant

A cloud tenant is a group or an instance that provides common access with the specific privileges to an object, software, or data source.

Clustered system

A clustered system, which was known as a cluster, is a group of up to eight IBM Storage Systems canisters (two in each system) that presents a single configuration, management, and service interface to the user.

Cold extent

A cold extent is an extent of a volume that does not get any performance benefit if it is moved from a hard disk drive (HDD) to a flash drive. A cold extent also refers to an extent that must be migrated onto an HDD if it is on a flash drive.

Compression

Compression is a function that removes repetitive characters, spaces, strings of characters, or binary data from the data that is being processed and replaces characters with control characters. Compression reduces the amount of storage space that is required for data.

Compression accelerator

A compression accelerator is hardware onto which the work of compression is offloaded from the microprocessor.

Configuration node

While the cluster is operational, a single node in the cluster is appointed to provide configuration and service functions over the network interface. This node is termed the configuration node. This configuration node manages the data that describes the clustered-system configuration and provides a focal point for configuration commands. If the configuration node fails, another node in the cluster transparently assumes that role.

Consistency group

A consistency group is a group of copy relationships between virtual volumes or data sets that are maintained with the same time reference so that all copies are consistent in time. A consistency group can be managed as a single entity.

Container

A container is a software object that holds or organizes other software objects or entities.

Contingency capacity

For thin-provisioned volumes that are configured to automatically expand, the contingency capacity is the unused real capacity that is maintained. For thin-provisioned volumes that are not configured to automatically expand, it is the difference between the used capacity and the new real capacity.

Copied state

Copied is a FlashCopy state that indicates that a copy was triggered after the copy relationship was created. The Copied state indicates that the copy process is complete, and the target disk has no further dependency on the source disk. The time of the last trigger event is normally displayed with this status.

Counterpart SAN

A counterpart SAN is the non-redundant portion of a redundant SAN. A counterpart SAN provides all of the connectivity of the redundant SAN, but without the 100% redundancy. IBM Storage canisters are typically connected to a “redundant SAN” that is made up of two counterpart SANs. A counterpart SAN is often called a SAN fabric.

Cross-volume consistency

A consistency group property that ensures consistency between volumes when an application issues dependent write operations that span multiple volumes.

Customer-replaceable unit

An assembly or part that can be replaced in its entirety by a user when any one of its components fails.

Data consistency

Data consistency is a characteristic of the data at the target site where the dependent write order is maintained to ensure the recoverability of applications.

Data deduplication

Data deduplication is a method of reducing storage needs by eliminating redundant data. Only one instance of the data is retained on storage media. Other instances of the same data are replaced with a pointer to the retained instance.

Data encryption key

The data encryption key is used to encrypt data. It is created automatically when an encrypted object, such as an array, a pool, or a child pool, is created. It is stored in secure memory and it cannot be viewed or changed. The data encryption key is encrypted by using the master access key.

Data migration

Data migration is the movement of data from one physical location to another physical location without the disruption of application I/O operations.

Data reduction

Data reduction is a set of techniques that can be used to reduce the amount of physical storage that is required to store data. An example of data reduction includes data deduplication and compression. See also “Data Reduction Pool” and “Capacity” on page 897.

Data reduction savings

The total amount of usable capacity that is saved in a system, pool, or volume through the application of an algorithm such as compression or deduplication on the written data. This saved capacity is the difference between the written capacity and the used capacity. See also “Data reduction”

Data Reduction Pool

Data Reduction Pools (DRPs) are specific types of pools where more control over volumes capacity is given to specific hosts (for example VMware vStorage application programming interfaces (APIs) for Array Integration (VAI), vSphere APIs for Storage Awareness (VASA), and Microsoft Offloaded Data Transfer (ODX)). These hosts can return unused space for reuse. With standard pools, the system is not aware of any unused space on host-allocated volumes. See also “Data reduction”.

Data reduction savings

See “Capacity” on page 897.

Deduplication

See “Data deduplication” on page 901.

Dependent write operation

A write operation that must be applied in the correct order to maintain cross-volume consistency.

Directed maintenance procedure

The fix procedures, which are also known as directed maintenance procedures (DMPs), ensure that you fix any outstanding errors in the error log. To fix errors, from the Monitoring window, click **Events**. The Next Recommended Action is displayed at the top of the Events window. Select **Run This Fix Procedure** and follow the instructions.

Discovery

The automatic detection of a network topology change, for example, new and deleted nodes or links.

Disk tier

MDisks (logical unit numbers (LUNs)) that are presented to the IBM Storage cluster likely have different performance attributes because of the type of disk or RAID array on which they are installed. The MDisks can be on 15,000 RPM Fibre Channel (FC) or serial-attached Small Computer System Interface (SCSI) (SAS) disk, nearline (NL) SAS, or Serial Advanced Technology Attachment (SATA), or even flash drives. Therefore, a storage tier attribute is assigned to each MDisk, and the default is `generic_hdd`.

Distributed redundant array of independent disks

An alternative RAID scheme where the number of drives that are used to store the array can be greater than the equivalent, typical RAID scheme. The same data stripes are distributed across a greater number of drives, which increases the opportunity for parallel I/O and improves the overall performance of array. See also “Rebuild area” on page 916.

Domain name server

A server program that supplies name-to-address conversion by mapping domain names to IP addresses.

Domain Name System

The distributed database system that maps domain names to IP addresses.

Drive technology

A category of a drive that pertains to the method and reliability of the data storage techniques being used on the drive. Possible values include enterprise (ENT) drive, NL drive, or solid-state drive (SSD).

Dual Inline Memory Module

A Dual Inline Memory Module (DIMM) is a small circuit board with memory-integrated circuits containing signal and power pins on both sides of the board.

Here are some terms that are associated with DIMMs:

- ▶ **Channel:** The memory modules are installed into matching banks, which are usually color-coded on the system board. These separate channels enable the memory controller to access each memory module. For the Intel Cascade Lake architecture, there are six DIMM Memory channels per CPU, and each memory channel has two DIMMs. The memory bandwidth is tied to each of these channels, and the speed of access for the memory controller is shared across the pair of DIMMs in that channel.
- ▶ **Slot:** Generally, the physical slot that a DIMM can fit into, but in this context, a slot is DIMM0 or DIMM1, which refers to the first or second slot within a channel on the system board. There are two slots per memory channel on the IBM SAN Volume Controller (SVC) SV2 hardware. On the system board, DIMM0 is the blue slot and DIMM1 is the black slot within each channel.
- ▶ **Rank:** A single-rank DIMM has one set of memory chips that is accessed while writing to or reading from the memory. A dual-rank DIMM is like having two single-rank DIMMs on the same module, with only one rank accessible at a time. A quad-rank DIMM is, effectively, two dual-rank DIMMs on the same module. The 32G DIMMS are dual rank.

Easy Tier

Easy Tier is a volume performance function within the IBM Storage family that provides automatic data placement of a volume's extents in a multitiered storage pool. The pool normally contains a mix of flash drives and HDDs. Easy Tier measures host I/O activity on the volume's extents and migrates hot extents onto the flash drives to ensure the maximum performance.

Effective capacity

See "Capacity" on page 897.

Encryption key

The encryption key, also known as master access key, is created and stored on USB flash drives or on a key server when encryption is enabled. The master access key is used to decrypt the data encryption key.

Encryption key manager / server

An internal or external system that receives and then serves existing encryption keys or certificates to a storage system.

Encryption recovery key

An encryption key that enables a method to recover from an encryption deadlock situation where the normal encryption key servers are not available.

Encryption of data-at-rest

Encryption of data-at-rest is the inactive encryption data that is stored physically on the storage system.

Evaluation mode

Evaluation mode is an Easy Tier operating mode in which the host activity on all the volume extents in a pool are "measured" only. No automatic extent migration is performed.

Event (error)

An event is an occurrence of significance to a task or system. Events can include the completion or failure of an operation, user action, or a change in the state of a process.

Event code

An event code is a value that is used to identify an event condition to a user. This value might map to one or more event IDs or to values that are presented on the service window. This value is used to report error conditions to IBM and to provide an entry point into the service guide.

Event ID

An event ID is a value that is used to identify a unique error condition that was detected by the IBM Storage System. An event ID is used internally in the cluster to identify the error.

Excluded condition

The excluded condition is a status condition. It describes an MDisk that the IBM Storage System decided is no longer sufficiently reliable to be managed by the cluster. The user must issue a command to include the MDisk in the cluster-managed storage.

Extent

An extent is a fixed-size unit of data that is used to manage the mapping of data between MDisks and volumes. The size of the extent can range 16 MB - 8 GB.

External storage

External storage refers to MDisks that are SCSI LUs that are presented by storage systems that are attached to and managed by the clustered system.

Failback

Failback is the restoration of an appliance to its initial configuration after the detection and repair of a failed network or component.

Failover

Failover is an automatic operation that switches to a redundant or standby system or node in a software, hardware, or network interruption. See also “Failback”.

Feature activation code

An alphanumeric code that activates a licensed function on a product. See also “License key” on page 910.

Fibre Channel

FC is a technology for transmitting data between computer devices. It is especially suited for attaching computer servers to shared storage devices and for interconnecting storage controllers and drives. See also “Zoning” on page 923.

Fibre Channel Arbitrated Loop

Fibre Channel Arbitrated Loop (FC-AL) is an implementation of the FC standards that uses a ring topology for the communication fabric, as described in American National Standards Institute (ANSI) INCITS 272-1996 (R2001). In this topology, two or more FC end points are interconnected through a looped interface.

Fibre Channel connection

A Fibre Channel connection (IBM FICON) is an FC communication protocol for IBM mainframe computers and peripheral devices.

Fibre Channel over IP

Fibre Channel over IP (FCIP) is network storage technology that combines the features of the Fibre Channel Protocol (FCP) and the IP to connect distributed SANs over large distances.

Fibre Channel port logins

FC port logins refer to the number of hosts that can see any one V7000 port. The IBM Storage System has a maximum limit per node port (N_Port) of FC logins that are allowed.

Fibre Channel Protocol

FCP is the serial SCSI command protocol that is used on FC networks.

Field-replaceable unit

Field-replaceable units (FRUs) are individual parts that are replaced entirely when any one of the unit's components fails. They are held as spares by the IBM service organization.

File Transfer Protocol

In TCP/IP, FTP is an application layer protocol that uses TCP and telnet services to transfer bulk-data files between machines or hosts.

Fix procedure

A maintenance procedure that runs within the product application and provides step-by-step guidance to resolve an error condition.

FlashCopy

FlashCopy refers to a point-in-time (PiT) copy where a virtual copy of a volume is created. The target volume maintains the contents of the volume at the PiT when the copy was established. Any subsequent write operations to the source volume are not reflected on the target volume.

FlashCopy mapping

A FlashCopy mapping is a continuous space on a direct-access storage volume that is occupied by or reserved for a particular data set, data space, or file.

FlashCopy relationship

See "FlashCopy mapping" on page 905.

FlashCopy service

FlashCopy service is a copy service that duplicates the contents of a source volume on a target volume. In the process, the original contents of the target volume are lost. See also "Point-in-time copy" on page 914.

Flash drive

A data storage device, which is typically removable and rewriteable, that uses solid-state memory to store persistent data. See also "Flash module".

Flash module

A modular hardware unit containing flash memory, one or more flash controllers, and associated electronics. See also "Flash drive".

Front end and back end

The IBM Storage System takes MDisks to create pools of capacity from which volumes are created and presented to application servers (hosts). The volumes that are presented to the hosts are in the front end of an IBM Storage System.

Full restore operation

A copy operation where a local volume is created by reading an entire a volume snapshot from cloud storage.

Full snapshot

A type of volume snapshot that contains all the volume data. When a full snapshot is created, an entire copy of the volume data is transmitted to the cloud.

General Parallel File System

General Parallel File System (GPFS) is a high-performance shared-disk file system that can provide data access from nodes in a clustered system environment.

Gigabyte

A gigabyte (GB) is, for processor storage, real and virtual storage, and channel volume, two to the power of 30 or 1,073,741,824 bytes. For disk storage capacity and communications volume, it is 1,000,000,000 bytes.

Global Mirror

GM is a method of asynchronous replication that maintains data consistency across multiple volumes within or across multiple systems. GM is used where distances between the source site and target site cause increased latency beyond what the application can accept.

Global Mirror with Change Volumes

Change volumes are used to record changes to the primary and secondary volumes of a Remote Copy (RC) relationship. A FlashCopy mapping exists between a primary and its change volume, and a secondary and its change volume.

GPFS cluster

A system of nodes that are defined as being available for use by GPFS file systems.

GPFS snapshot

A PiT copy of a file system or file set.

Grain

A grain is the unit of data that is represented by a single bit in a FlashCopy bitmap (64 kibibytes (KiB) or 256 KiB) in the IBM Storage System. A grain is also the unit to extend the real size of a thin-provisioned volume (32 KiB, 64 KiB, 128 KiB, or 256 KiB).

Graphical user interface

A graphical user interface (GUI) is a computer interface that presents a visual metaphor of a real-world scene, often of a desktop, by combining high-resolution graphics, pointing devices, menu bars and other menus, overlapping windows, icons, and the object-action relationship.

Hop

One segment of a transmission path between adjacent nodes in a routed network.

Host

A physical or virtual computer system that hosts computer applications, with the host and the applications using storage.

Host bus adapter

A host bus adapter (HBA) is an interface card that connects a server to the SAN environment through its internal bus system, for example, PCIe. Typically, it is referred to the FC adapters.

Host cluster

A configured set of physical or virtual hosts that share one or more storage volumes to increase scalability or availability of computer applications.

Host ID

A host ID is a numeric identifier that is assigned to a group of host FC ports or internet Small Computer Systems Interface (iSCSI) hostnames for LUN mapping. For each host ID, SCSI IDs are mapped to volumes separately. The intent is to have a one-to-one relationship between hosts and host IDs, although this relationship cannot be policed.

Host mapping

Host mapping refers to the process of controlling which hosts have access to specific volumes within a cluster. Host mapping is equivalent to LUN masking.

Host object

A logical representation of a host within a storage system that is used to represent the host for configuration tasks.

Host zone

A zone that is defined in the SAN fabric in which the hosts can address the system.

Hot extent

A hot extent is a frequently accessed volume extent that gets a performance benefit if it is moved from an HDD onto a flash drive.

Hot spare node (IBM SAN Volume Controller only)

A hot spare node is an online SVC node that is defined in a cluster but not in any I/O group. During a failure of any online node in any I/O group of clusters, it is automatically swapped with this spare node. After the recovery of an original node finishes, the spare node returns to the standby spare status. This feature is not available for IBM FlashSystem 7200.

IBM FlashCore Module

The IBM FlashCore Module (FCM) is a family of high-performance flash drives. The FCM design uses the Non-Volatile Memory Express (NVMe) protocol, a Peripheral Component Interconnect Express (PCIe) Gen3 interface, and high-speed NAND memory to provide high throughput and input/output operations per second (IOPS) and low latency. FCM drives are available in different capacities. Hardware-based data compression and self-encryption are supported. The FCM drives are accessible from the front of the enclosure.

IBM HyperSwap

Pertaining to a function that provides continuous, transparent availability against storage errors and site failures, and is based on synchronous replication.

IBM Real-time Compression Appliance

IBM Real-time Compression Appliance® is an IBM integrated software function for storage space efficiency. The Random Access Compression Engine (RACE) compresses data on volumes in real time with minimal effect on performance.

IBM Remote Support Server and Client

IBM Remote Support Client is a software toolkit that is in IBM Storage System and opens a secured tunnel to the IBM Remote Support Server. IBM Remote Support Server is in the IBM network and collects key health check and troubleshooting information that is required by IBM support personnel.

IBM SAN Volume Controller

SVC is an appliance that is designed for attachment to various host computer systems. The SVC performs block-level virtualization of disk storage. IBM Spectrum Virtualize is a software engine of SVC (and IBM Storage System family) that performs block-level virtualization of disk storage.

IBM Security Key Lifecycle Manager

IBM Security Key Lifecycle Manager centralizes, simplifies, and automates the encryption key management process to help minimize risk and reduce operational costs of encryption key management.

Image mode

Image mode is an access mode that establishes a one-to-one mapping of extents in the storage pool (existing LUN or (image mode) MDisk) with the extents in the volume. See also “Managed mode” on page 911 and “Unconfigured mode” on page 921.

Image volume

An image volume is a volume in which a direct block-for-block conversion exists from the MDisk to the volume.

I/O group

Each pair of SVC cluster nodes is known as an input/output (I/O) group. An I/O group has a set of volumes that are associated with it that are presented to host systems. Each SVC node is associated with exactly one I/O group. The nodes in an I/O group provide a failover and failback function for each other.

Incremental restore operation

A copy operation where a local volume is modified to match a volume snapshot by reading from cloud storage only the parts of the volume snapshot that differ from the local volume.

Incremental snapshot

A type of volume snapshot where the changes to a local volume relative to the volume's previous snapshot are stored on cloud storage.

Input/output operations per second

A standard computing benchmark that is used to determine the best configuration settings for servers.

Input/output throttling rate

The maximum rate at which an I/O transaction is accepted for a volume.

Internal storage

Internal storage refers to an array of MDisks and drives that are held in IBM Storage System enclosures.

Internet Protocol

Internet Protocol (IP) is a protocol that routes data through a network or interconnected networks. This protocol acts as an intermediary between the higher protocol layers and the physical network.

Internet Small Computer Systems Interface

The iSCSI is a protocol that is used by a host system to manage iSCSI targets and iSCSI discovery. iSCSI initiators use the internet Storage Name Service (iSNS) protocol to locate the appropriate storage resources.

Internet Storage Name Service

The iSNS Protocol that is used by a host system to manage iSCSI targets and the automated iSCSI discovery, management, and configuration of iSCSI and FC devices. It was defined in Request for Comments (RFC) 4171.

Inter-switch link hop

An inter-switch link (ISL) is a connection between two switches and counted as one ISL hop. The number of hops is always counted on the shortest route between two N-ports (device connections). In an IBM Storage System environment, the number of ISL hops is counted on the shortest route between the pair of canisters that are farthest apart. The IBM Storage System supports a maximum of three ISL hops.

iSCSI alias

An alternative name for the iSCSI-attached host.

iSCSI initiator

An initiator functions as an iSCSI client. An initiator typically serves the same purpose to a computer as a SCSI bus adapter would, except that, instead of physically cabling SCSI devices (such as HDDs and tape changers), an iSCSI initiator sends SCSI commands over an IP network.

iSCSI name

A name that identifies an iSCSI target adapter or an iSCSI initiator adapter. An iSCSI name can be an iSCSI Qualified Name (IQN) or an extended-unique identifier (EUI). Typically, this identifier has the following format: `iqn.datecode.reverse domain`.

iSCSI Qualified Name

IQN refers to special names that identify both iSCSI initiators and targets. IQN is one of the three name formats that is provided by iSCSI. The IQN format is `iqn.<yyy-mm>.<reversed domain name>`. For example, the default for an IBM Storage System canister can be in the following format:

```
iqn.1986-03.com.ibm:2076.<clustername>.<nodename>
```

iSCSI session

The interaction (conversation) between an iSCSI Initiator and an iSCSI Target.

iSCSI target

An iSCSI target is a storage resource on an iSCSI server.

Just a bunch of disks

Just a bunch of disks (JBOD) is a group of HDDs that are not configured according to the RAID system to increase fault tolerance and improve data access performance.

Key server

- ▶ A server that negotiates the values that determine the characteristics of a dynamic virtual private network (VPN) connection that is established between two endpoints.
- ▶ See “Encryption key manager / server” on page 903.

Latency

The time interval between the initiation of a send operation by a source task and the completion of the matching receive operation by the target task. More generally, latency is the time between a task initiating data transfer and the time that transfer is recognized as complete at the data destination.

Least recently used

Least recently used (LRU) pertains to an algorithm that is used to identify and make available the cache space that contains the data that was least recently used.

Licensed capacity

The amount of capacity on a storage system that a user is entitled to configure.

License key

An alphanumeric code that activates a licensed function on a product.

License key file

A file that contains one or more licensed keys.

Lightweight Directory Access Protocol

Lightweight Directory Access Protocol (LDAP) is an open protocol that uses TCP/IP to provide access to directories that support an X.500 model. It does not incur the resource requirements of the more complex X.500 directory access protocol. For example, LDAP can be used to locate people, organizations, and other resources in an internet or intranet directory.

Local and remote fabric interconnect

The local fabric interconnect and the remote fabric interconnect are the SAN components that are used to connect the local and remote fabrics. Depending on the distance between the two fabrics, they can be single-mode optical fibers that are driven by long wave gigabit interface converters (GBICs) or small form factor pluggable (SFP), or more sophisticated components, such as channel extenders or special SFP modules that are used to extend the distance between SAN components.

Local fabric

The local fabric is composed of SAN components (switches, cables, and other components) that connect the components (nodes, hosts, and switches) of the local cluster together.

Logical drive

See “Volume” on page 922.

Logical unit and logical unit number

The LU is defined by the SCSI standards as a LUN. LUN is an abbreviation for an entity that exhibits disk-like behavior, such as a volume or an MDisk.

LUN masking

A process where a host object can detect more LUNs than it is intended to use, and the device-driver software masks the LUNs that are not to be used by this host.

Machine signature

A string of characters that identifies a system. A machine signature might be required to obtain a license key.

Managed disk

An MDisk is a SCSI disk that is presented by a RAID controller and managed by IBM Storage Systems. The MDisk is not visible to host systems on the SAN.

Managed mode

An access mode that enables virtualization functions to be performed. See also “Image mode” on page 908 and “Virtualized storage” on page 921.

Management node

A node that is used for configuring, administering, and monitoring a system.

Master volume

In most cases, the volume that contains a production copy of the data and that an application accesses. See also “Auxiliary volume” on page 896, and “Relationship” on page 917.

Maximum replication delay

Maximum replication delay is the number of seconds that MM or GM replication can delay a write operation to a volume.

MDisk group (Storage Pool)

See “Storage pool (MDisk group)” on page 919.

Media Access Control

In networking, the lower of two sublayers of the Open Systems Interconnection model data link layer. The Media Access Control (MAC) sublayer handles access to shared media, such as whether token passing or contention is used.

Megabytes per second

Megabytes per second (MBps) is a unit of data transfer rate equal to 1024 * 1024 bytes.

Metro Global Mirror

Metro Mirror Global (MGM) is a cascaded solution where MM synchronously copies data to the target site. This MM target is the source volume for GM that asynchronously copies data to a third site. This solution can provide disaster recovery (DR) with no data loss at GM distances when the intermediate site does not participate in the disaster that occurs at the production site.

Metro Mirror

MM is a method of synchronous replication that maintains data consistency across multiple volumes within the system. MM is used when the write latency that is caused by the distance between the source site and target site is acceptable to application performance.

Mirrored volume

A mirrored volume is a single virtual volume that has two physical volume copies. The primary physical copy is known within the IBM Storage System as copy 0 and the secondary copy is known within the IBM Storage System as copy 1.

N_Port ID Virtualization

N_Port ID Virtualization (NPIV) is an FC feature whereby multiple FC N_Port IDs can share a single physical N_Port.

Namespace Globally Unique Identifier

The Namespace Globally Unique Identifier (NGUID) is defined in the Identify Namespace data structure. The NGUID is composed of an IEEE organizationally unique identifier (OUI), an extension identifier, and a vendor-specific extension identifier. The extension identifier and vendor-specific extension identifier are both assigned by the vendor and can be considered as a single field. NGUID is defined in big endian format. The OUI field differs from the OUI Identifier, which is in little endian format.

Nearline SAS drive

A drive that combines the high capacity data storage technology of a SATA drive with the benefits of a SAS interface for improved connectivity.

Node

A single processing unit within a system. For redundancy, multiple nodes are typically deployed to make up a system.

Node canister

A node canister is a hardware unit that includes the node hardware, fabric and service interfaces, and SAS expansion ports. Node canisters are recognized on IBM Storage System products. In SVC, all these components are spread within the whole system chassis, so node canisters in SVC are not considered, but the node as a whole.

Node rescue

The process by which a node with no valid software is installed on its HDD, and can copy software from another node that is connected to the same FC fabric.

Non-Volatile Memory Express

NVMe or Non-Volatile Memory Host Controller Interface Specification (NVMHCIS) is an open logical-device interface specification for accessing non-volatile storage media that is attached through a PCIe bus.

NVMe Qualified Name

NVMe Qualified Names (NQNs) are used to uniquely describe a host or NVM subsystem for identification and authentication. The NQN for the NVM subsystem is specified in the Identify Controller data structure. An NQN is permanent for the lifetime of the host or NVM subsystem.

Object-Based Access Control

See “Ownership Groups”.

Object storage

Object storage is a general term that refers to the entity in which cloud object storage organizes, manages, and stores units of storage or just *objects*.

Overprovisioned

See “Capacity” on page 897.

Overprovisioned ratio

See “Capacity” on page 897.

Oversubscription

Oversubscription refers to the ratio of the sum of the traffic on the initiator N-port connections to the traffic on the most heavily loaded ISLs, where more than one connection is used between these switches. Oversubscription assumes a symmetrical network, and a specific workload that is applied equally from all initiators and sent equally to all targets. A symmetrical network means that all the initiators are connected at the same level, and all the controllers are connected at the same level.

Ownership Groups

The Ownership Groups feature provides a method of implementing a multi-tenant solution on the system. Ownership groups enable the allocation of storage resources to several independent tenants with the assurance that one tenant cannot access resources that are associated with another tenant. Ownership groups restrict access for users in the ownership group to only those objects that are defined within that ownership group.

Parent pool

Parent pools receive their capacity from MDisks. All MDisks in a pool are split into extents of the same size. Volumes are created from the extents that are available in the pool. You can add MDisks to a pool at any time either to increase the number of extents that are available for new volume copies or to expand existing volume copies. The system automatically balances volume extents between the MDisks to provide the best performance to the volumes. See also “Child pool” on page 899.

Partner node

The other node that is in the I/O group to which this node belongs.

Partnership

In MM or GM operations, the relationship between two clustered systems. In a clustered-system partnership, one system is defined as the local system and the other system as the remote system.

Performance group

A collection of volumes that is assigned the same performance characteristics. See also “Performance policy”.

Performance policy

A policy that specifies performance characteristics, for example quality of service (QoS). See also “Pool”.

Point-in-time copy

A PiT copy is an instantaneous copy that the FlashCopy service makes of the source volume. See also “FlashCopy service” on page 905.

Pool

See “Storage pool (MDisk group)” on page 919.

Pool pair

Two storage pools that are required to balance workload. Each storage pool is controlled by a separate node.

Preferred node

When you create a volume, you can specify a preferred node. Many of the multipathing driver implementations that the system supports use this information to direct I/O to the preferred node. The other node in the I/O group is used only if the preferred node is not accessible. If you do not specify a preferred node for a volume, the system selects the node in the I/O group that has the fewest volumes to be the preferred node. After the preferred node is chosen, it can be changed only when the volume is moved to a different I/O group. The management GUI provides a wizard that moves volumes between I/O groups without disrupting host I/O operations.

Preparing phase

Before you start the FlashCopy process, you must prepare a FlashCopy mapping. The preparing phase flushes a volume’s data from cache in preparation for the FlashCopy operation.

Primary volume

In a stand-alone MM or GM relationship, the target of write operations that are issued by the host application. See also “Relationship” on page 917.

Priority flow control

Priority flow control (PFC) is a link-level flow control mechanism that is based on IEEE standard 802.1Qbb. PFC operates on individual priorities. Instead of pausing all traffic on a link, PFC is used to selectively pause traffic according to its class.

Private fabric

Configure one SAN per fabric so that it is dedicated for node-to-node communication. This SAN is referred to as a private SAN.

Provisioned capacity

See “Capacity” on page 897.

Provisioning group

A provisioning group is an object that represents a set of MDisks that share physical resources. Provisioning groups are used for capacity reporting and monitoring of overprovisioned storage resources.

Public fabric

A public fabric is where you configure one SAN per fabric so that it is dedicated for host attachment, storage system attachment, and RC operations. This SAN is referred to as a public SAN. You can configure the public SAN to enable IBM Storage System node-to-node communication also. You can optionally use the `-localportfcmask` parameter of the `chsystem` command to constrain the node-to-node communication to use only the private SAN.

Qualifier

- ▶ A value that provides more information about a class, association, indication, method, method parameter, instance, property, or reference.
- ▶ A modifier that makes a name unique.

Queue depth

The number of input/output (I/O) operations that can be run in parallel on a device.

Quorum disk

A disk that contains a reserved area that is used exclusively for system management. The quorum disk is accessed when it is necessary to determine which half of the clustered system continues to read and write data. Quorum disks can either be MDisks or drives.

Quorum index

The quorum index is the pointer that indicates the order that is used to resolve a tie. Nodes attempt to lock the first quorum disk (index 0), followed by the next disk (index 1), and then the last disk (index 2). The tie is broken by the node that locks them first.

Quota

The amount of disk space and number of files and directories that are assigned as upper limits for a specified user, group of users, or file set.

Random Access Compression Engine

The RACE engine compresses data on volumes in real time with minimal effect on performance. See “Compression” on page 900 or “IBM Real-time Compression Appliance” on page 908.

RAID controller

See “Node canister” on page 912.

Raw capacity

See “Capacity” on page 897.

Real capacity

Real capacity is the amount of storage that is allocated to a volume copy from a storage pool. See also “Capacity” on page 897.

Redundant array of independent disks

RAID refers to two or more physical disk drives that are combined in an array in a certain way, which incorporates a RAID level for failure protection or better performance. The most common RAID levels are 0, 1, 5, 6, and 10. Some storage administrators refer to the RAID group as traditional RAID (TRAID). For distributed redundant array of independent disks (DRAID), see “Distributed redundant array of independent disks” on page 902.

RAID 0

A data striping technique, which is commonly called RAID Level 0 or RAID 0 because of its similarity to common, RAID, data-mapping techniques. However, it includes no data protection, so the appellation RAID is a misnomer. RAID 0 is also known as data striping.

RAID 1

RAID 1 is a mirroring technique that is used on a storage array in which two or more identical copies of data are maintained on separate mirrored disks.

RAID 10

A collection of two or more physical drives that present to the host an image of one or more drives. In the event of a physical device failure, the data can be read or regenerated from the other drives in the RAID due to data redundancy.

RAID 5

RAID 5 is an array that has a data stripe, which includes a single logical parity drive. The parity check data is distributed across all the disks of the array.

RAID 6

RAID 6 is a RAID level that has two logical parity drives per stripe, which are calculated with different algorithms. Therefore, this level can continue to process read and write requests to all the array’s virtual disks (virtual disks (VDisks)) in the presence of two concurrent disk failures.

Real capacity

The amount of storage that is allocated to a volume copy from a storage pool.

Rebuild area

Reserved capacity that is distributed across all drives in a RAID. If a drive in the array fails, the lost array data is systematically restored into the reserved capacity, returning redundancy to the array. The duration of the restoration process is minimized because all drive members simultaneously participate in restoring the data. See also “Distributed redundant array of independent disks” on page 902.

Reclaimable (or reclaimed) capacity

Reclaimable data is the capacity that is no longer needed. Reclaimable capacity is created when data is overwritten and the new data is stored in a new location, when data is marked as unneeded by a host by using the SCSI **UNMAP** command, or when a volume is deleted.

Recovery key

See “Encryption recovery key” on page 903.

Redundant storage area network

A redundant SAN is a SAN configuration in which there is no single point of failure (SPOF). Therefore, data traffic continues no matter what component fails. Connectivity between the devices within the SAN is maintained (although possibly with degraded performance) when an error occurs. A redundant SAN design is normally achieved by splitting the SAN into two independent counterpart SANs (two SAN fabrics). In this configuration, if one path of the counterpart SAN is destroyed, the other counterpart SAN path keeps functioning. See also “Counterpart SAN” on page 901.

Relationship

In MM or GM, a relationship is the association between a master volume and an auxiliary volume. These volumes also have the attributes of a primary or secondary volume. See also “Auxiliary volume” on page 896, “Master volume” on page 911, “Primary volume” on page 914, and “Secondary volume” on page 917.

Reliability, availability, and serviceability

Reliability, availability, and serviceability (RAS) are a combination of design methodologies, system policies, and intrinsic capabilities that, when taken together, balance improved hardware availability with the costs that are required to achieve it.

Reliability is the degree to which the hardware remains free of faults. Availability is the ability of the system to continue operating despite predicted or experienced faults. Serviceability is how efficiently and nondisruptively broken hardware can be fixed.

Remote Copy

See “Global Mirror” on page 906 and “Metro Mirror” on page 912.

Remote fabric

The remote fabric is composed of SAN components (switches, cables, and other components) that connect the components (nodes, hosts, and switches) of the remote cluster together. Significant distances can exist between the components in the local cluster and those components in the remote cluster.

SCSI initiator

The SCSI initiator is the system component that initiates communications with attached targets.

SCSI target

A device that acts as a subordinate to a SCSI initiator and consists of a set of one or more LUs, each with an assigned LUN. The LUs on the SCSI target are typically I/O devices.

Secondary volume

Pertinent to RC, the volume in a relationship that contains a copy of data that is written by the host application to the primary volume.

Secure Copy Protocol

Secure Copy Protocol (SCP) is the secure transfer of computer files between a local and a remote host or between two remote hosts by using the Secure Shell (SSH) protocol.

Secure Sockets Layer certificate

Secure Sockets Layer (SSL) is the standard security technology for establishing an encrypted link between a web server and a browser. This link ensures that all data that is passed between the web server and browsers remain private. To create an SSL connection, a web server requires an SSL certificate.

Sequential volume

A volume that uses extents from a single MDisk.

Serial-attached SCSI

Serial-attached SCSI (SAS) is a method that is used in accessing computer peripheral devices that employs a serial (1 bit at a time) means of digital data transfer over thin cables. The method is specified in the ANSI standard that is called SAS. In the business enterprise, SAS is useful for access to mass storage devices, external HDDs.

Service Location Protocol

The Service Location Protocol (SLP) is an internet service discovery protocol that enables computers and other devices to find services in a local area network (LAN) without prior configuration. It was defined in RFC 2608.

Simple Network Management Protocol

Simple Network Management Protocol (SNMP) is a set of protocols for monitoring systems and devices in complex networks. Information about managed devices is defined and stored in a Management Information Base (MIB).

Small Computer System Interface

SCSI is an ANSI-standard electronic interface with which PCs can communicate with peripheral hardware, such as disk drives, tape drives, CD-ROM drives, printers, and scanners, faster and more flexibly than with previous interfaces.

Snapshot

A snapshot is an image backup type that consists of a PiT view of a volume.

Solid-state drive

An SSD or flash drive is a disk that is made from solid-state memory and therefore has no moving parts. Most SSDs use NAND-based flash memory technology. It is defined to the IBM Storage System as a disk tier generic_ SSD.

Space efficient

See “Thin provisioning” on page 920.

Space-efficient virtual disk

See “Thin-provisioned volume” on page 920.

Spare

An extra storage component, such as a drive or tape, that is predesignated for use as a replacement for a failed component.

Spare drive

A drive that is reserved in an array for rebuilding a failed drive in a RAID. If a drive fails in a RAID, a spare drive from within that device adapter pair is selected to rebuild it.

Spare goal

The optimal number of spares that are needed to protect the drives in the array from failures. The system logs a warning event when the number of spares that protect the array drops below this number.

Space-efficient volume

See “Thin-provisioned volume” on page 920.

Stand-alone relationship

In FlashCopy, MM, and GM, relationships that do not belong to a consistency group and that have a null consistency-group attribute.

Statesave

Binary data collection that is used for a problem determination by IBM service support.

Storage area network

A SAN is a dedicated storage network that is tailored to a specific environment, which combines servers, systems, storage products, networking products, software, and services.

Storage-class memory

Storage-class memory (SCM) is a type of NAND flash that includes a power source to ensure that data is not lost due to a system crash or power failure. SCM treats non-volatile memory as DRAM and includes it in the memory space of the server. Access to data in that space is quicker than access to data in local, PCI-connected SSDs, direct-attached HDDs, or external storage arrays. SCM read/write technology is up to 10 times faster than NAND flash drives and is more durable.

Storage capacity unit

Storage capacity unit (SCU) is an SVC license metric that measures the managed capacity so that the price is differentiated by the technology that is used to store the data.

Storage node

A component of a storage system that provides internal storage or a connection to one or more external storage systems.

Storage pool (MDisk group)

A storage pool is a collection of storage capacity, which is made up of MDisks, that provides the pool of storage capacity for a specific set of volumes. A storage pool can contain more than one tier of disk, which is known as a multitier storage pool, which is a prerequisite of Easy Tier automatic data placement.

Striped

Pertaining to a volume that is created from multiple MDisks that are in the storage pool. Extents are allocated on the MDisks in the order that is specified.

Support Assistance

A function that is used to provide support personnel remote access to the system to perform troubleshooting and maintenance tasks.

Symmetric virtualization

Symmetric virtualization is a virtualization technique in which the physical storage, in the form of a RAID, is split into smaller chunks of storage that are known as extents. These extents are then concatenated by using various policies to make volumes. See also “Asymmetric virtualization” on page 896.

Synchronous replication

Synchronous replication is a type of replication in which the application write operation is made to both the source volume and target volume before control is given back to the application. See also “Asynchronous replication” on page 896.

Syslog

A standard for transmitting and storing log messages from many sources to a centralized location to enhance system management.

T10 DIF

T10 DIF is a *Data Integrity Field* (DIF) extension to SCSI to enable end-to-end protection of data from a host application to physical media.

Thin-provisioning savings

See “Capacity” on page 897.

Thin-provisioned volume

A thin-provisioned volume is a volume that allocates storage when data is written to it.

Thin provisioning

Thin provisioning refers to the ability to define storage, usually a storage pool or volume, with a “logical” capacity size that is larger than the actual physical capacity that is assigned to that pool or volume. Therefore, a thin-provisioned volume is a volume with a virtual capacity that differs from its real capacity.

Throttles

Throttling is a mechanism to control the amount of resources that are used when the system is processing I/Os on supported objects. The system supports throttles on hosts, host clusters, volumes, copy offload operations, and storage pools. If a throttle limit is defined, the system either processes the I/O for that object or delays the processing of the I/O to free resources for more critical I/O operations.

Throughput

A measure of the amount of information that is transmitted over a network in a period. Throughput is measured in bits per second (bps), kilobits per second (Kbps), or megabits per second (Mbps).

Tie-breaker

When a cluster is split into two groups of nodes, the role of tie-breaker in a quorum device decides which group continues to operate as the system and handle all I/O requests.

Transparent Cloud Tiering

Transparent Cloud Tiering (TCT) is a separately installable feature of IBM Spectrum Scale that provides a native cloud storage tier.

Trial license

A temporary entitlement to use a licensed function.

Total capacity savings

See “Capacity” on page 897.

Unconfigured mode

An access mode in which an external storage MDisk is not configured in the system, so no operations can be performed. See also “Image mode” on page 908 and “Managed mode” on page 911.

Unique identifier

A unique identifier (UID) is an identifier that is assigned to storage system LUs when they are created. It is used to identify the LU regardless of the LUN, the status of the LU, or whether alternative paths exist to the same device. Typically, a UID is used only once.

Usable capacity

The amount of capacity that is provided for storing data on a system, pool, array, or MDisk after formatting and RAID techniques are applied.

Used capacity

The amount of usable capacity that is taken up by data or capacity in a system, pool, array, or MDisk after data reduction techniques are applied.

VDisk-to-host mapping

See “Host mapping” on page 907.

Virtual capacity

The amount of storage that is available. In a thin-provisioned volume, the virtual capacity can be different from the real capacity. In a standard volume, the virtual capacity and real capacity are the same.

Virtual disk

See “Volume” on page 922.

Virtualization

In the storage industry, virtualization is a concept in which a pool of storage is created that contains several storage systems. Storage systems from various vendors can be used. The pool can be split into volumes that are visible to the host systems that use them. See also “Capacity licensing” on page 898.

Virtualized storage

Virtualized storage is physical storage that has virtualization techniques that are applied to it by a virtualization engine.

Virtual local area network

Virtual local area network (VLAN) tagging separates network traffic at the layer 2 level for Ethernet transport. The system supports VLAN configuration on both IPv4 and IPv6 connections.

Virtual storage area network

A virtual storage area network (VSAN) is a logical fabric entity that is defined within the SAN. It can be defined on a single physical SAN switch or across multiple physical switched or directors. In VMware terminology, the VSAN is defined as a logical layer of storage capacity that is built from physical disk drives that are attached directly into the Elastic Sky X Integrated (ESXi) hosts. This solution is not considered within the scope of this publication.

Vital product data

Vital product data (VPD) is information that uniquely defines system, hardware, software, and microcode elements of a processing system.

Volume

A volume is an IBM Storage System logical device that appears to host systems that are attached to the SAN as a SCSI disk. Each volume is associated with exactly one I/O group. A volume has a preferred node within the I/O group.

Volume copy

A volume copy is a physical copy of the data that is stored on a volume. Mirrored volumes have two copies. Non-mirrored volumes have one copy.

Volume protection

To prevent active volumes or host mappings from inadvertent deletion, the system supports a global setting that prevents these objects from being deleted if the system detects that they have recent I/O activity. When you delete a volume, the system checks to verify whether it is part of a host mapping, FlashCopy mapping, or an RC relationship. In these cases, the system fails to delete the volume unless the **-force** parameter is specified. Using the **-force** parameter can lead to unintentional deletions of volumes that are still active. Active means that the system detected recent I/O activity to the volume from any host.

Volume snapshot

A collection of objects on a cloud storage account that represents the data of a volume at a particular time.

Worldwide ID

A worldwide ID (WWID) is a name identifier that is unique worldwide and that is represented by a 64-bit value that includes the IEEE-assigned OUI.

Worldwide name

Worldwide name (WWN) is a 64-bit, unsigned name identifier that is unique.

Worldwide node name

Worldwide node name (WWNN) is a unique 64-bit identifier for a host containing an FC port. See also "Worldwide port name" on page 923.

Worldwide port name

Worldwide port name (WWPN) is a unique 64-bit identifier that is associated with an FC adapter port. The WWPN is assigned in an implementation-independent and protocol-independent manner. See also “Worldwide node name” on page 922.

Write-through mode

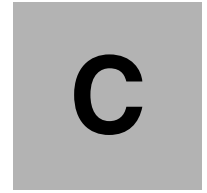
Write-through mode is a process in which data is written to a storage device while the data is cached.

Written capacity

See “Capacity” on page 897.

Zoning

The grouping of multiple ports to form a virtual and private storage network. Ports that are members of a zone can communicate with each other, but are isolated from ports in other zones. See also “Fibre Channel” on page 904.



Command-line interface setup

This appendix describes access configuration to the command-line interface (CLI) by using the local Secure Shell (SSH) authentication method.

CLI setup

The IBM Spectrum Virtualize system features a powerful CLI that offers more options and flexibility as compared to the GUI. This appendix describes how to configure a management system by using the SSH protocol to connect to the IBM Spectrum Virtualize system for issuing commands by using the CLI.

For more information about the CLI, see [IBM Documentation](#).

Note: If a task completes in the GUI, the associated CLI command is always displayed in the details, as shown throughout this book.

In the IBM Spectrum Virtualize GUI, authentication is performed by supplying a username and password. The CLI uses SSH to connect from a host to the IBM Spectrum Virtualize system. A private and a public key pair or username and password is necessary.

Using SSH keys with a passphrase is more secure than a login with a username and password because authenticating to a system requires the private key and the passphrase. By using the other method, only the password is required to obtain access to the system.

When SSH keys are used without a passphrase, it becomes easier to log in to a system because you must provide only the private key when performing the login and you are not prompted for password. This option is less secure than using SSH keys with a passphrase.

To enable CLI access with SSH keys, complete the following steps:

1. Generate a public key and a private key as a pair.
2. Upload a public key to the IBM Spectrum Virtualize system by using the GUI.
3. Configure a client SSH tool to authenticate with the private key.
4. Establish a secure connection between the client and the system.

SSH is the communication vehicle between the management workstation and the IBM Spectrum Virtualize system. The SSH client provides a secure environment from which to connect to a remote machine. It uses the principles of public and private keys for authentication.

SSH keys are generated by the SSH client software. The SSH keys include a public key, which is uploaded and maintained by the storage system, and a private key, which is kept private on the workstation that is running the SSH client. These keys authorize specific users to access the administration and service functions on the system.

Each key pair is associated with a user-defined ID string that can consist of up to 256 characters. Up to 100 keys can be stored on the system. New IDs and keys can be added, and unwanted IDs and keys can be deleted. To use the CLI, an SSH client must be installed on that system. To use the CLI with SSH keys, the SSH client is required. An SSH key pair also must be generated on the client system, and the client's SSH public key must be stored on the IBM Spectrum Virtualize systems.

Basic setup on a Windows host

The SSH client on a Windows host that is used in this book is PuTTY. A PuTTY key generator can also be used to generate the private and public key pair. The PuTTY client can be downloaded at no cost from [Download PuTTY](#).

Download the following tools:

- ▶ PuTTY SSH client: `putty.exe`
- ▶ PuTTY key generator: `puttygen.exe`

Generating a public and private key pair

To generate a public and private key pair, complete the following steps:

1. Start the PuTTY key generator to generate the public and private key pair, as shown in Figure C-1.

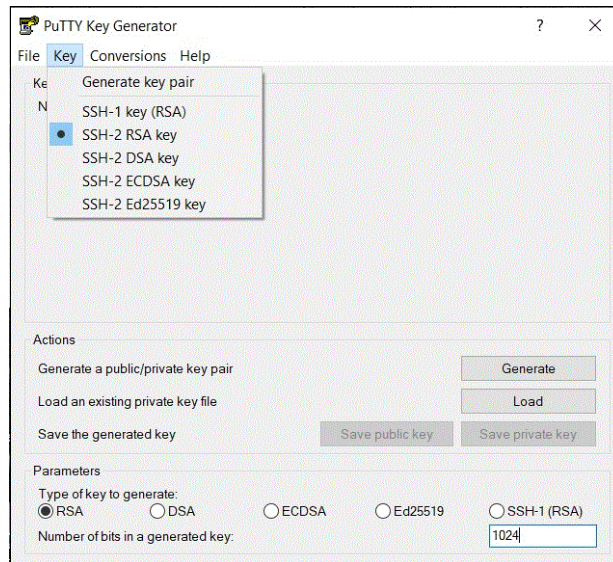


Figure C-1 PuTTY key generator

Select the following options:

- **SSH-2 RSA**
- Number of bits in a generated key: **1024**

Note: Larger SSH keys, such as 2048 bits, are also supported.

2. Click **Generate** and move the cursor over the blank area to generate keys (see Figure C-2).

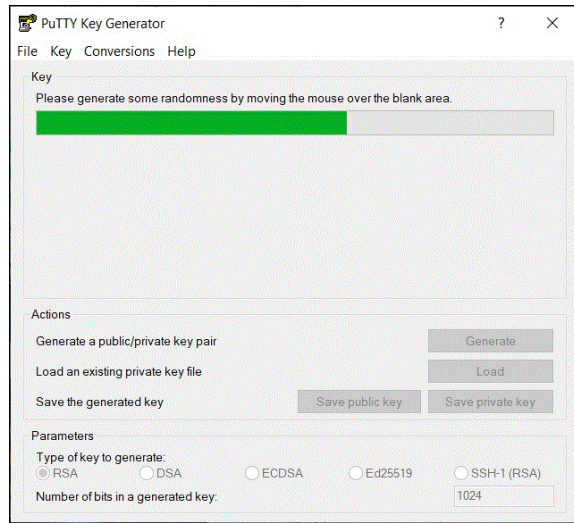


Figure C-2 Generating keys

To generate keys: The blank area that is indicated by the message is the large blank rectangle in the GUI inside the Key field. Continue to move the mouse pointer over the blank area until the progress bar reaches the far right. This action generates random characters based on the cursor location to create a unique key pair.

3. After the keys are generated, save them for later use. Click **Save public key**.
4. You are prompted to enter a name (for example, `sshkey.pub`) and a location for the public key (for example, `C:\Keys\`). Enter this information and click **Save**.

Ensure that you record the SSH public key name and location because this information must be specified later.

Public key extension: By default, the PuTTY key generator saves the public key with no extension. Use the string `pub` for naming the public key. For example, add the extension `.pub` to the name of the file to easily differentiate the SSH public key from the SSH private key.

5. Click **Save private key**. A warning message is displayed (see Figure C-3). Click **Yes** to save the private key without a passphrase.

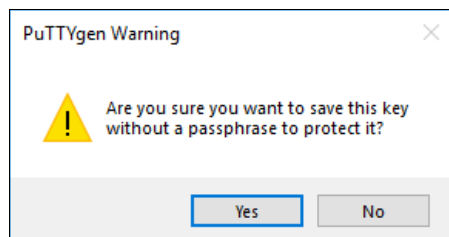


Figure C-3 Confirming the security warning

Note: It is possible to use a passphrase for an SSH key. Although this action increases security, it generates an extra step to log in with the SSH key because it requires the passphrase input.

6. When prompted, enter a name (for example, `sshkey.ppk`), select a secure place as the location, and click **Save**.

Private Key Extension: The PuTTY key generator saves the PuTTY private key (PPK) with the `.ppk` extension. This is a proprietary format for PuTTY and the keys are not interchangeable with OpenSSH clients. There is a utility to convert keys between PuTTY and OpenSSH if you want to use the same keys between the two environments.

7. Close the PuTTY key generator.

Uploading the SSH public key to the IBM Storage System

After you create your SSH key pair, upload your SSH public key onto the IBM Storage System. Complete the following steps:

1. Open the user section in the GUI, as shown in Figure C-4.

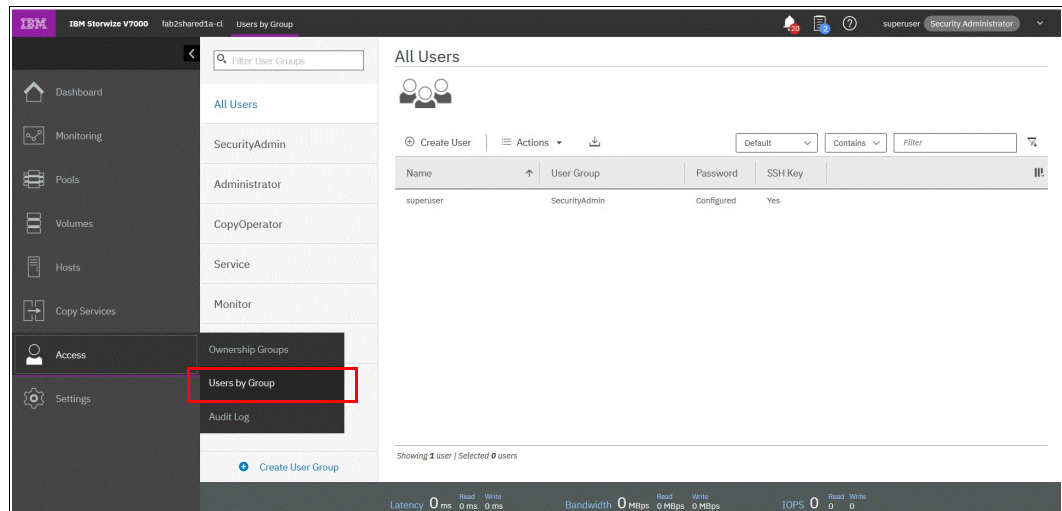


Figure C-4 Opening the user section

- Right-click the username for which you want to upload the key and click **Properties** (see Figure C-5).

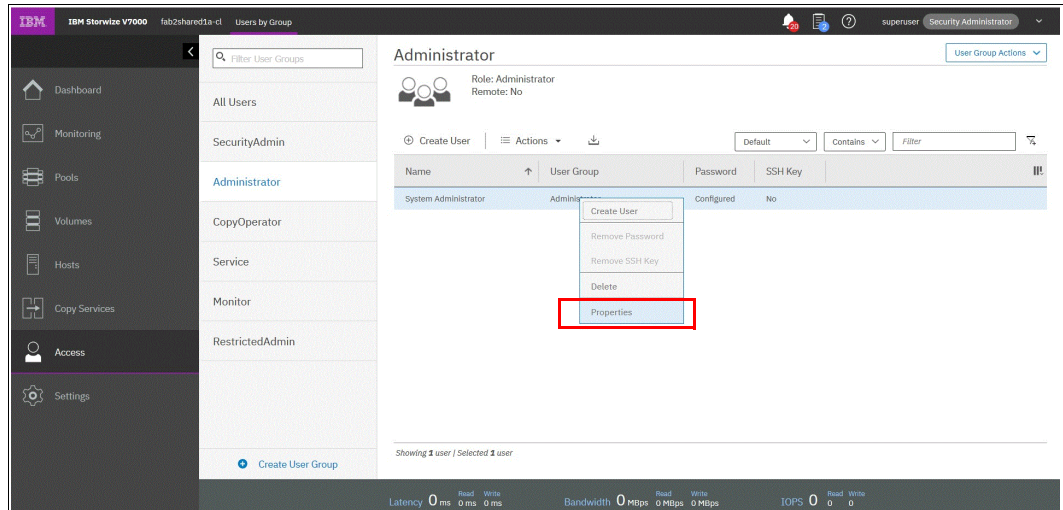


Figure C-5 User properties

- To upload the public key, click **Browse**, open the folder where you stored the public SSH key, and select the key.
- Click **OK** and the key is uploaded, as shown in Figure C-6.

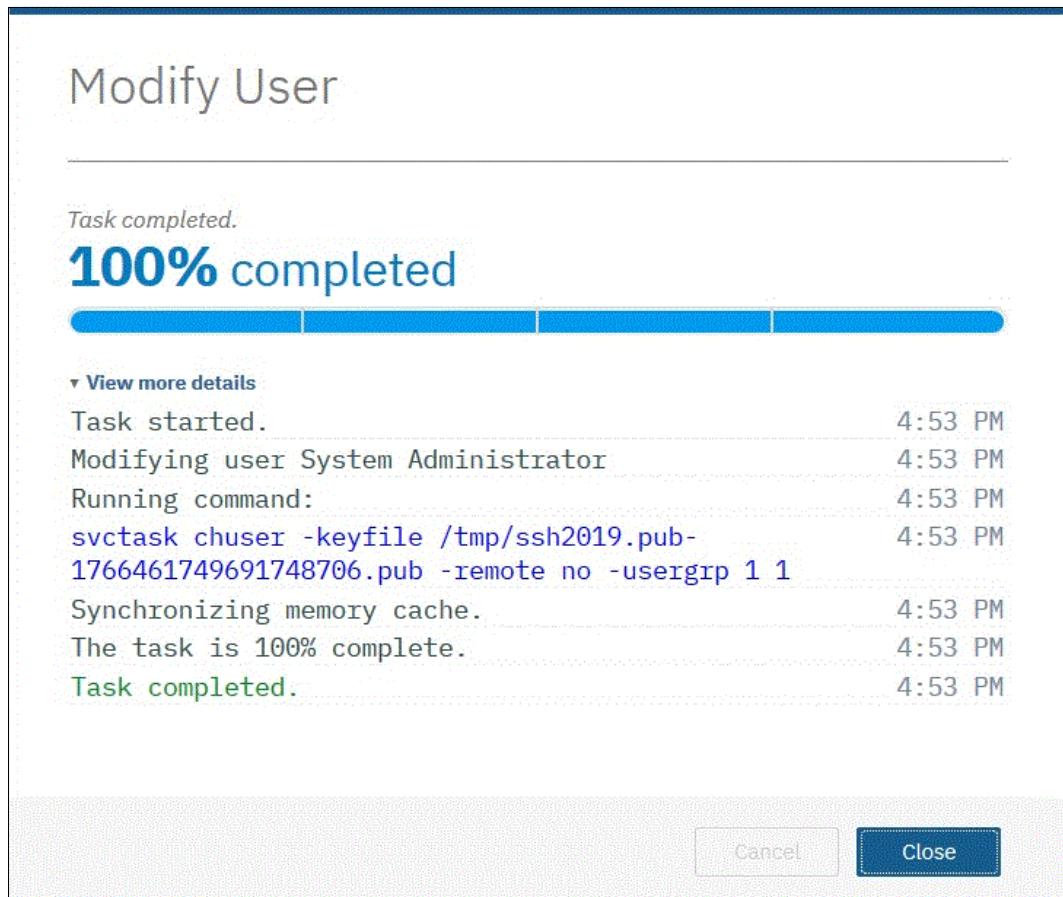


Figure C-6 Confirming the SSH key upload

5. Check in the GUI to ensure that the SSH key is imported successfully (see Figure C-7).

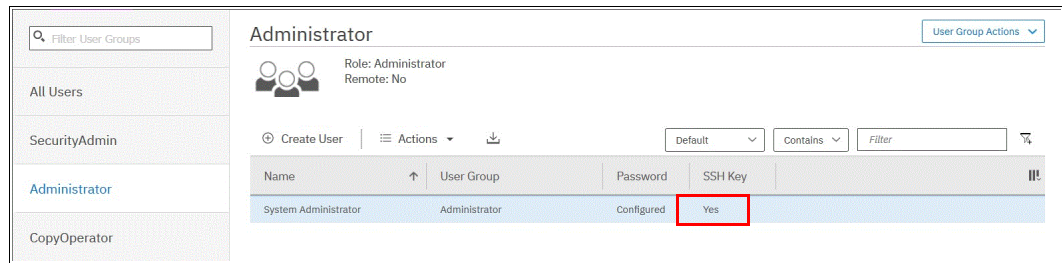


Figure C-7 Key successfully imported

Configuring the SSH client

Before the CLI can be used, the SSH client must be configured. Complete the following steps:

1. Start PuTTY. The PuTTY Configuration window opens (see Figure C-8).

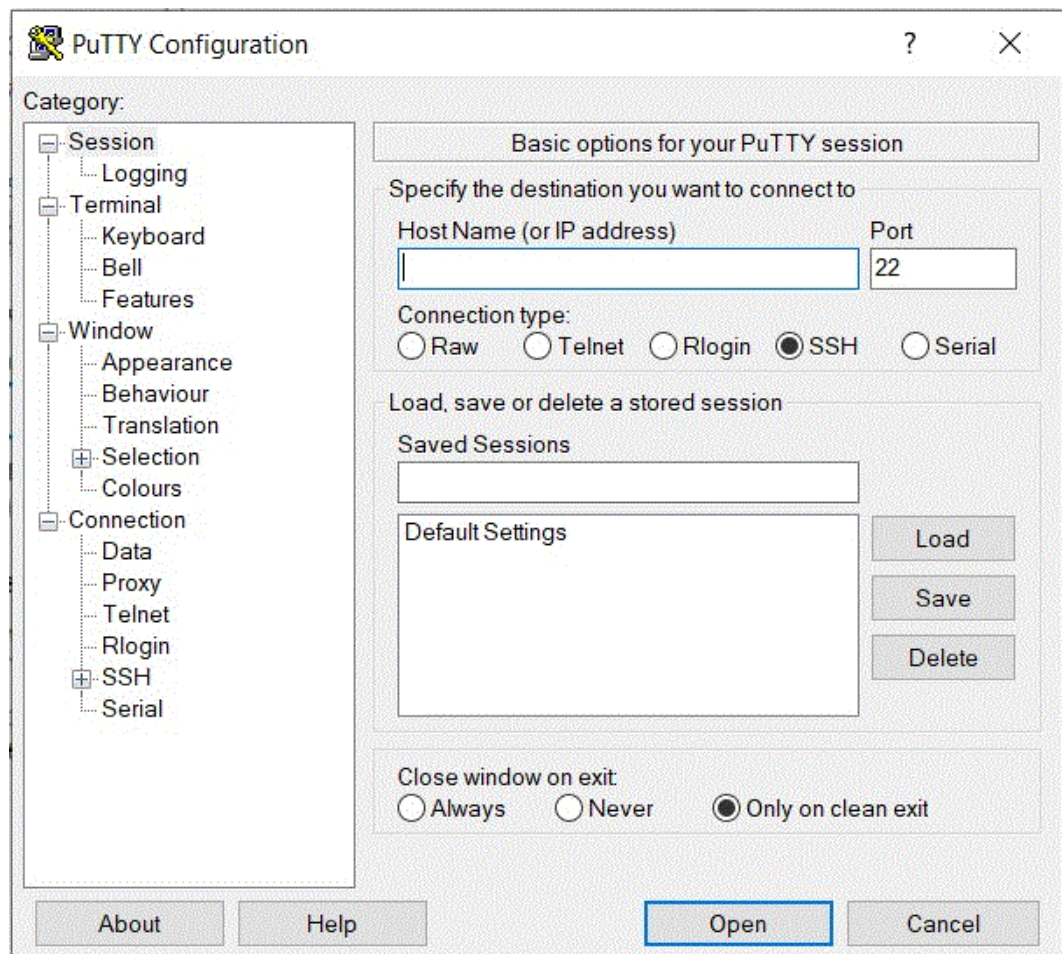


Figure C-8 PuTTY Configuration

2. In the upper right, select **SSH** as the connection type. In the “Close window on exit” section, select **Only on clean exit** (see Figure C-9 on page 932), which ensures that if any connection errors occur that they are displayed on the user’s window.

- In the Category window, on the left side of the PuTTY Configuration window, select **Connection** → **Data**, as shown on Figure C-9. In the “Auto-login username” field, enter the IBM Spectrum Virtualize user ID that was used when uploading the public key. The admin account was used in the example that is shown in Figure C-5 on page 930.

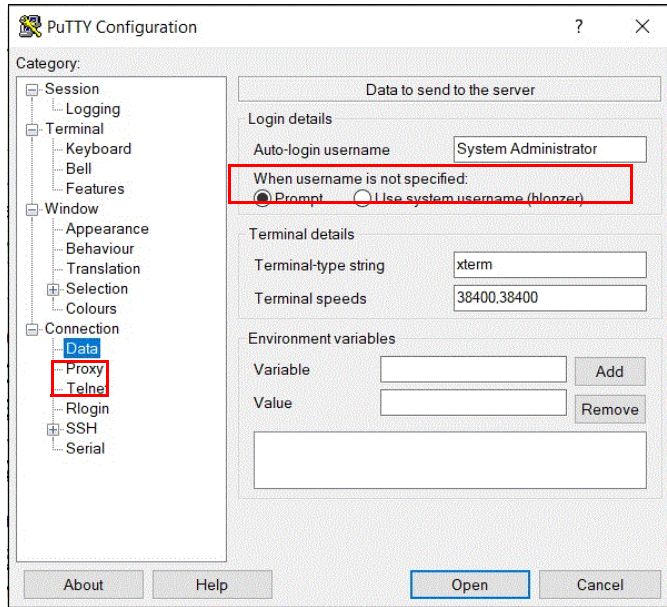


Figure C-9 PuTTY Auto-login username

- In the Category window, on the left side of the PuTTY Configuration window (see Figure C-10), select **Connection** → **SSH** to open the PuTTY SSH Configuration window. In the SSH protocol version section, select **2**.

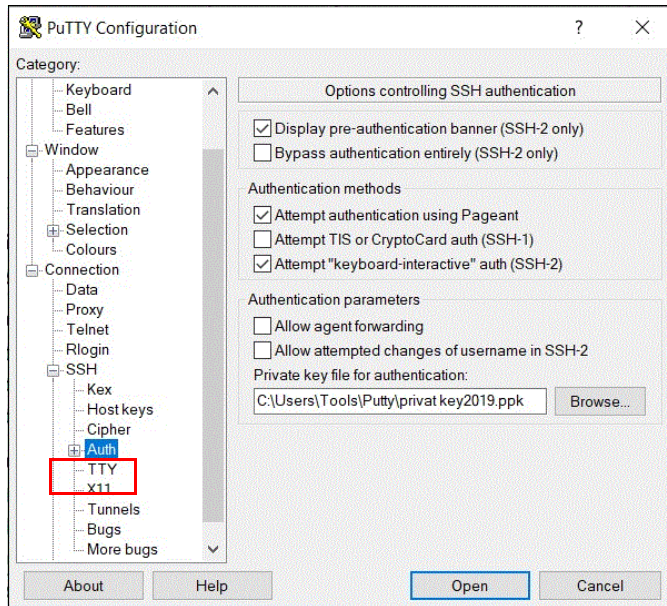


Figure C-10 SSH protocol Version 2

- In the Category window on the left, select **Connection** → **SSH** → **Auth**. More options are displayed for controlling SSH authentication.

- In the “Private key file for authentication” field in Figure C-11, browse to or enter the fully qualified directory path and file name of the SSH client private key file that was created (in this example, C:\Users\Tools\Putty\privatekey2019.ppk is used).

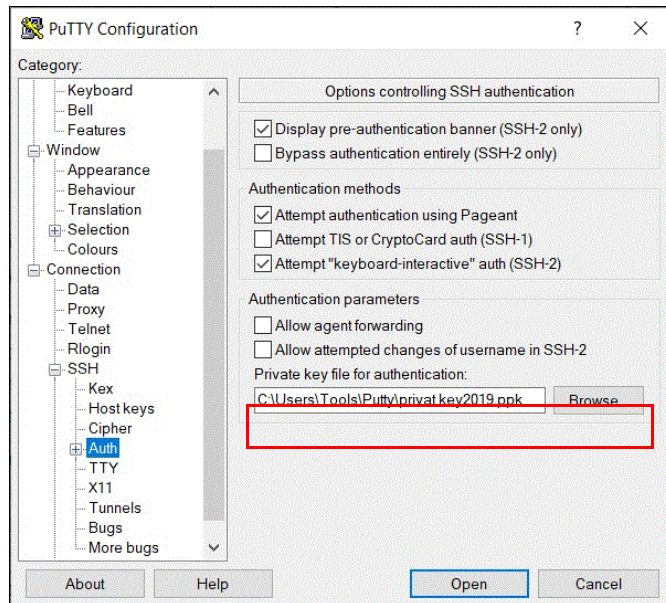


Figure C-11 SSH authentication

- In the Category window, click **Session** to return to the “Basic options for your PuTTY session” view.
- Enter the following information in the fields in the right pane (see Figure C-12):
 - Host Name (or IP address): Specify the hostname or system IP address of the IBM Spectrum Virtualize system.
 - Saved Sessions: Enter a session name.
- Click **Save** to save the new session (Figure C-12).

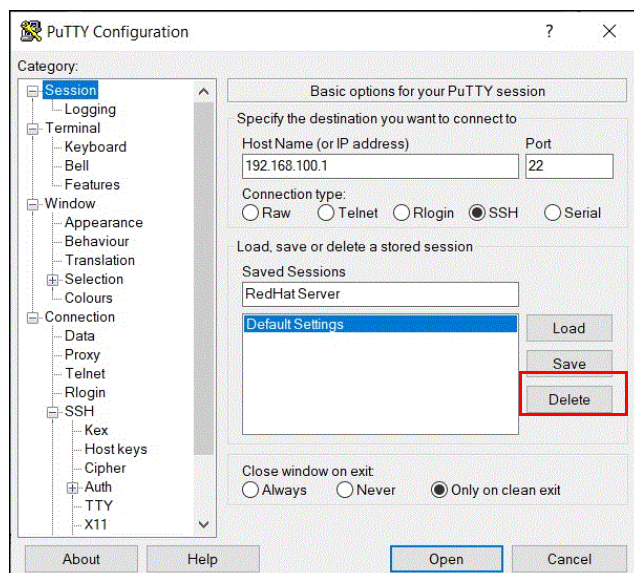


Figure C-12 Session information

10. Select the new session and click **Open** to connect to the IBM Spectrum Virtualize system, as shown in Figure C-13.

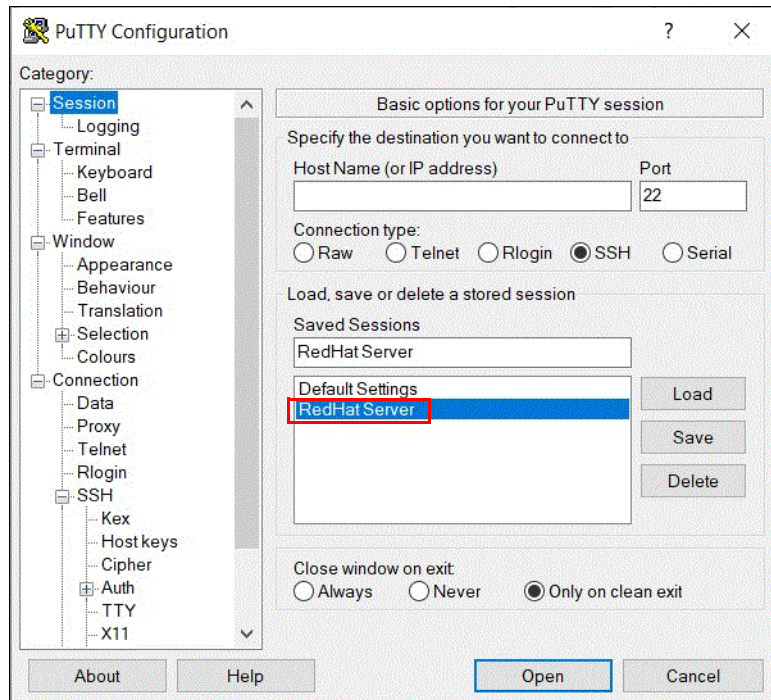


Figure C-13 Connecting to a system

11. If a PuTTY Security Alert opens as shown in Figure C-14, confirm it by clicking **Yes**.

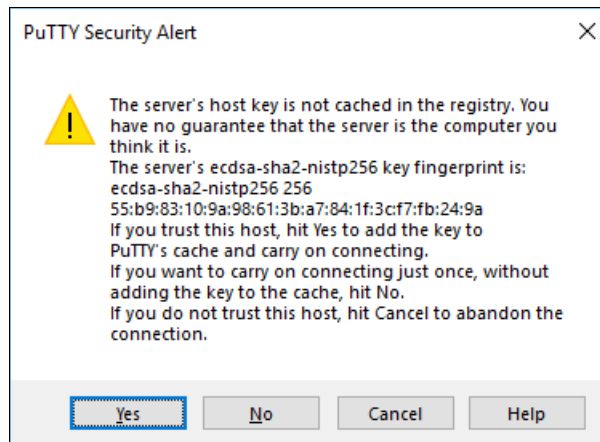


Figure C-14 Confirming the security alert

12. As shown in Figure C-15, PuTTY now connects to the system automatically by using the user ID that was specified earlier, without prompting for password.

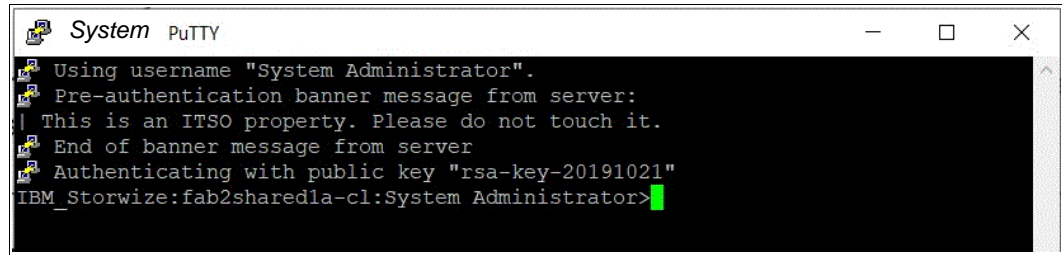


Figure C-15 PuTTY login

The CLI is now configured for IBM Spectrum Virtualize system administration.

Basic setup on a Mac, UNIX, or Linux host

The OpenSSH client is the most common tool that is used on Mac, UNIX, or Linux operating systems (OSs). It is installed by default on most of these types of OSs. If OpenSSH is not installed on your system, download it from [OpenSSH: Portable Release](#).

The OpenSSH suite consists of various tools. The following tools are used to generate the SSH keys, transfer the SSH keys to a remote system, and establish a connection to IBM Spectrum Virtualize device by using SSH:

- ▶ ssh: OpenSSH SSH client
- ▶ ssh-keygen: Tool to generate SSH keys
- ▶ scp: Tool to transfer files between hosts

Generating a public and private key pair

To generate a public and private key pair to connect to an IBM Spectrum Virtualize system without entering the user password, run the **ssh-keygen** tool, as shown in Example C-1.

Example: C-1 SSH keys generation with ssh-keygen

```
# ssh-keygen -t rsa -b 1024
Generating public/private rsa key pair.
Enter file in which to save the key (/.ssh/id_rsa): /.ssh/sshkey
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /.ssh/sshkey.
Your public key has been saved in /.ssh/sshkey.pub.
The key fingerprint is:
55:5e:5e:09:db:a4:11:01:b9:57:96:74:0c:85:ed:5b root@hostname.ibm.com
The key's randomart image is:
+--[ RSA 1024]-----+
|          .+=B0*|
|         + oB*+|
|        . oo+o |
|       . . . E |
|      S  .  o |
|              |
|              |
+-----+
#
```

In `ssh-keygen`, the parameter `-t` refers to the type of SSH key (RSA in Example C-1 on page 935) and `-b` is the size of SSH key in bits (in Example C-1 on page 935, 1024 bits was used).

You also must specify the path and name for the SSH keys. The name that you provide is the name of the private key. The public key has the same name, but with extension `.pub`. In Example C-1 on page 935, the path is `/.ssh/`, the name of the private key is `sshkey`, and the name of the public key is `sshkey.pub`.

Note: Using a passphrase for the SSH key is optional. If a passphrase is used, security is increased, but more steps are required to log in with the SSH key because the user must enter the passphrase.

Uploading the SSH public key to the IBM Storage System

To upload the new SSH public key to IBM Spectrum Virtualize by using the GUI, see “Uploading the SSH public key to the IBM Storage System” on page 929.

To upload the public key by using the CLI, complete the following steps:

1. On the SSH client (for example, AIX or Linux host), run `scp` to copy the public key to the IBM Storage System. The basic syntax for the command is:

```
scp <file> <user>@<hostname_or_IP_address>:<path>
```

The directory `/tmp` in the IBM Spectrum Virtualize active configuration node can be used to store the public key temporarily. Example C-2 shows the command to copy the newly generated public key to the IBM Spectrum Virtualize system.

Example: C-2 SSH public key copy to an IBM Storage System

```
# scp /.ssh/sshkey.pub admin@192.168.100.1:/tmp/  
Password:*****  
sshkey.pub  
100% 241    0.2KB/s   00:00  
#
```

2. Log in to the storage system by using SSH and run the `chuser` command (as shown in Example C-3) to associate the public SSH key with a user.

Example: C-3 Importing the SSH public key to a user

```
IBM_Storage System:ITS0:admin>chuser -keyfile /tmp/sshkey.pub admin  
IBM_Storage System:ITS0:admin>lsuser admin  
id 4  
name admin  
password yes  
ssh_key yes  
remote no  
usergrp_id 1  
usergrp_name Administrator  
IBM_Storage System:ITS0:admin>
```

When running the `lsuser` command as shown in Example C-3, it is indicated that a user has a configured SSH key in the field `ssh_key`.

Connecting to an IBM Spectrum Virtualize system

Now that the SSH key is uploaded to the IBM Spectrum Virtualize system and assigned to a user account, you can connect to the device by running the `ssh` command with the following options:

```
ssh -i <SSH_private_key> <user>@<IP_address_or_hostname>
```

Example C-4 shows the SSH command that is running from an AIX server and connecting to the storage system with an SSH private key and no password prompt.

Example: C-4 Connecting to IBM Storage System with an SSH private key

```
# ssh -i /.ssh/sshkey admin@192.168.100.1  
IBM_Storage System:ITS0:admin>
```

Related publications

The publications that are listed in this section are considered suitable for a more detailed description of the topics that are covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide more information about the topics in this document. Some publications that are referenced in this list might be available in softcopy only.

- ▶ *IBM FlashSystem 5000 Family Products*, SG24-8449
- ▶ *IBM FlashSystem 9100 Architecture, Performance, and Implementation*, SG24-8425
- ▶ *IBM FlashSystem 9200 and 9100 Best Practices and Performance Guidelines*, SG24-8448
- ▶ *IBM System Storage SAN Volume Controller, IBM Storwize V7000, and IBM FlashSystem 7200 Best Practices and Performance Guidelines*, SG24-7521
- ▶ *Implementing the IBM Storwize V5000 Gen2 (including the Storwize V5010, V5020, and V5030) with IBM Spectrum Virtualize V8.2.1*, SG24-8162
- ▶ *Implementing the IBM Storwize V7000 with IBM Spectrum Virtualize V8.2.1*, SG24-7938
- ▶ *Implementing the IBM System Storage SAN Volume Controller with IBM Spectrum Virtualize V8.2.1*, SG24-7933
- ▶ *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430

You can search for, view, download, or order these documents and other Redbooks, Redpapers, web docs, drafts, and additional materials, at the following website:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Abbreviations and acronyms

AD	Active Directory	ESCC	EMEA Storage Competence Center
AES	Advanced Encryption Standard	ESM	Enclosure Services Manager
AI	artificial intelligence	ESX	Elastic Sky X
ALUA	Asymmetric Logical Unit Access	ESXi	Elastic Sky X Integrated
ANA	Asymmetric Namespace Access	ETS	Enhanced Transmission Selection
ANSI	American National Standards Institute	EUI	extended-unique identifier
API	application programming interface	FC-NVMe	Non-Volatile Memory Express over Fibre Channel
ASCII	American Standard Code for Information Interchange	FC	Fibre Channel
ATM	asynchronous transfer mode	FCIP	Fibre Channel over IP
BBU	battery backup unit	FCM	IBM FlashCore Module
CA	certificate authority	FCoE	Fibre Channel over Ethernet
CBC	Cipher Block Chaining	FCP	Fibre Channel Protocol
CHAP	Challenge Handshake Authentication Protocol	FICON	Fibre Channel connection
CIMOM	Common Information Model Object Manager	FQDN	fully qualified domain name
CLI	command-line interface	FRU	field-replaceable unit
CoW	copy-on-write	GbE	Gb Ethernet
CRU	customer-replaceable unit	GBIC	gigabit interface converter
CSP	cloud service provider	GBps	gigabytes per second
CSU	Customer Setup Unit	Gbps	gigabits per second
CSV	comma-separated value	GM	Global Mirror
DCBx	Data Center Bridging Exchange	GMCV	Global Mirror with Change Volumes
DES	Data Encryption Standard	GPFS	General Parallel File System
DIF	Data Integrity Field	GUI	graphical user interface
DIMM	Dual Inline Memory Module	HA	high availability or highly available
DMP	directed maintenance procedure	HBA	host bus adapter
DNS	Domain Name System or domain name server	HDD	hard disk drive
DR	disaster recovery	HIC	host interface card
DRAID	distributed redundant array of independent disks	HSM	hardware security module
DRET	Data Reduction Estimation Tool	laaS	infrastructure as a service
DRP	Data Reduction Pool	IBM	International Business Machines Corporation
DSFA	Data Storage Feature Activation	IDA	Information Dispersal Algorithm
DWPD	Drive Write Per Day	IOPS	input/output operations per second
ECS	Enterprise Class Support	IQN	iSCSI Qualified Name
ECuRep	Enhanced Customer Data Repository	iSCSI	internet Small Computer Systems Interface
ENT	enterprise	iSER	iSCSI Extensions for RDMA
		ISL	inter-switch link
		iSNS	internet Storage Name Service
		ITIL	IT Infrastructure Library

iWARP	internet Wide Area RDMA Protocol	OBAC	object-based access control
JBOD	just a bunch of disks	ODX	Microsoft Offloaded Data Transfer
KB	kilobytes	OS	operating system
Kbps	kilobits per second	OUI	organizationally unique identifier
KiB	kibibyte	PaaS	platform as a service
LAN	local area network	PCIe	Peripheral Component Interconnect Express
LBA	logical block address	PDU	power distribution unit
LDAP	Lightweight Directory Access Protocol	PEM	Privacy Enhanced Mail
LED	light-emitting diode	PFC	priority flow control
LFF	large form factor	PiT	point in time or point-in-time
LIC	License Internal Code	PMP	Project Management Professional
LRU	least recently used	PMR	Problem Management Record
LSA	Log Structured Array	POST	power-on self-test
LU	logical unit	PPK	PuTTY private key
LUN	logical unit number	PSU	power supply unit
LVM	logical volume manager	QoS	quality of service
MAC	Media Access Control	RACE	Random Access Compression Engine
Mb	megabits	RAID	redundant array of independent disks
MBps	megabytes per second	RAIDs	Redundant Arrays of Independent Disks
Mbps	megabits per second	RAS	reliability, availability, and serviceability
MDisk	managed disk	RC	Remote Copy
MES	miscellaneous equipment specification	RDMA	Remote Direct Memory Access
MGM	Metro Mirror Global	REST	Representational State Transfer
MIB	Management Information Base	RFC	Request for Comments
MiB	mebibytes	RHEL	Red Hat Enterprise Linux
MM	Metro Mirror	RoCE	RDMA over Converged Ethernet
MSDSM	Microsoft Device Specific Module	RoW	redirect-on-write
MT	machine type	RPO	recovery point objective
MTBF	mean time between failures	RPORT	remote port
MTM	machine type and model	RtC	real-time compression
MTU	maximum transmission unit	RTO	recovery time objective
N_Port	node port	RTT	round-trip time
NAA	Network Address Authority	SaaS	software as a service
NDVM	Non-disruptive Volume Move	SAN	storage area network
NGUID	Namespace Globally Unique Identifier	SAS	serial-attached SCSI
NIC	network interface controller	SAT	Service Assistant Tool
NL	nearline	SATA	Serial Advanced Technology Attachment
NPIV	N_Port ID Virtualization	SBB	Storage Bridge Bay
NQN	NVMe Qualified Name	SCM	storage-class memory
NTP	Network Time Protocol	SCP	Secure Copy Protocol
NVMe-oF	NVMe over Fabrics		
NVMe	Non-Volatile Memory Express		

SCSI	Small Computer System Interface	VLAN	virtual local area network
SCU	storage capacity unit	VM	virtual machine
SDD	Subsystem Device Driver	VMFS	Virtual Machine File System
SDDDSM	Subsystem Device Driver Device Specific Module	VPD	vital product data
SDN	software-defined network	VPN	virtual private network
SEM	Secondary Expander Module	VSAN	virtual storage area network
SFF	small form factor	VSR	Variable Stripe RAID
SFP	small form factor pluggable	VVOL	VMware vSphere Virtual Volume
SLA	service-level agreement	WWID	worldwide ID
SLP	Service Location Protocol	WWN	worldwide name
SME	subject matter expert	WWNN	worldwide node name
SMTP	Simple Mail Transfer Protocol	WWPN	worldwide port name
SNMP	Simple Network Management Protocol		
SPOF	single point of failure		
SPOFs	single points of failure		
SRM	System Resource Manager		
SSD	solid-state drive		
SSH	Secure Shell		
SSIC	System Storage Interoperation Center		
SSL	Secure Sockets Layer		
SSR	IBM System Services Representative		
STAT	Storage Tier Advisor Tool		
SVC	SAN Volume Controller		
T0	time-zero		
TB	terabytes		
TBps	terabytes per second		
TCO	total cost of ownership		
TCT	Transparent Cloud Tiering		
TLC	Triple Level Cell		
TPGS	Target Port Group Support		
TRAIID	traditional RAID		
UDID	unit device identifier		
UID	unique identifier		
UPN	User Principal Name		
URL	Uniform Resource Locator		
VAAI	vStorage APIs for Array Integration		
VASA	vSphere APIs for Storage Awareness		
VC	virtual connection		
VDisk	virtual disk		
VIOS	Virtual I/O Server		



Implementing IBM FlashSystem with IBM Spectrum Virtualize V8.4

SG24-8492-00

ISBN 0738459364



(1.5" spine)
1.5" <-> 1.998"
789 <-> 1051 pages



SG24-8492-00

ISBN 0738459364

Printed in U.S.A.

Get connected

