Redbooks

ibm.com/redbooks

Implementing the IBM Storwize V7000 with IBM Spectrum Virtualize V8.2.1

Jon Tate

Jack Armstrong

Tiago Bastos

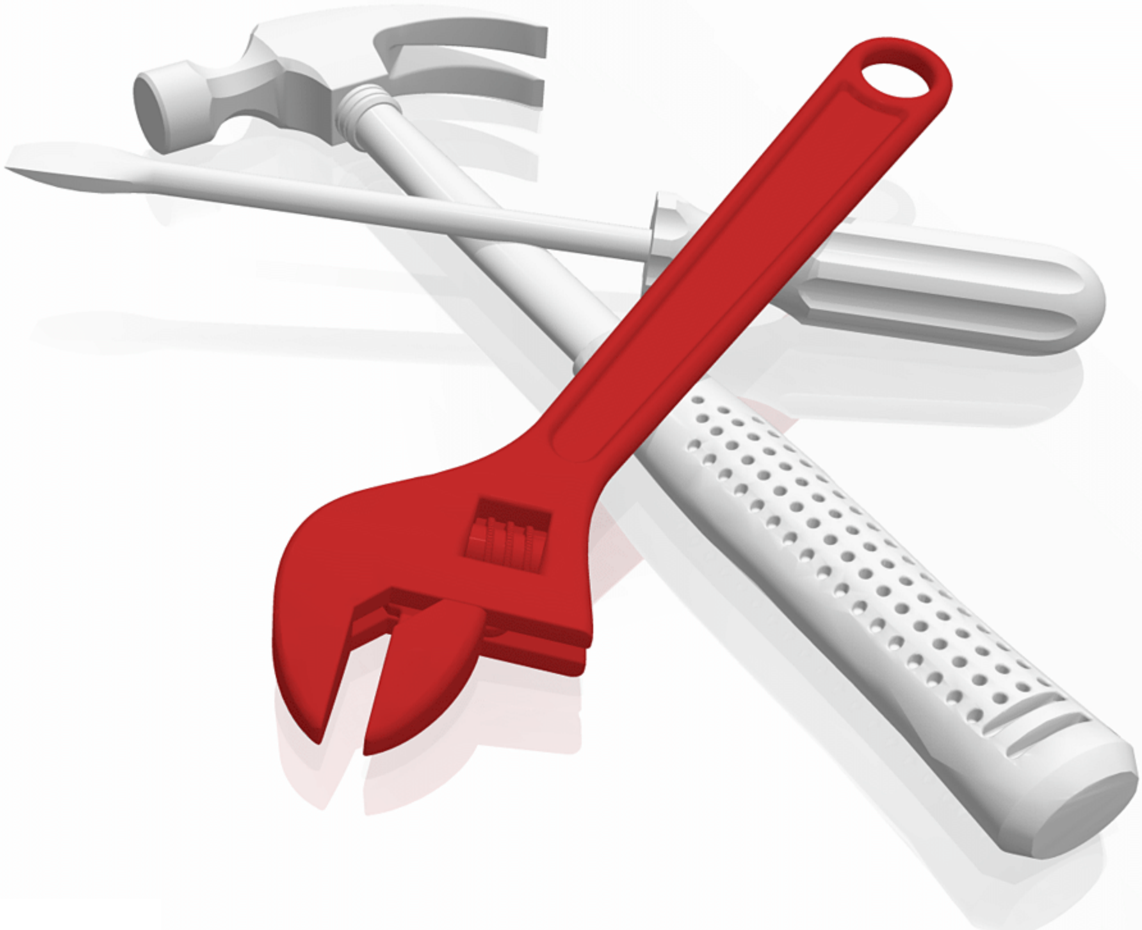
Pawel Brodacki

Frank Enders

Sergey Kubin

Danilo Miyasiro

Rodrigo Suzuki



IBM

Redbooks

International Technical Support Organization

Implementing the IBM Storwize V7000 with IBM Spectrum Virtualize V8.2.1

June 2019

SG24-7938-07

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

Eighth Edition (June 2019)

This edition applies to IBM Spectrum Virtualize V8.2.1 and the associated hardware and software detailed within. Note that the screen captures included within this book might differ from the generally available (GA) version, because parts of this book were written with pre-GA code.

Contents

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 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®
Concert®
DB2®
Db2®
DS4000®
DS6000™
DS8000®
Easy Tier®
FlashCopy®
Global Technology Services®
HyperSwap®
IBM®
IBM Cloud™
IBM FlashSystem®
IBM Spectrum™
IBM Spectrum Accelerate™
IBM Spectrum Control™
IBM Spectrum Protect™
IBM Spectrum Scale™
IBM Spectrum Storage™
IBM Spectrum Virtualize™
Informix®
Insight®
PowerHA®
Rational®
Rational Team Concert™
Real-time Compression™
Redbooks®
Redbooks (logo) ®
Storwize®
System Storage®
Tivoli®
XIV®

The following terms are trademarks of other companies:

SoftLayer, are trademarks or registered trademarks of SoftLayer, Inc., an IBM Company.

ITIL is a Registered Trade Mark of AXELOS Limited.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

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.

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

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 Storwize® V7000 solution powered by IBM Spectrum™ Virtualize. 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 IBM Storwize V7000 solution incorporates some of the top IBM technologies that are typically found only in enterprise-class storage systems, which raises 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 the IBM Storwize V7000 and IBM Spectrum Virtualize™ V8.2.1 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 of the Storwize V7000, how to implement it, and how to take advantage of its industry-leading functions and features.

Authors

This book was produced by a team of specialists from around the world working at the IBM Redbooks, Poughkeepsie Center.

Jon Tate is a Project Manager for IBM System Storage® SAN Solutions at the ITSO, San Jose Center. Before joining the ITSO in 1999, he worked in the IBM Technical Support Center, providing Level 2/3 support for IBM mainframe storage products. Jon has 33 years of experience in storage software and management, services, and support. He is an IBM Certified IT Specialist, an IBM SAN Certified Specialist, and is Project Management Professional (PMP) certified. He is also the UK Chairman of the Storage Networking Industry Association (SNIA).

Jack Armstrong is a Storage Support Specialist for IBM Systems Group based in Hursley, UK. He joined IBM as part of the Apprenticeship Scheme in 2012 and has 6 years of experience working with Storage, providing support to thousands of customers across Europe and beyond. He also works with IBM Enhanced Technical Support Services to help clients to expand and improve their storage environments.

Tiago Bastos is a SAN and Storage Disk specialist for IBM Brazil. He has over 17 years experience in the IT arena, is an IBM Certified Master IT Specialist, and is certified on the Storwize portfolio. He works on Storage as a Service (SaaS) implementation projects. His areas of expertise include planning, configuring and troubleshooting IBM DS8000®, Storwize V5000 and V7000, FlashSystem 900, IBM SAN Volume Controller, and XIV®.

Pawel Brodacki is an Infrastructure Architect with 20 years of experience in IT, working for IBM in Poland since 2003. His main focus for the last 5 years has been on virtual infrastructure architecture from storage to servers to software defined networks. Before changing profession to Architecture, he was an IBM Certified IT Specialist working on various infrastructure, virtualization, and disaster recovery projects. His experience includes SAN, storage, highly available systems, disaster recovery solutions, IBM System x and Power servers, and several types of operating systems (Linux, IBM AIX®, and Microsoft Windows). Pawel has 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.

Frank Enders has worked for the last 12 years for EMEA Storwize Level 2 support in Germany, and his duties include pre- and post-sales support. He has worked for IBM Germany for more than 22 years and started as a technician in disk production for IBM Mainz and changed to magnetic head production four years later. In 2001, he changed his role and continued working for IBM within ESCC Mainz as a team member of the Installation Readiness team for products such as the IBM DS8000, IBM DS6000™, and the IBM System Storage SAN Volume Controller. During that time, he studied for four years to earn a degree in Electrical Engineering.

Sergey Kubin is a subject-matter expert for IBM Storage and SAN support in IBM Russia. He has worked with IBM Technology Support Services for 12 years, providing L1 and L2 support on Spectrum Virtualize, SAN, DS4000/DS5000, and N Series storage for IBM customers in Russia, CEE, and EMEA. He is IBM Certified Specialist for Storwize Family Technical Solutions.

Danilo Miyasiro is a SAN and Disk Storage Specialist for IBM Global Technology Services® in Brazil. He graduated in Computer Engineering at State University of Campinas, Brazil, and has more than 10 years of experience in IT. As a storage subject matter expert for several international customers, he works on designing, implementing, and supporting storage solutions. He is an IBM Certified Specialist for DS8000 and the Storwize family, and also holds certifications from the ITIL Foundation and other storage products.

Rodrigo Suzuki is a SAN Storage specialist at IBM Brazil Global Technology Services in Hortolandia. Currently, Rodrigo is a subject matter expert account focal, and has been working on projects and support for international customers.

He has 24 years of IT Industry experience with the last 9 years in the SAN Storage Disk area. He also 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, NetApp NCDA, IBM Storwize V7000 Technical Solutions V2, and ITIL certified.

Thanks to the authors of the previous edition of this book:

Erwan Auffret, Pawel Brodacki, Libor Miklas, Glen Routley, James Whitaker

Thanks to the following people for their contributions to this project:

Christopher Bulmer
Debbie Butts
Carlos Fuente
Evelyn Perez
Matt Smith
IBM Hursley, UK

James Whitaker
Imran Imtiaz
Adam Lyon-Jones
IBM Manchester, UK

Jordan Fincher
Karen Brown
Mary Connell
Navin Manohar
Terry Niemeyer
IBM US

Special thanks to the Broadcom Inc. staff in San Jose, California for their support of this residency in terms of equipment and support in many areas:

Sangam Racherla
Brian Steffler
Marcus Thordal
Broadcom Inc.

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 ITSO 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, International Technical Support Organization
  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

Summary of changes

This section describes the technical changes made in this edition of the book and in previous editions. This edition might also include minor corrections and editorial changes that are not identified.

Summary of Changes
for SG24-7938-07
for Implementing the IBM Storwize V7000 with IBM Spectrum Virtualize V8.2.1
as created or updated on November 7, 2019.

June 2019, Eighth Edition

This revision includes the following new and changed information.

New information

- Add new look GUI
- Hot Spare node
- RAS line items

Changed information

- Added new GUI windows throughout

1

Introduction to storage virtualization

This chapter defines the concept of *storage virtualization* and provides an overview of its application in addressing the challenges of modern storage environment.

This chapter includes the following topics:

1.1 Storage virtualization terminology

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.

IBM describes 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 desirable characteristics.
- It 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 function with services and can be nested or applied to multiple layers of a system.

Storage virtualization is also defined in the Storage Networking Industry Association's (SNIA) shared storage model version 2, as shown in Figure 1-1.

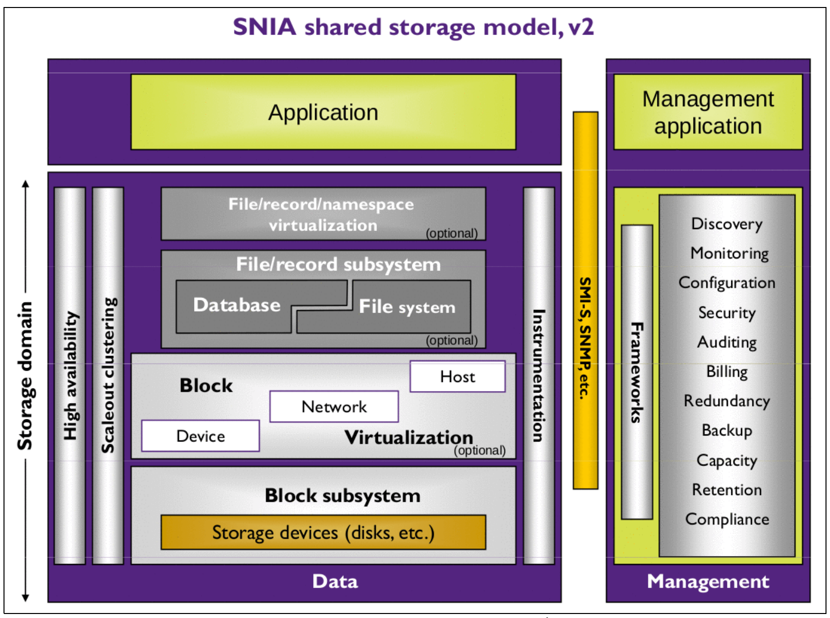


Figure 1-1 SNIA shared storage model, version 2[1]

The model consists of the following layers:

- Application: The user of the storage domain.
- Storage domain:
  - File/record/namespace virtualization and File/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).

[1] Source: Storage Networking Industry Association.

File/record/namespace virtualization and *File/record subsystem* layers translate records or files that are required by applications to vectors of blocks, which are the language of the *Block Virtualization* layer. *Block Virtualization* layer, in turn, 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 IBM SAN Volume Controller and the IBM Storwize family. The IBM SAN Volume Controller is implemented as a clustered appliance in the storage network layer. The IBM Storwize family is deployed as modular storage that can virtualize its internally and externally attached storage.

IBM Spectrum Virtualize uses Small Computer System Interface (SCSI) protocol to communicate with its clients and presents storage space in form of SCSI logical units (LUs), which are identified by SCSI logical unit numbers (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 of a host must translate these abstractions to the language of storage; that is, vectors of storage blocks that are identified by logical block addresses 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 in turn are made out of managed disks (MDisks); that is, 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.

The hierarchy of objects, from a file system block down to a physical block on a physical drive, is shown in Figure 1-2.

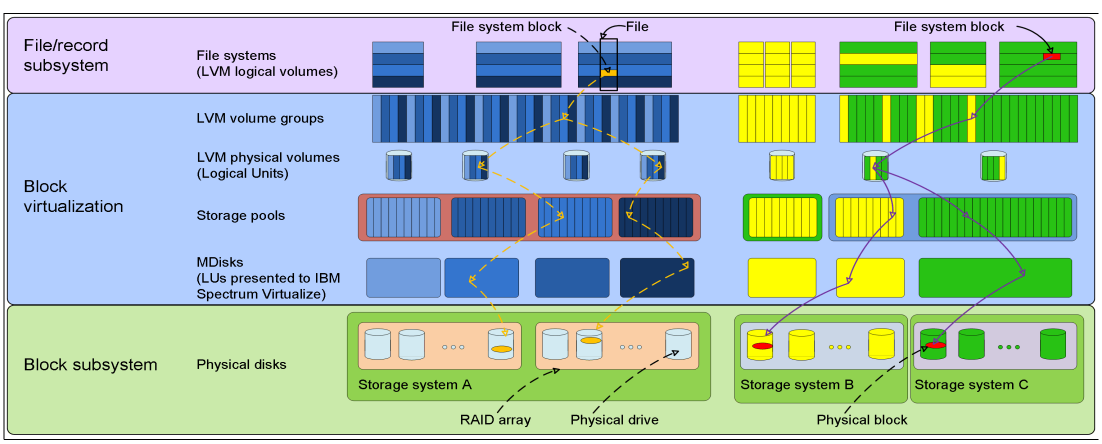


Figure 1-2 Block-level virtualization overview

Storage virtualization allows us to manage the mapping between logical blocks within an LU presented to a host, and blocks on physical drives. This mapping can be as simple, or as complicated, as required by a given use case. 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, possibly in different geographical locations.

Importantly, the mapping can be dynamic: with EasyTier functionality, IBM Spectrum Virtualize can automatically change underlying storage, to which given 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, to creating a point-in-time copy with IBM FlashCopy®, to migrating data across physical storage systems. Importantly, all the functionality that is presented to the storage consumers is independent from 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.

Storage virtualization can, and is, being implemented on many layers. Figure 1-2 on page 3 shows an example where a file system block gets mirrored by the host's operating system (left side of the figure) by using features of the logical volume manager (LVM), or by IBM Spectrum Virtualize system at the storage pool level (as shown on the right side of Figure 1-2 on page 3). Although the result is similar (the data block is written to two different arrays), the effort that is required for per-host configuration is disproportionately larger than for a centralized solution with organization-wide storage virtualization that is done on a dedicated system and managed from a single GUI.

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, in a way that is transparent for the storage consumers and without over-straining the storage infrastructure. The migration can be done within a specific storage system (from one set of disks to another), 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 the IBM Spectrum Virtualize removes dependencies on capabilities and intercompatibility 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 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 way of 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, such as thin provisioning, Data Reduction Pools (DRP), deduplication, and IBM Real-time Compression™ (RtC). Today, open systems typically use less than 50% of the provisioned storage capacity. IBM Spectrum Virtualize, can enable significant savings, increase effective capacity of storage systems up to five times, and decrease the floor space, power, and cooling required by the storage system.

IBM Storwize V7000 is a scalable solution running on a highly available platform that can use diverse back-end storage systems to provide all the benefits to various attached hosts.

1.2 Benefits of using IBM Spectrum Virtualize

The storage virtualization functions of IBM Spectrum Virtualize are a powerful tool in the hands of storage administrators. However, for an organization to fully realize benefits of storage virtualization, its implementation must be the result of a process that begins with identifying the organization's goals. For a storage virtualization project to be a success, the organization must identify what it wants to achieve before it starts to think how to implement the solution.

Today, organizations are searching for affordable and efficient ways to store, use, protect, and manage their data. Additionally, a storage environment is required to have an easy to manage interface and be sufficiently flexible to support wide range of applications, servers, and mobility requirements. Although business demands change quickly, some recurring client concerns drive adoption of storage virtualization, including the following examples:

- Growing data center costs
- Inability of IT organizations to respond quickly to business demands
- Poor asset usage
- Poor availability and resultant unsatisfactory (for the clients) or challenging (for the providers) service levels
- Lack of skilled staff for storage administration

The importance of addressing the complexity of managing storage networks is clearly visible in the results of industry analyses of the total cost of ownership (TCO) of storage networks. Typically, storage costs are only about 20% of the TCO. Most of the remaining costs relate to managing storage systems.

In a non-virtualized storage environment, every system is an "island" that must be managed separately. In large SAN environments, the sheer number of separate and different management interfaces and lack of unified view of the whole environment jeopardizes an organization's ability to manage their storage as a single entity and to maintain the current view of the system state.

IBM Storwize V7000, running IBM Spectrum Virtualize software, reduces the number of separate environments that must be managed down to a single system. After the initial configuration of the back-end storage subsystems, all of the day-to-day storage management operations are performed by way of a single graphical user interface. At the same time, administrators gain access to the rich functionality set provided by IBM Spectrum Virtualize, whether particular features are natively available on the virtualized storage systems.

With IBM Spectrum Virtualize, free space does not need to be maintained and managed in each storage subsystem, which further increases capacity usage. Today, it is typical that open systems run at less than 50% of the usable capacity that is provided by the RAID disk subsystems.

The use of the installed raw capacity in the disk subsystems shows usage numbers of less than 35%, depending on the RAID level that is used. A block-level virtualization solution, such as IBM Spectrum Virtualize, can allow a significant increase to approximately 75 - 80%.

1.3 Latest changes and enhancements

IBM Spectrum Virtualize V8.2.1 is another step in the product line development that brings new features and enhancements. The major software changes in subsequent code releases are described in this section.

V8.2.1 of the Spectrum Virtualize code brought the following changes:

- Full IP-based quorum: Support for administrators looking to consolidate their infrastructure over Ethernet.
- Support for iSCSI extensions over RDMA (iSER) host attach enhancements.
- Support for host login information for Ethernet-attached hosts.
- 64 K host mappings: Increased from the previous limitation of 20 K.
- IBM Spectrum Insights: Provides a call home protocol that uses IBM IP to deliver a more robust path and higher bandwidth and higher frequency data transmission, with end-to-end confirmation of receipt.
- Improvements to call home with email notifications.
- Added support for SafeNet KeySecure encryption key server.
- Single copy VDisk expand with format: Enables administrators to expand a VDisk without migrating the data off and back on.
- Support for RDMA-based connections between nodes. This feature requires 25 GbE adapters to be installed on the nodes. Expands the host connectivity options for those storage arrays not initially covered by 8.2.
- NVMe over Fibre Channel support on 16 Gb Fibre Channel adapters: Extends the simplicity, efficiency, and end-to-end NVMe model where NVMe commands and structures are transferred end to end, requiring no translations

1.4 Summary

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

IBM Spectrum Virtualize running on IBM Storwize V7000 is a mature, ninth-generation virtualization solution that uses open standards and complies with the SNIA storage model. IBM Storwize V7000 is an appliance-based, mid-range storage, and in-band block virtualization engine that moves 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.

2

System overview

This chapter provides an overview of IBM Spectrum Virtualize software and IBM Storwize V7000 architecture and components. The chapter shows the software elements that build the IBM Storwize V7000 platform and provides an overview of the useful management and support tools that helps to maintain and operate the IBM Storwize V7000.

This chapter includes the following topics:

2.1 IBM Spectrum Virtualize

IBM Spectrum Virtualize is a software-enabled storage virtualization engine that provides a single point of control for storage resources within the data centers. IBM Spectrum Virtualize is a core software engine of well-established and industry-proven IBM storage virtualization solutions. These solutions include IBM SAN Volume Controller and all versions of IBM Storwize family of products (IBM Storwize V3700, IBM Storwize V5000, IBM Storwize V7000, and IBM FlashSystem® V9000).

Naming: With the introduction of the IBM Spectrum Storage™ family, the software that runs on IBM SAN Volume Controller and IBM Storwize family products is called *IBM Spectrum Virtualize*. The name of the underlying hardware platform remains intact.

The objectives of IBM Spectrum Virtualize are to manage storage resources in your IT infrastructure, and to ensure that they are used to the advantage of your business. These processes take place quickly, efficiently, and in real time, while avoiding increases in administrative costs.

Although IBM Spectrum Virtualize is a core software engine of the entire family of IBM Storwize products (see Figure 2-1), the information in this book is intentionally related to the deployment considerations of IBM Storwize V7000 Gen2 and Gen2+.

Throughout this book, the term *IBM Storwize V7000* is used to refer to both models of the IBM Storwize V7000 when the text applies similarly to both.

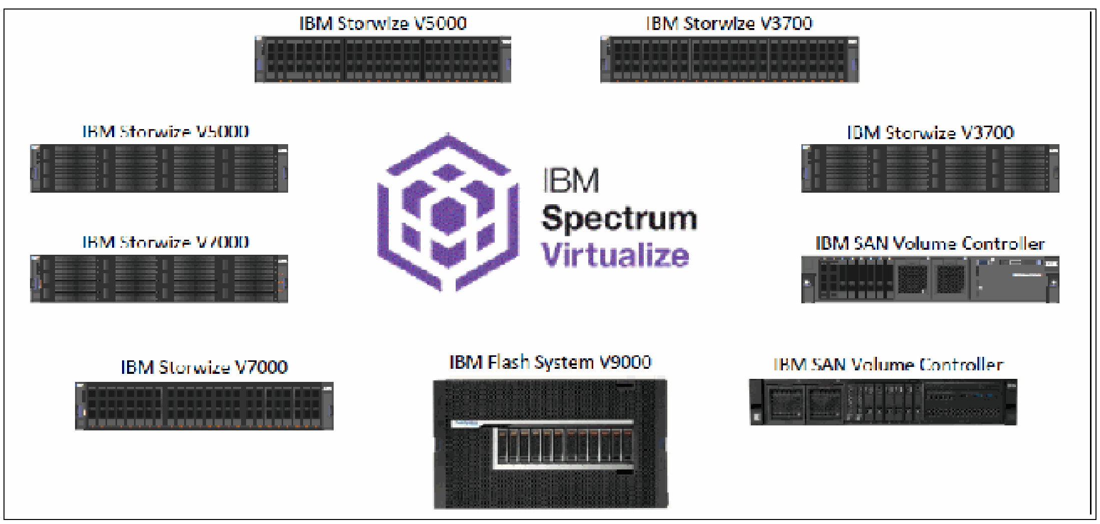


Figure 2-1 IBM Spectrum Virtualize software

2.2 Storage virtualization

Storage virtualization, is one of the foundations of building a flexible and reliable infrastructure solution that enables companies to better align their technologies needs. Storage virtualization enables an organization to achieve affordability and manageability by implementing storage pools across several physically separate disk systems (which might be from different vendors).

Storage can then be deployed from these pools, and migrated transparently between pools without interruption to the attached host systems and their applications. Storage virtualization provides a single set of tools for advanced functions, such as instant copy and remote mirroring solutions, which enables faster and seamless deployment of storage regardless of the underlying hardware.

Because the storage virtualization represented by IBM Spectrum Virtualize is a software-enabled function, it offers more features that are typically not available on a regular pure storage subsystem, including (but not limited to) the following features:

- Data compression
- Software and hardware encryption
- IBM Easy Tier® for workload balancing
- Thin provisioning
- Mirroring and copy services
- Interface to Cloud Service Providers

Figure 2-2 shows these features.

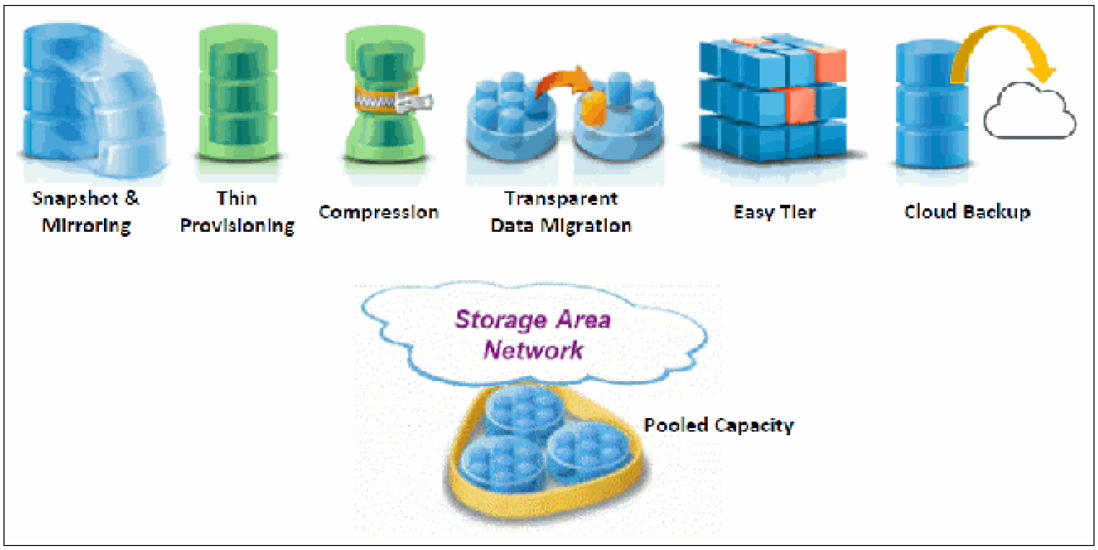


Figure 2-2 IBM Spectrum Storage virtualization

The following reasons are the five top reasons to choose software-defined storage virtualization infrastructure and benefit from IBM Spectrum Virtualize:

- Lower license cost. Avoid purchasing licenses from multiple storage vendors for advanced features (replication, tiering, snapshots, compression, and so on) of each external storage subsystem. Manage them centrally from IBM Spectrum Virtualize.
- Feed more data on physical disks. External SAN disk arrays have physical boundaries. Although one subsystem might run out of space, another has free capacity. Virtualization removes these boundaries.
- Choose lower-cost disk arrays. Less-demanding applications can easily run on cheaper disks with lower performance. IBM Spectrum Virtualize automatically and transparently moves data up and down between low-performance and high-performance disk arrays (tiering).
- Cloud FlashCopy. IBM Spectrum Virtualize has an interface to the external cloud service providers. IBM Spectrum Cloud FlashCopy supports full and incremental backup and restore from cloud snapshots.

- Quick adoption of new technologies. IBM Spectrum Virtualize seamlessly integrates invention in storage technologies, such as new array types and new disk vendors.
- Extended high availability (HA). Cross-site virtualization, workload migration, and copy services enhance options for deployment of HA scenarios or disaster recovery (DR) solutions (IBM SAN Volume Controller Enhanced Stretched Cluster and IBM HyperSwap®).

2.3 IBM Storwize V7000 overview

IBM Storwize V7000 solution incorporates IBM Spectrum Virtualize software and provides a modular storage system that includes the capability to virtualize its internal and external SAN-attached storage. IBM Storwize V7000 solution is built on IBM Spectrum Virtualize.

IBM Storwize V7000 system provides several configuration options that are aimed at simplifying the implementation process. These configuration options conform to different implementation scenarios regarding the size of your data center, and SAN and local area network (LAN) topology. IBM Storwize V7000 system is a clustered, scalable, midrange storage system that is easy to deploy and use.

Figure 2-3 shows a high-level overview of IBM Storwize V7000.

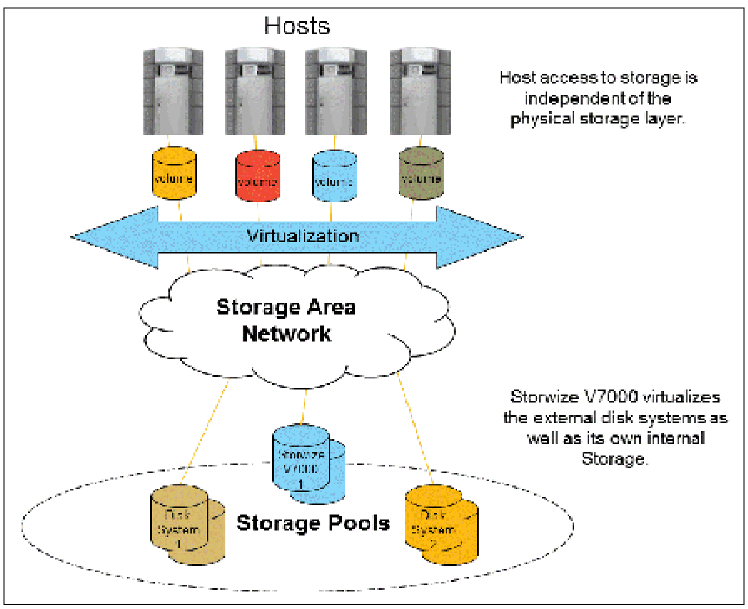


Figure 2-3 IBM Storwize V7000 overview

The IBM Spectrum Virtualize software that runs on IBM Storwize V7000 provides a GUI that enables storage to be deployed quickly and efficiently. The GUI is provisioned by IBM Spectrum Virtualize code and there is no need for a separate console.

The management GUI contains a series of preestablished configuration options that are called *presets*, and that use common settings to quickly configure objects on the system. Presets are available for creating volumes and IBM FlashCopy mappings, and for setting up a RAID configuration, including traditional RAIDs and the new feature of distributed RAID.

An IBM Storwize V7000 solution provides a choice of up to 760 disk drives per system or 1024 disk drives per clustered system (by using dense drawers). The solution uses SAS cables and connectors to attach to the optional expansion enclosures.

The IBM Storwize V7000 system supports a range of external disk systems similar to what IBM SAN Volume Controller supports today. A view of an IBM Storwize V7000 control enclosure is shown in Figure 2-4.

Figure 2-4 Top-front view of a Storwize V7000 control enclosure

The IBM Storwize V7000 solution consists of 1 - 4 control enclosures and optionally, up to 36 expansion enclosures. It supports the intermixing of the different expansion enclosures. Each enclosure contains two canisters. Control enclosures contain two node canisters, and expansion enclosures contain two expansion canisters.

2.3.1 IBM Storwize V7000 models

The IBM Storwize V7000 consists of enclosures and drives. An enclosure contains two canisters that are seen as part of the enclosure, although they can be replaced independently.

More information: For the most up-to-date information about features, benefits, and specifications of IBM Storwize V7000 models, see this web page.

The information in this IBM Redbooks publication is valid at the time of this writing and covers IBM Spectrum Virtualize V8.2. As IBM Storwize V7000 matures, expect to see new features and enhanced specifications.

The IBM Storwize V7000 models are listed in Table 2-1.

Table 2-1 IBM Storwize V7000 models

Model	Cache	Fibre Channel (FC) / iSCSI / SAS ports	Drive slots	Power supply
2076-AF1 (with two node canisters Gen2+)	64, 128, or 256 gigabytes (GB)	16 x 16 Gb / 6 x 1 Gb + 8x 10 Gb / 4 x 12 Gb	24 x 2.5-inch (All Flash)	Integrated dual power supplies with battery

Model	Cache	Fibre Channel (FC) / iSCSI / SAS ports	Drive slots	Power supply
2076-624 (with two node canisters Gen2+)	64, 128, or 256 gigabytes (GB)	16 x 16 Gb / 6 x 1 Gb + 8x 10 Gb / 4 x 12 Gb	24 x 2.5-inch	Integrated dual power supplies with battery
2076-524 (with two node canisters Gen2)	32 or 64 gigabytes (GB)	4 x 16 Gb / 4 x 1 Gb + 4 x 10 Gb / 4 x 12 Gb	24 x 2.5-inch	Integrated dual power supplies with battery
2076-212 (with two expansion canisters)	Not applicable (N/A)	-- / -- / 4 x 12 Gb	12 x 3.5-inch	Integrated dual power supplies
2076-224 (with two expansion canisters)	N/A	-- / -- / 4 x 12 Gb	24 x 2.5-inch	Integrated dual power supplies
2076-12F (with two expansion canisters Gen2)	N/A	-- / -- / 4 x 12 Gb	12 x 3.5-inch	Integrated dual power supplies (attaches to 2076-524 and 2076-624 only)
2076-24F (with two expansion canisters Gen2)	N/A	-- / -- / 4 x 12 Gb	24 x 2.5-inch	Integrated dual power supplies (attaches to 2076-524 and 2076-624 only)
2076-92F (with two expansion canisters Gen2)	N/A	-- / -- / 4 x 12 Gb	92 x 3.5-inch	Integrated dual power supplies (attaches to 2076-524 and 2076-624 only)

Note: The first generation of control enclosures (2076 - models 112, 124, 312, and 324) were withdrawn from marketing. However, expansion enclosures 2076-212 and 2076-224 can still be ordered (see Table 2-1) because they attach to those control enclosures only. Intermixing control enclosures with expansion enclosures of different generations is not a supported combination, and is refused by IBM Spectrum Virtualize software.

The first generation of IBM Storwize V7000 hardware is not supported by IBM Spectrum Virtualize V8.1. Any attempt to upgrade to V8.1 is rejected by the software. The last supported version for first-generation Storwize V7000 is V7.8.

2.3.2 IBM Storage Utility offerings

The IBM 2076 Model U7A is the IBM Storwize V7000 with a three-year warranty to be used in the Storage Utility Offering space. These models are physically and functionally identical to the Storwize V7000 model 624, except for target configurations and variable capacity billing. The variable capacity billing uses IBM Spectrum Control™ Storage Insights to monitor the system usage, which allows allocated storage usage above a base subscription rate to be billed per TB, per month.

Allocated storage is identified as storage that is allocated to a specific host (and unusable to other hosts), whether data is written. For thin-provisioning, the data that is written is considered used. For thick provisioning, total allocated volume space is considered used.

IBM Storage Utility Offerings include the IBM FlashSystem 900 (9843-UF3), IBM Storwize V5030 (2078-U5A), and Storwize V7000 (2076-U7A) storage utility models that enable variable capacity usage and billing.

These models provide a fixed total capacity, with a base and variable usage subscription of that total capacity. IBM Spectrum Control Storage Insights is used to monitor the system capacity usage. It is used to report on capacity that is used beyond the base subscription capacity, which is referred to as variable usage. The variable capacity usage is billed on a quarterly basis. This feature enables customers to grow or shrink their usage, and pay for configured capacity only.

IBM Storage Utility Offering models are provided for customers who can benefit from a variable capacity system, where billing is based on provisioned space above the base. The base subscription is covered by a three-year lease that entitles the customer to use the base capacity at no additional cost. If storage needs increase beyond the base capacity, usage is billed based on the average daily provisioned capacity per terabyte, per month, on a quarterly basis.

Example

A customer has a Storwize V5030 utility model with 2 TB nearline disks, for a total system capacity of 48 TB. The base subscription for such a system is 16.8 TB. During the months where the average daily usage is below 16.8 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 25 TB, 42.6 TB, and 22.2 TB in three consecutive months, Storage Insights calculates the overage (rounding to the nearest terabyte), as listed in Table 2-2.

Table 2-2 Overage calculation

Average daily	Base	Overage	To be billed
25	16.8	8.2	8
42.6	16.8	25.8	26
22.2	16.8	5.4	5

The capacity that is billed at the end of the quarter is a total of 39 TB-months in this example.

The usage data that is collected by Storage Insights is used by IBM to determine the physical data that is provisioned in the system. This data is compared to the base system capacity subscription, and any provisioned capacity beyond that base subscription is billed per terabyte, per month, on a quarterly basis. The calculated usage is based on the average use over a specific month.

In a highly variable environment, such as managed or cloud service providers, this feature enables the system to be used only as much as is necessary during each month. Usage might increase or decrease, and is billed accordingly. Provisioned capacity is considered capacity that is reserved by the system.

In thick-provisioned environments (available on FlashSystem 900 and Storwize), this capacity is allocated to a host whether it has data that is written. For thin-provisioned environments (available on the Storwize), this data is written and used. This difference occurs because of the different ways in which thick and thin provisioning use disk space.

These systems are available worldwide, but specific client and program differences exist by location. Contact your IBM Business Partner or sales person for more information.

2.3.3 IBM Storwize V7000 functions

The following functions are available with the current release of IBM Spectrum Virtualize:

- Thin provisioning

  Traditional fully allocated volumes allocate real physical disk capacity for an entire volume, even if that capacity is never used. Thin-provisioned volumes allocate real physical disk capacity only when data is written to the logical volume.

- Volume mirroring

  Provides a single volume image to the attached host systems while maintaining pointers to two copies of data in separate storage pools. Copies can be on separate disk storage systems that are being virtualized. If one copy is failing, IBM Storwize V7000 provides continuous data access by redirecting input/output (I/O) to the remaining copy. When the copy becomes available, automatic resynchronization occurs.

- FlashCopy

  Provides a volume level point-in-time copy function for any storage being virtualized by IBM Spectrum Virtualize. This function creates copies for backup, parallel processing, testing, and development, and has the copies available almost immediately.

  IBM Storwize V7000 includes the following IBM FlashCopy functions enabled by IBM Spectrum Virtualize V8.1:

  – Full or incremental copy

    This function copies only the changes from the source or target data since the last FlashCopy operation, and enables completion of point-in-time online backups much more quickly than the use of traditional FlashCopy.

  – Multitarget FlashCopy

    IBM Storwize V7000 supports copying of up to 256 target volumes from a single source volume. Each copy is managed by a unique mapping and in general, each mapping acts independently and is not affected by other mappings sharing the source volume.

  – Cascaded FlashCopy

    This function is used to create copies of copies, and supports full, incremental, or nocopy operations.

  – Reverse FlashCopy

    This function enables data from an earlier point-in-time copy to be restored with minimal disruption to the host.

  – FlashCopy nocopy with thin provisioning

    This function provides a combination of using thin-provisioned volumes and FlashCopy together to help reduce disk space requirements when making copies.

This option has two variations:

- Space-efficient source and target with background copy

  Copies only the allocated space.

- Space-efficient target with no background copy

  Copies only the space that is used for changes between the source and target, and is referred to as a *snapshot*.

This function can be used with multitarget, cascaded, and incremental FlashCopy.

- Consistency groups

  Consistency groups address the issue where application data is on multiple volumes. By placing the FlashCopy relationships into a consistency group, commands can be issued against all of the volumes in the group. This action provides a consistent point-in-time copy of all of the data, even if it is on a physically separate volume.

  FlashCopy mappings can be members of a consistency group, or they can be operated in a stand-alone manner, not as part of a consistency group. FlashCopy commands can be issued to a FlashCopy consistency group, which affects all FlashCopy mappings in the consistency group, or to a single FlashCopy mapping if it is not part of a defined FlashCopy consistency group.

- Metro Mirror

  Provides a *synchronous* remote mirroring function up to approximately 300 km (186.4 miles) between sites. Because the host I/O only completes after the data is cached at both locations, performance requirements might limit the practical distance. Metro Mirror provides fully synchronized copies at both sites with zero data loss after the initial copy is completed. Metro Mirror can operate between multiple IBM Storwize systems and IBM SAN Volume Controllers.

- Global Mirror

  Provides a long-distance *asynchronous* remote mirroring function up to approximately 8,000 km (4,971 miles) between sites. With Global Mirror, the host I/O completes locally and the changed data is sent to the remote site later. This feature maintains a consistent recoverable copy of data at the remote site, which lags behind the local site. Global Mirror can operate between multiple IBM Storwize systems and IBM SAN Volume Controllers.

- External virtualization

  IBM Storwize V7000 provides a data migration function that can be used to import external storage systems into the IBM Storwize V7000 system. The following tasks can be accomplished:

  - Move volumes nondisruptively onto a newly installed storage system.
  - Move volumes to rebalance a changed workload.
  - Migrate data from other back-end storage to IBM Storwize V7000-managed storage.

- Software Encryption

  IBM Storwize V7000 Gen2 and IBM Storwize Gen2+ provide optional encryption of data-at-rest functionality, which protects against the potential exposure of sensitive user data and user metadata that is stored on discarded, lost, or stolen storage devices. Encryption can only be enabled and configured on enclosures that support encryption.

  Starting with IBM Spectrum Virtualize V7.6, IBM Spectrum Virtualize offers software-enabled encryption (hardware encryption was introduced in V7.4), which includes encryption of internal or external storage. Encryption keys now can be stored either on USB flash drives, IBM Security Key Lifecycle Manager server, or both.

- IBM Easy Tier

  Provides a mechanism to seamlessly migrate hot spots to the most appropriate tier within the IBM Storwize V7000 solution. This migration can be to internal drives within the IBM Storwize V7000 or to external storage systems that are virtualized by IBM Storwize V7000. Independently on Easy Tier, IBM Storwize V7000 provides an automatic storage pool balancing that is enabled by default and requires no license.

- IBM Real-time Compression

  IBM Real-time Compression technology is based on the Random Access Compression Engine (RACE). RACE is a part of the software stack of IBM Spectrum Virtualize V6.4 and later. This integration does not alter the behavior of the system, so previously features are supported for compressed volumes. Starting with IBM Spectrum Virtualize V7.6, two software-enabled RACE engines sharing HW resources for compression are required.

2.3.4 IBM Storwize V7000 licensing

With the broad range of technical features and capabilities of the IBM Storwize V7000 Gen2 and Gen2+, including Copy Services, External Virtualization, Easy Tier, and Real-time Compression, IBM simplified the licensing model to include these new features. The IBM Storwize V7000 offers two ways of license procurement:

- Fully flexible
- Bundled (license packages)

The license model is based on a *license-per-enclosure* concept that is familiar from the first generation of IBM Storwize V7000. However, the second generation offers more flexibility to match your specific needs.

Upgrade: Installing or upgrading the code on the *first* generation of IBM Storwize V7000 to the V8.2 does not change your license model or license needs.

The conceptual model of the licensing in IBM Storwize V7000 Gen2 and Gen2+ is shown in Figure 2-5.

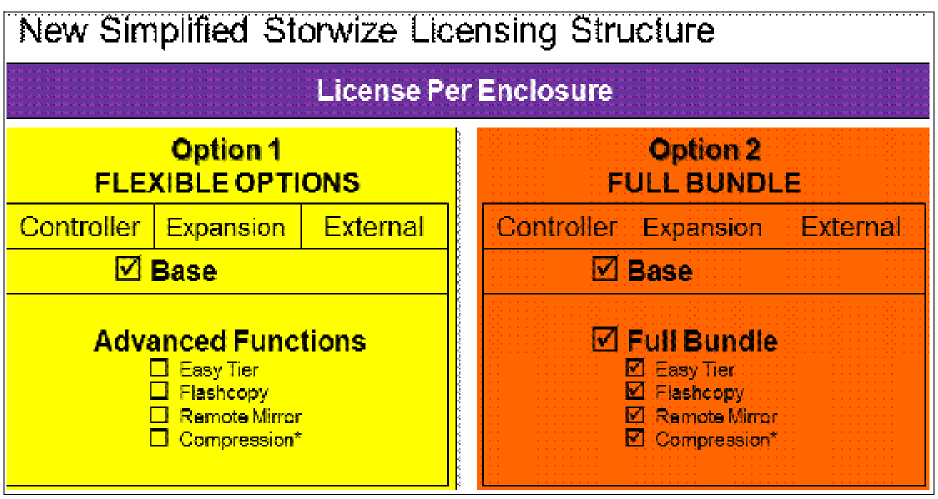


Figure 2-5 Licensing model on IBM Storwize V7000 Gen2 and Gen2+

The base module is represented by *IBM Spectrum Virtualize family* and is mandatory for every controller, enclosure, or externally managed controller unit. Additional licensed features can be purchased on-demand as a full software bundle or each feature separately.

Any additional licenses must be procured per every enclosure where they are planned to be used. For example, if you plan to build a disk array with compression enabled in one enclosure only (no matter if this enclosure is the control or expansion enclosure), you must purchase only one compression license.

2.4 IBM Storwize V7000 hardware

Along with the previous release of the IBM Spectrum Virtualize V7.3, IBM introduced a hardware refresh for the IBM Storwize V7000 platform (code name Gen2) with another update in V7.8 (Gen2+). These improvements are further enhanced with the most recent version of IBM Spectrum Virtualize software V8.2.

This section introduces and partially repeats these hardware changes and software improvements that are associated with these updates. They include the following changes and improvements:

- New internal component layout, such as canister and ports
- Integrated battery pack within node canisters
- Enhanced scalability and flexibility with 16 Gbps I/O adapters
- Improved Real-time Compression engine with hardware assistance
- Extended disk drive support
- Cache upgrade option to 128 GB per canister

To meet these objectives, the base hardware configuration of the IBM Storwize V7000 Gen2 and Gen2+ were substantially improved to support more advanced processors, more memory, and faster data interfaces.

2.5 IBM Storwize V7000 components

The IBM Storwize V7000 is a modular, midrange virtualization RAID storage subsystem that uses the IBM Spectrum Virtualize software engine. It has the following benefits:

- Brings enterprise technology to midrange storage
- Specialty administrators are not required
- Client setup and service are simplified
- The system can grow incrementally as storage capacity and performance needs change
- Multiple storage tiers are in a single system with nondisruptive migration between them
- Simple integration can be done into the server environment

The IBM Storwize V7000 consists of a set of drive enclosures. Control enclosures contain disk drives and two nodes (an I/O group), which are attached to the SAN fabric or 10-gigabit Ethernet (GbE) fast Ethernet. Expansion enclosures contain drives and are attached to control enclosures.

The simplest use of the IBM Storwize V7000 is as a traditional RAID subsystem. The internal drives are configured into RAID arrays, and virtual disks are created from those arrays. With IBM Spectrum Virtualize software, this usage is extended by the deployment of distributed RAID arrays, which shrink the reconstruction time of failed drives.

IBM Storwize V7000 supports spinning disks and flash drives. When different tiers are used, the IBM Storwize V7000 uses IBM Easy Tier to automatically place volume hot spots on better-performing storage. Even without the Easy Tier enabled, the Storage Pool balancing is available and enabled by default to equally distribute workload equally across all MDisks in that tier.

2.5.1 Hosts

A host system is a server that is connected to IBM Storwize V7000 through a Fibre Channel connection, Fibre Channel over Ethernet (FCoE), or through an internet SCSI (iSCSI) connection.

Hosts are defined to IBM Spectrum Virtualize by identifying their worldwide port names (WWPNs) for Fibre Channel hosts. For iSCSI hosts, they are identified by using their iSCSI names. The iSCSI names can be iSCSI qualified names (IQNs) or extended unique identifiers (EUIs).

2.5.2 Host cluster

Host cluster is a host object in the IBM Storwize V7000. A host cluster is a combination of two or more servers that is connected to IBM Storwize V7000 through a Fibre Channel, FCoE, or iSCSI connection. Host cluster object can see the same set of volumes.

2.5.3 Nodes or canisters

An IBM Storwize V7000 can have 2 - 8 hardware components, called *nodes*, *node canisters*, or *canisters*. They perform the virtualization of internal and external volumes and the cache and copy services (remote copy) functions. A clustered system consists of 1 - 4 node pairs.

One of the canisters within the system is known as the *configuration node*. This node manages configuration activity for the clustered system. If this node fails, the system automatically nominates another node to become the configuration node.

Because the active nodes are installed in pairs, each node provides a failover function to its partner node if a node fails.

2.5.4 I/O groups

Within IBM Storwize V7000, 1 - 4 pairs of node canisters are known as *I/O groups*. The IBM Storwize V7000 with installed IBM Spectrum Virtualize supports eight node canisters in the clustered system, which provides four I/O groups.

When a host server performs I/O to one of its volumes, all the I/Os for a specific volume are directed to the I/O group. Also, under normal conditions, the I/Os for that specific volume are always processed by the same node within the I/O group.

One node of the I/O group acts as a preferred node for its own specific subset of the total number of volumes that the I/O group presents to the host servers (a maximum of 2048 volumes per I/O group). However, each node also acts as a failover node for its partner node within the I/O group, so a node takes over the I/O workload from its partner node, if required, with no effect to the server's application.

In an IBM Storwize V7000 environment, the use of active/active architecture and the I/O handling for a volume can be managed by both nodes of the I/O group. Therefore, servers that are connected through Fibre Channel connectors *must* use multipath device drivers handle this capability.

The IBM Storwize V7000 I/O groups are connected to the SAN so that all application servers accessing volumes from the I/O group have access to them. Up to 2048 host server objects can be defined in four I/O groups.

Important: The active/active architecture provides availability to process I/Os for both controller nodes. It enables the application to continue running smoothly, even if the server has only one access route or path to the storage controller. This type of architecture eliminates the path and LUN thrashing typical of an active/passive architecture.

2.5.5 Cache

The primary benefit of storage cache is to improve I/O response time. Reads and writes to a magnetic disk drive experience seek time and latency time at the drive level, which can result in 1 ms - 10 ms of response time (for an enterprise-class disk).

Cache is allocated in 4 KiB segments. A *segment* holds part of one track. A *track* is the unit of locking and destaging granularity in the cache. The cache virtual track size is 32 KiB (eight segments). A track might be only partially populated with valid pages. The IBM Storwize V7000 combines writes up to a 256 KiB track size if the writes are in the same tracks before destage. For example, if 4 KiB is written into a track, another 4 KiB is written to another location in the same track.

Therefore, the blocks that are written from the IBM Storwize V7000 to the disk subsystem can be any size between 512 bytes - 256 KiB. The large cache and advanced cache management algorithms allow it to improve on the performance of many types of underlying disk technologies. The system’s capability to manage, in the background, the destaging operations that are incurred by writes (in addition to still supporting full data integrity) assists with system’s capability in achieving good database performance.

The cache is separated into two layers: An upper cache and a lower cache.

Figure 2-6 shows the separation of the upper and lower cache.

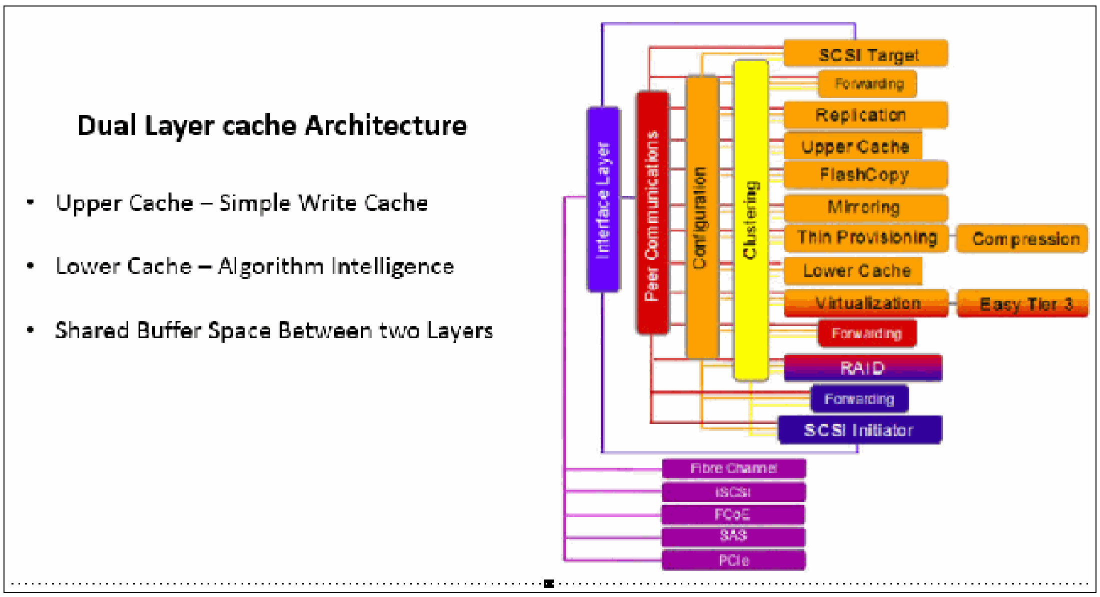


Figure 2-6 Separation of upper and lower cache

The upper cache provides the following functionality, which enables the system to streamline data write performance:

- Fast write response times to the host by being as high up in the I/O stack as possible
- Partitioning

The lower cache delivers the following additional functions:

- Ensures that the write cache between two nodes is in sync
- Caches partitioning to ensure that a slow back end cannot use the entire cache
- Uses a destage algorithm that adapts to the amount of data and the back-end performance
- Provides read caching and prefetching

Combined, the two levels of cache also deliver the following functions:

- Pins data when the LUN goes offline
- Provides enhanced statistics for IBM Tivoli® Storage Productivity Center, and maintains compatibility with an earlier version
- Provides trace for debugging
- Reports medium errors
- Resynchronizes cache correctly and provides the atomic write functionality
- Ensures that other partitions continue operation when one partition becomes 100% full of pinned data
- Supports fast-write (two-way and one-way), flush-through, and write-through
- Integrates with T3 recovery procedures
- Supports two-way operation
- Supports none, read-only, and read/write as user-exposed caching policies
- Supports flush-when-idle
- Supports expanding cache as more memory becomes available to the platform
- Supports credit throttling to avoid I/O skew and offer fairness/balanced I/O between the two nodes of the I/O Group
- Enables switching of the preferred node without needing to move volumes between I/O Groups

Depending on the size, age, and technology level of the disk storage system, the total available cache in the system can be larger, smaller, or about the same as the cache that is associated with the disk storage. Because hits to the cache can occur in either the IBM Storwize V7000 or at the disk controller level of the overall system, the system as a whole can take advantage of the larger amount of cache wherever the cache is located. Therefore, if the storage controller level of the cache has the greater capacity, expect hits to this cache to occur, in addition to hits in the IBM Storwize V7000 cache.

Also, regardless of their relative capacities, both levels of cache tend to play an important role in enabling sequentially organized data to flow smoothly through the system. The IBM Storwize V7000 cannot increase the throughput of the underlying disks in all cases because this increase depends on both the underlying storage technology and the degree to which the workload exhibits *hotspots* or sensitivity to cache size or cache algorithms.

IBM Spectrum Virtualize V7.3 introduced a major upgrade to the cache code. In association with IBM Storwize V7000 Gen2 hardware, it provided another cache capacity upgrade.

At V8.1, these limits were enhanced with IBM Storwize V7000 Gen2+. Before this release, the IBN SAN Volume Controller memory manager (PLMM) only addressed 64 GB of memory.

In V8.1, the underlying PLMM was rewritten and the structure size increased. The cache size can be now upgraded up to 128 GB per canister and the entire memory can now be used.

However, the write cache is still assigned to a maximum of 12 GB and compression cache to a maximum of 34 GB. The remaining installed cache is used as read cache (including allocation for features, such as FlashCopy, Global, or Metro Mirror).

> **Note:** When installing cache in your IBM Storwize V7000 Gen2+ and upgrading to V8.1, the error message "`1199 Detected hardware needs activation`" appears in the GUI event log with an associated error code `0x841` in CLI. Run the command **chnodehw <node_id> -force** for each affected canister in sequence or perform the fix procedure by using the GUI.

2.5.6 Clustered system

A clustered system consists of 1 - 4 pairs of nodes. All configuration, monitoring, and service tasks are performed at the system level, and the configuration settings are replicated across all node canisters in the clustered system. To facilitate these tasks, one or two management IP addresses are set for the system.

A process is provided to back up the system configuration data to disk so that the clustered system can be restored if a disaster occurs. This method does not back up application data, only the IBM Storwize V7000 system configuration information.

> **System configuration backup:** After backing up the system configuration, save the backup data outside of the SAN. If you cannot access the IBM Storwize V7000, you do not have access to the backup data if it is on the SAN.

For the purposes of remote data mirroring, two or more clustered systems (IBM Spectrum Virtualize systems starting from software V7.1) must form a partnership before creating relationships between mirrored volumes.

> **Important:** IBM Storwize V7000 V6.3 introduced the **`layer`** parameter. It can be changed by running **`chsystem`** by using the command-line interface (CLI). The default is the `storage` layer. You must change it to `replication` if you need to set up a copy services relationship between the IBM Storwize family and IBM SAN Volume Controller.

One node is designated as the configuration node canister. It is the only node that activates the system IP address. If the configuration node canister fails, the system chooses a new configuration node and the new configuration node takes over the system IP addresses.

The system can be configured by using either the IBM Spectrum Virtualize GUI-based management software, the CLI, or through an application, such as IBM Spectrum Control, that uses the IBM Storwize V7000 Common Information Model object manager (CIMOM).

2.5.7 HyperSwap

IBM HyperSwap function is a high availability feature that provides dual-site, active-active access to a volume. Active-active volumes have a copy in one site and a copy at another site. Data that is written to the volume is automatically sent to both copies so that either site can provide access to the volume if the other site becomes unavailable. Active-active 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.

Relationships can be grouped into consistency groups as with Metro Mirror and Global Mirror relationships. The consistency groups fail over consistently as a group based on the state of all copies in the group. An image that can be used if a disaster occurs is maintained at each site. When the system topology is set to HyperSwap, each IBM Storwize V7000 controller and hostmap object in the system configuration must have a site attribute set to `1` or `2`.

This site must be the same site as the site of the controllers that provide the managed disks to that I/O group. When managed disks are added to storage pools, their site attributes must match. This requirement ensures that each copy in the active-active relationship is fully independent and is at a distinct site.

2.5.8 Dense expansion drawers

Dense expansion drawers, or dense drawers, are optional 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 92 drives that are positioned in four rows of 14 and another three rows of 12 mounted drives assemblies. Two Secondary Expander Modules (SEM) are centrally located in the chassis. One 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 Storwize V7000 node or another expansion enclosure. SAS connector 2 is optional because it is used to attach to more expansion enclosures.

Each IBM Storwize V7000 can support up to four dense drawers per SAS chain.

Figure 2-7 shows a dense expansion drawer.

Figure 2-7 IBM dense expansion drawer

2.5.9 Expansion Enclosures

The following types of Expansion Enclosures are available:

- IBM Storwize V7000 large form factor (LFF) Expansion Enclosure Model 12F
- IBM Storwize V7000 small form factor (SFF) 24F

IBM Storwize V7000 Gen2 LFF 12F 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 mount enclosure with AC power supplies

IBM Storwize V7000 SFF Expansion Enclosure Model 24F 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 Expansion Enclosure is a 2U enclosure that includes the following components:

- A total of 24 2.5-inch 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.
- RS232 port on the back panel (3.5 mm stereo jack), which is used for configuration during manufacturing.

The front of an Expansion Enclosure is shown in Figure 2-8.

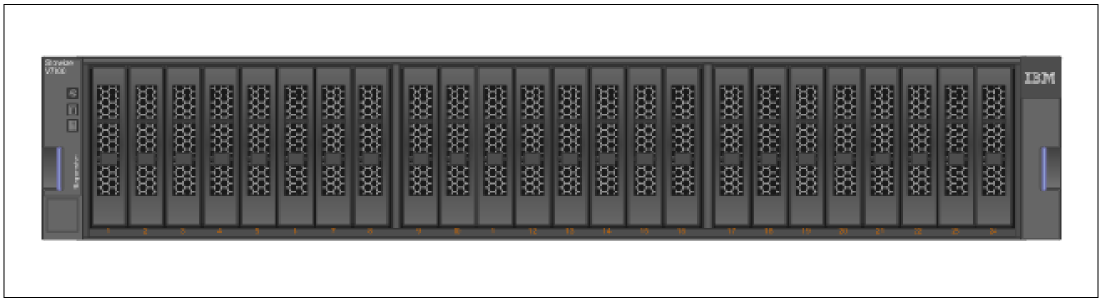


Figure 2-8 Front of IBM Storwize V7000 Expansion Enclosure

Figure 2-9 shows a rear view of an expansion enclosure.

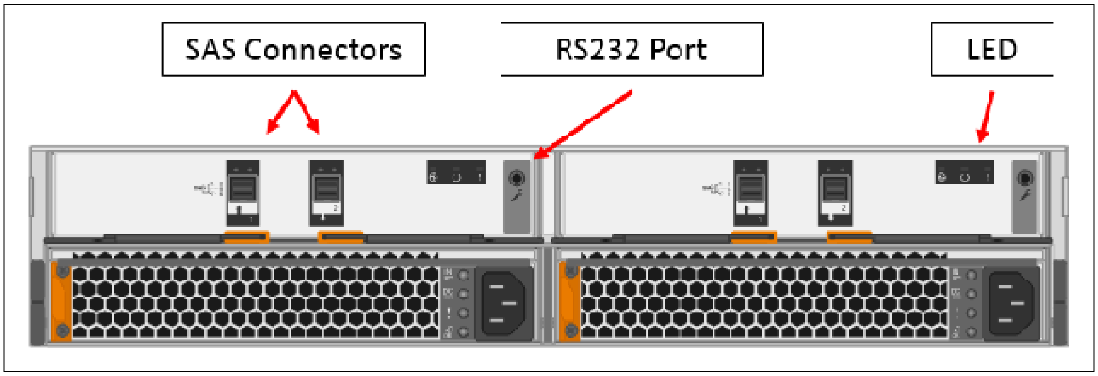


Figure 2-9 Rear of IBM Storwize V7000 expansion enclosure

SAS chain limitations

When attaching expansion enclosures to the control enclosure, you are not limited by type of the enclosure (if it meets all generation level restrictions). The only limitation for each SAS chain is its *chain weight*. Each type of enclosure defined its own chain weight:

- Enclosures 12F and 24F have a chain weight of 1
- Enclosure 92F has a chain weight of 2.5

The maximum chain weight is 10.

For example, you can combine seven 24F and one 92F expansions (7x1 + 1x2.5 = 9.5 chain weight), or two 92F enclosures, one 12F, and four 24F (2x2.5 + 1x1 + 4x1 = 10 chain weight).

An example of chain weight 4.5 with one 24F, one 12F, and one 92F enclosures, all correctly cabled, is shown in Figure 2-10.

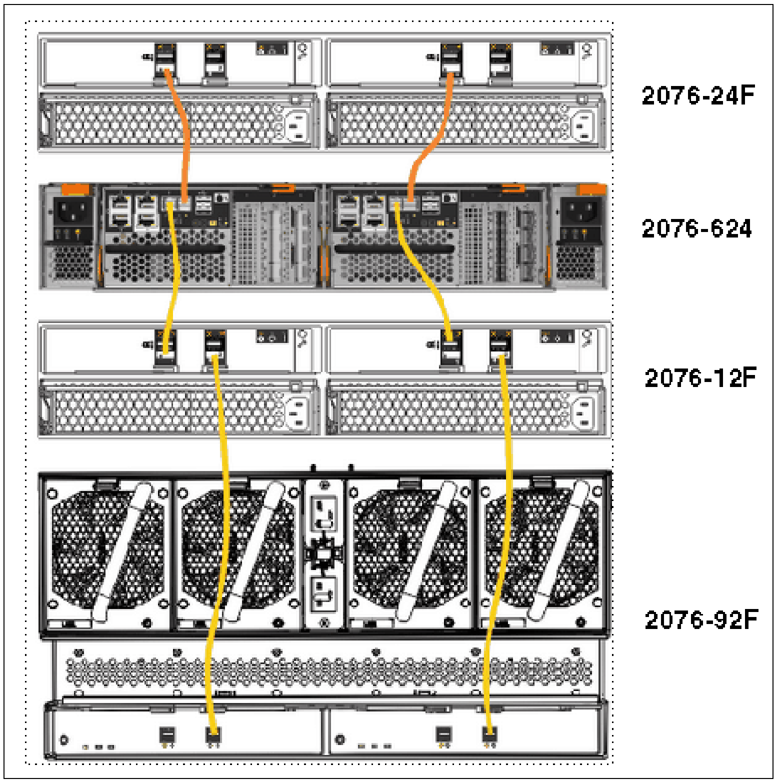


Figure 2-10 Connecting SAS cables while complying with maximum chain weight

2.5.10 RAID

When planning your network, consideration must be given to the type of RAID configuration. The Storwize V7000 supports either the traditional array configuration or the distributed array.

The IBM Storwize V7000 setup contains several internal disk drives, but these disk drives cannot be directly added to the storage pools. Disk drives must be included in a RAID to provide protection against the failure of individual drives.

These disk drives are referred to as *members* of the array. Each array has a RAID level. RAID levels provide various degrees of redundancy and performance, and have various restrictions regarding the number of members in the array.

Apart from traditional disk arrays, IBM Spectrum Virtualize V7.6 introduced distributed RAID. Distributed RAID improves recovery time of failed disk drives in an array by distributing spare capacity between primary disks, rather than dedicating a whole spare drive for replacement.

IBM Storwize V7000 supports hot spare drives. When an array member drive fails, the system automatically replaces the failed member with a hot spare drive and rebuilds the array to restore its redundancy. Candidate and spare drives can be manually exchanged with array members.

Each array has a set of goals that describe the location and performance of each array. A sequence of drive failures and hot spare takeovers can leave an array unbalanced, with members that do not match these goals. The system automatically rebalances such arrays when the appropriate drives are available.

2.5.11 Read Intensive Flash Drives

Generally, two types of SSDs for Enterprise Storage are available: The Multi-level cell (MLC) and Single-level cell (SLC).

The most common SSD technology is MLC. They are found in consumer products, such as portable electronic devices. However, they are also present in some enterprise storage products. Enterprise class SSDs are built on mid- to high-endurance multi-level cell flash technology, known as mainstream endurance SSD.

MLC SSDs uses multi-cell to store data and feature the Wear Leveling method, which is the process to evenly spread the data across all memory cells on the SSD. This method helps to eliminate potential hotspots caused by repetitive Write-Erase cycles.

SLC SSDs use a single cell to store one bit of data, which makes them faster.

To support particular business demands, IBM Spectrum Virtualize qualified the use of Read Intensive (RI) Flash Drives (based on SSD technology) with applications where the read operations are significantly high.

RI Flash Drives are available as an optional purchase product to IBM SAN Volume Controller and IBM Storwize Family.

2.5.12 Managed disks

A managed disk (MDisk) is the unit of storage that IBM Spectrum Virtualize virtualizes. This unit might be a logical volume on an external storage array that is presented to the IBM Storwize V7000, or a RAID consisting of internal drives. IBM Spectrum Virtualize can then allocate these MDisks into various storage pools. An MDisk is not visible to a host system on the storage area network because it is internal or zoned only to the IBM Storwize V7000 system.

An MDisk has the following modes:

- Array

  Array mode MDisks are constructed from drives by using the RAID function. Array MDisks are always associated with storage pools.

- Unmanaged

  Unmanaged MDisks are not being used by the system. This situation might occur when an MDisk is first imported into the system, for example.

- Managed

  Managed MDisks are assigned to a storage pool and provide extents so that volumes can use it.

- Image

  Image MDisks are assigned directly to a volume with a one-to-one mapping of extents between the MDisk and the volume. This situation is normally used when importing logical volumes into the clustered system that already have data on them, which ensures that the data is preserved as it is imported into the clustered system.

2.5.13 Quorum disks

A quorum disk is an MDisk or a managed drive that contains a reserved area for use exclusively by the system. In IBM Storwize V7000, internal drives can be considered as quorum candidates. The clustered system uses quorum disks to break a tie when exactly half the nodes in the system remain after a SAN failure.

The clustered system automatically forms the quorum disk by taking a small amount of space from an MDisk. It allocates space from up to three different MDisks for redundancy, although only one quorum disk is active.

If the environment has multiple storage systems, allocate the quorum disk on different storage systems to avoid the possibility of losing all of the quorum disks because of a failure of a single storage system. It is possible to manage the quorum disks by using the CLI.

2.5.14 IP Quorum

In a HyperSwap configuration, there must be a third, independent site to house quorum devices. To use a quorum disk as the quorum device, this third site must use Fibre Channel connectivity with an external storage system. Sometimes, Fibre Channel connectivity is not possible. In a local environment, no extra hardware or networking, such as Fibre Channel or SAS-attached storage, is required beyond what is normally always provisioned within a system.

To use an IP-based quorum application as the quorum device for the third site, no Fibre Channel connectivity is used. Java applications are run on hosts at the third site. However, strict requirements on the IP network exist, and some disadvantages with the use of IP quorum applications.

Unlike quorum disks, all IP quorum applications must be reconfigured and redeployed to hosts when certain aspects of the system configuration change. These aspects include adding or removing a node from the system, or when node service IP addresses are changed.

For stable quorum resolutions, an IP network must provide the following requirements:

- Connectivity from the hosts to the service IP addresses of all nodes. If IP quorum is configured incorrectly, the network must also deal with possible security implications of exposing the service IP addresses because this connectivity can also be used to access the service GUI.
- Port 1260 is used by IP quorum applications to communicate from the hosts to all nodes.
- The maximum round-trip delay must not exceed 80 ms, which means 40 ms in each direction.
- A minimum bandwidth of 2 MBps is ensured for node-to-quorum traffic.

Even with IP quorum applications at the third site, quorum disks at site one and site two are required because they are used to store metadata. To provide quorum resolution, use the `mkquorumapp` command to generate a Java application that is copied from the system and run on a host at a third site. The maximum number of applications that can be deployed is five. Currently, supported Java runtime environments (JREs) are IBM Java 7.1 and IBM Java 8.

2.5.15 Storage pool

A storage pool is a collection of MDisks that are grouped to provide capacity for volumes. All MDisks in the pool are split into extents with the same size. Volumes are then allocated out of the storage pool and are mapped to a host system.

IBM Storwize V7000 object names: The names *must* begin with a letter and can be a maximum of 63 characters. Valid characters for names are uppercase letters (A-Z), lowercase letters (a-z), digits (0 - 9), underscore (_), period (.), hyphen (-), and space. The names must not begin or end with a space.

MDisks can be added to a storage pool at any time to increase the capacity of the storage pool. MDisks can belong in only one storage pool, and only MDisks in unmanaged mode can be added to the storage pool. When an MDisk is added to the storage pool, the mode changes from unmanaged to managed (and vice versa when you remove it).

Each MDisk in the storage pool is divided into several extents. The size of the extent is selected by the administrator at creation time of the storage pool, and cannot be changed later. The size of the extent ranges from 16 MB - 8 GB. The default extent size in V8.1 is 1024 MB.

The extent size directly affects the maximum volume size and storage capacity of the clustered system. A system can manage 4 million (4 x 1024 x 1024) extents. For example, a system with a 16 MB extent size can manage up to 16 MB x 4 MB = 64 TB of storage.

Use the same extent size for all storage pools in a clustered system, which is a prerequisite for supporting volume migration between two storage pools. If the storage pool extent sizes are not the same, you must use volume mirroring to copy volumes between storage pools.

Default extent size: The IBM Storwize V7000 GUI has a default extent size value of 1024 MB when you define a new storage pool. This setting can only be changed by using the CLI.

Figure 2-11 shows the storage pool components.

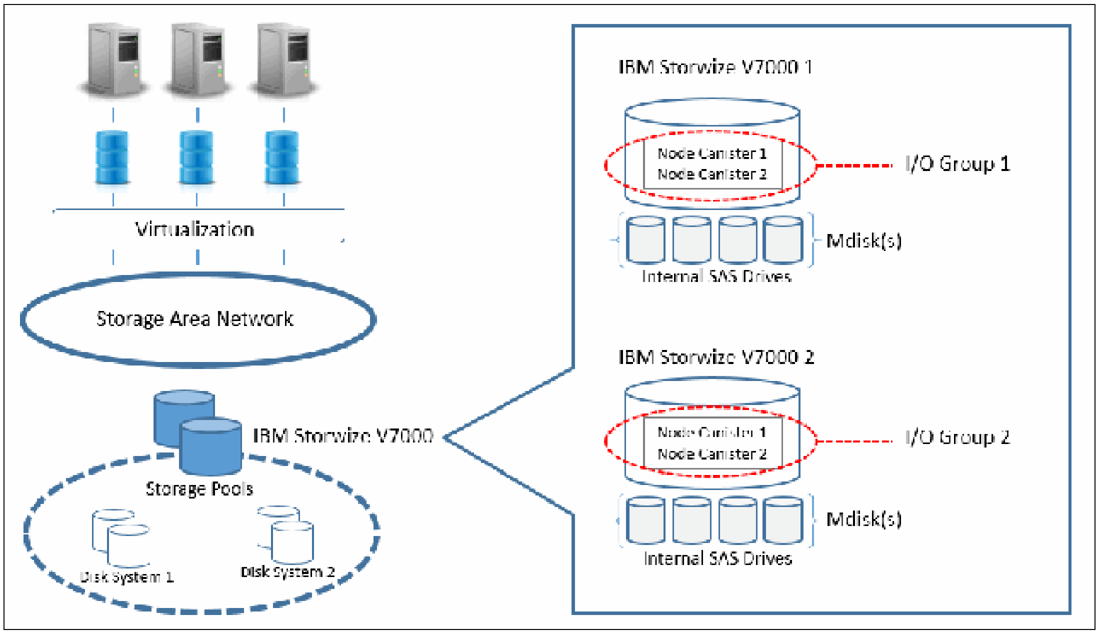


Figure 2-11 IBM Spectrum Virtualize virtualization components

2.5.16 Volumes

A volume is a logical disk that is presented to a host system by the clustered system. In our virtualized environment, the host system has a volume mapped to it by IBM Storwize V7000. IBM Spectrum Virtualize translates this volume into several extents, which are allocated across MDisks. The advantage of storage virtualization is that the host is “decoupled” from the underlying storage, so the virtualization appliance can move the extents without affecting the host system.

The host system cannot directly access the underlying MDisks in the same manner as it can access RAID arrays in a traditional storage environment.

The following types of volumes are available:

- Striped

  A striped volume is allocated one extent in turn from each MDisk in the storage pool. This process continues until the space required for the volume is satisfied.

  It is also possible to supply a list of MDisks to use.

Figure 2-12 shows how a striped volume is allocated, assuming that 10 extents are required.

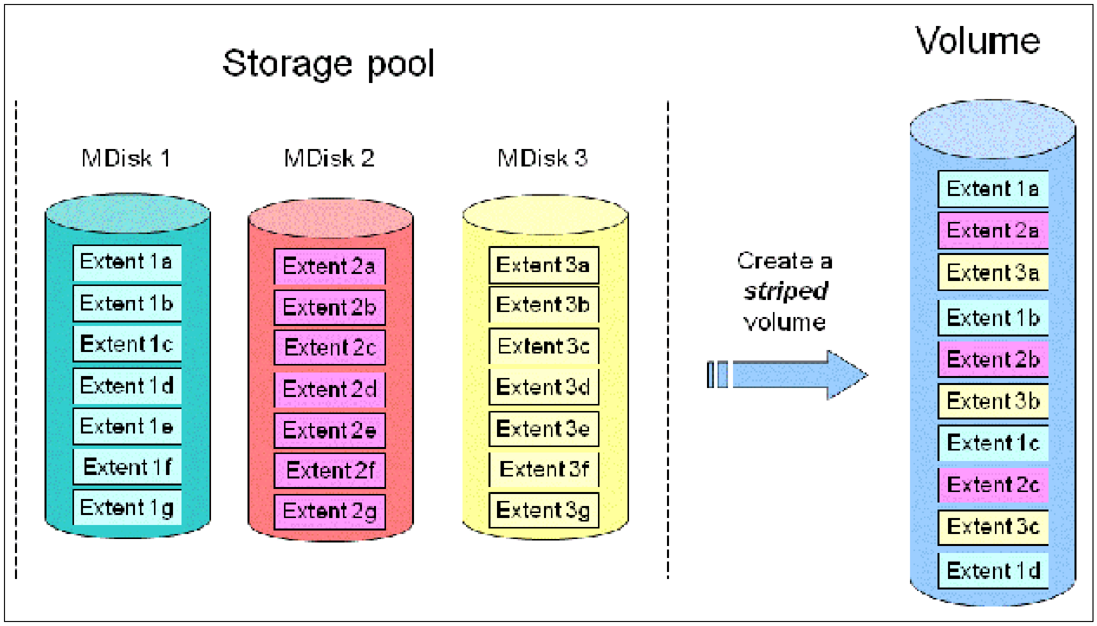


Figure 2-12 Striped volume

- Sequential

  A sequential volume is where the extents are allocated one after the other, from one MDisk to the next MDisk (see Figure 2-13).

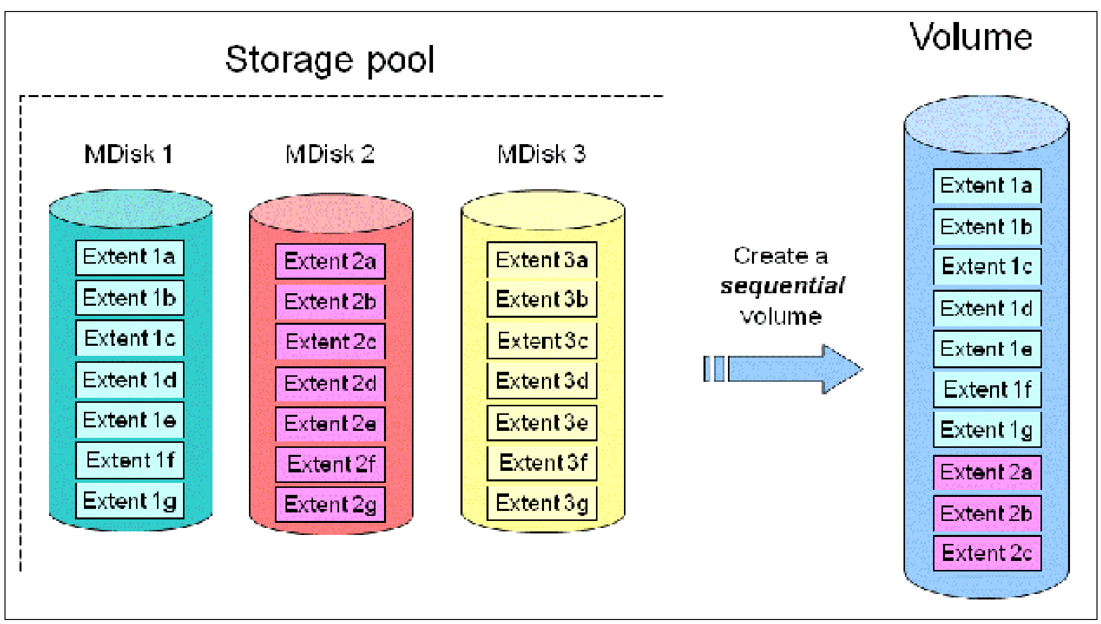


Figure 2-13 Sequential volume

- Image mode

  Image mode volumes are special volumes that have a direct relationship with one MDisk. They are used to migrate data into and out of the clustered system.

When the image mode volume is created, a direct mapping is made between extents that are on the MDisk and the extents that are on the volume. The LBA *x* on the MDisk is the same as the LBA *x* on the volume, which ensures that the data on the MDisk is preserved as it is brought into the clustered system (see Figure 2-14).

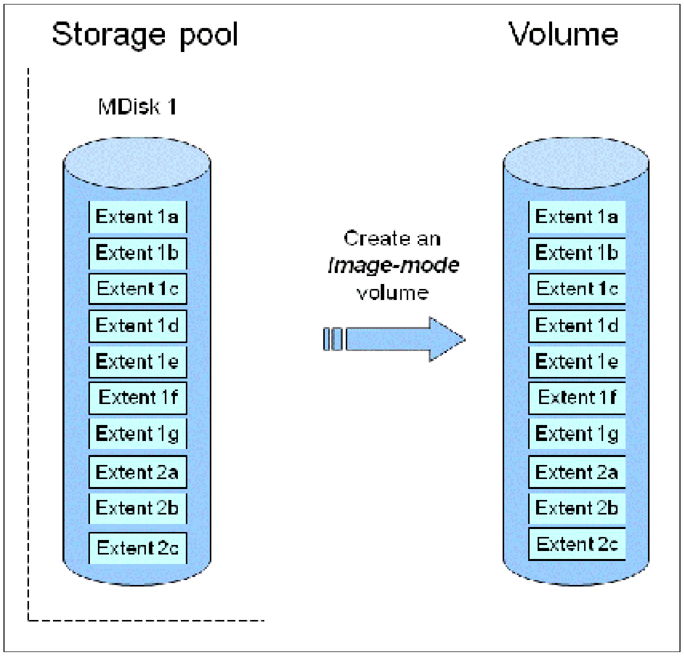


Figure 2-14 Image mode volume

Some virtualization functions are not available for image mode volumes, so it is often useful to migrate the volume into a new storage pool. After it is migrated, the MDisk becomes a managed MDisk.

If you add an MDisk that contains data to a storage pool, any data on the MDisk is lost. Ensure that you create image mode volumes from MDisks that contain data before adding MDisks to the storage pools.

2.5.17 Easy Tier

IBM Easy Tier is a performance optimization function that automatically migrates or moves the extents of a volume across two or more tiers. Easy Tier monitors the host I/O activity and latency on the extent of all volumes with the Easy Tier function turned on, in a multitiered storage pool, over a 24-hour period.

After this period, the IBM Easy Tier creates an extent migration plan that is based on this activity, and then dynamically moves high activity (or *hot*) extents to a higher disk tier within the storage pool. It also moves extent activity that dropped off (or *cooled*) from the high-tiered MDisk back to a lower-tiered MDisk.

The Easy Tier function can be turned on or off at the storage pool and the volume level. The *automatic storage pool balancing* is an integrated part of the IBM Easy Tier engine and is enabled by default on all pools. Storage Pool Balancing and IBM Easy Tier do not require more licenses within IBM Spectrum Virtualize that is running on IBM Storwize V7000. This load balancing feature is not considered an Easy Tier function, although it uses the same principles and mechanism.

It is possible to demonstrate the potential benefit of IBM Easy Tier in the environment before installing different drives. When IBM Easy Tier is turned on, the algorithm produces the statistic file that can be offloaded from IBM Storwize V7000.

The IBM Storage Tier Advisor Tool (STAT) can be used to create a summary report that uses the statistic file offloaded from IBM Storwize V7000.

The STAT tool is available at this website.

2.5.18 Encryption

IBM Storwize V7000 provides optional encryption of data at rest, which protects against the potential exposure of sensitive user data and user metadata that is stored on discarded, lost, or stolen storage devices. Encryption of system data and system metadata is not required, so system data and metadata are not encrypted.

Planning for encryption involves purchasing a licensed function and then activating and enabling the function on the system. A license is required to encrypt data that is stored on drives. When encryption is activated and enabled on the system, valid encryption keys must be present on the system when the system unlocks the drives or the user generates a key.

IBM Spectrum Virtualize V7.4, hardware encryption was introduced, with software encryption introduced in V7.6. Encryption keys can be managed by IBM SKLM or stored on USB flash drives that are attached to a minimum of one of the nodes. Since V8.1, a combination of SKLM and USB key repositories can be used.

IBM Security Key Lifecycle Manager is an IBM solution that provides the infrastructure and processes to locally create, distribute, backup, and manage the lifecycle of encryption keys and certificates. Before activating and enabling encryption, you must determine the method of accessing key information during times when the system requires an encryption key to be present.

When Security Key Lifecycle Manager is used as a key manager for encryption, you can run into a deadlock situation when the key servers are hosted on encrypted storage provided by the IBM Storwize V7000. That is, if the storage that is used by these SKLM servers is allocated from the encrypted pool on the same Storwize V7000. To avoid a deadlock situation, ensure that the IBM Storwize V7000 node canisters can communicate with an encryption server to get the unlock key after a power-on or restart scenario.

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 Institute of Electrical and Electronics Engineers (IEEE) 1619-2007 standard as XTS-AES-256. That data encryption key is protected by a 256-bit AES key wrap when stored in non-volatile form.

2.5.19 iSCSI

iSCSI is an alternative means of attaching hosts and external storage controllers to the IBM Storwize V7000. Within IBM Spectrum Virtualize release 7.7, IBM introduced software capabilities to allow the underlying virtualized storage to attach to IBM Storwize V7000 by using the iSCSI protocol.

In the simplest terms, iSCSI enables the transport of SCSI commands and data over an Internet Protocol network, based on IP routers and Ethernet switches. iSCSI is a block-level protocol that encapsulates SCSI commands into Transmission Control Protocol/Internet Protocol (TCP/IP) packets and uses an IP network, rather than requiring expensive FC host bus adapters (HBAs) and a SAN fabric infrastructure.

The major functions of iSCSI include encapsulation and the reliable delivery of CDB transactions between initiators and targets through the Internet Protocol network, especially over a potentially unreliable IP network.

Every iSCSI node in the network must have an iSCSI name and address. Consider the following points:

- An *iSCSI name* is a location-independent, permanent identifier for an iSCSI node. An iSCSI node has one iSCSI name that stays constant for the life of the node. The terms *initiator name* and *target name* also refer to an iSCSI name.
- An *iSCSI address* specifies not only the iSCSI name of an iSCSI node, but a location of that node. The address consists of a host name or IP address, a TCP port number (for the target), and the iSCSI name of the node. An iSCSI node can have any number of addresses, which can change at any time, particularly if they are assigned by way of Dynamic Host Configuration Protocol (DHCP). An IBM SAN Volume Controller node represents an iSCSI node and provides statically allocated IP addresses.

2.5.20 Real-time Compression

IBM Real-time Compression is an attractive solution to address increasing data storage requirements for power, cooling, and floor space. When applied, IBM Real-time Compression can significantly save storage space because more data is stored in rack space so fewer storage enclosures are required to store a data set. IBM Real-time Compression provides the following benefits:

- Compression for active primary data. IBM Real-time Compression can be used with active primary data.
- Compression for replicated/mirrored data. Remote volume copies can be compressed in addition to the volumes at the primary storage tier. This process also reduces storage requirements in Metro Mirror and Global Mirror destination volumes.
- No changes to the environment are required. IBM Real-time Compression is part of the storage system.
- Overall savings in operational expenses. More data is stored in a rack space, so fewer storage expansion enclosures are required to store a data set. This reduced rack space has the following benefits:
  - Reduced power and cooling requirements. More data is stored in a system, which requires less power and cooling per gigabyte or used capacity.
  - Reduced software licensing for more functions in the system. More data stored per enclosure reduces the overall spending on licensing.
- Disk space savings are immediate. The space reduction occurs when the host writes the data. This process is unlike other compression solutions, in which some or all of the reduction is realized only after a post-process compression batch job is run.

2.5.21 Data Reduction Pools

Data Reduction Pools (DRP) represent a significant enhancement to the storage pool concept. The reason is the virtualization layer is primarily a simple layer that runs the task of lookups between virtual and physical extents.

Data Reduction Pools is a new type of storage pool and implements techniques, such as thin-provisioning, compression, and deduplication to reduce the amount of physical capacity that is required to store data, and decrease the network infrastructure required. Savings in storage capacity requirements translate into reduction in the cost of storing the data.

The storage pools enable you to automatically de-allocate and reclaim capacity of thin-provisioned volumes that contain deleted data and for the first time, enable this reclaimed capacity to be reused by other volumes. Data reduction provides more performance from compressed volumes because of the new log structured pool that is implemented.

2.5.22 Deduplication

Data deduplication is one of the methods of reducing storage needs by eliminating redundant copies of a file. Data reduction is a way to decrease the storage disk and network infrastructure that is required, and optimize use of storage disks and improve data recovery infrastructure efficiency. Existing data or new data are standardized into chunks that are examined for redundancy. If data duplicates are detected, pointers are shifted to reference a single copy of the chunk and the extra duplicate data is then released.

Deduplication has benefits, such as storing more data per physical storage system, which saves energy by using fewer disk drives and decreasing the amount of data that must be sent across a network to another storage for backup replication and disaster recovery.

2.5.23 IP replication

IP replication was introduced in V7.2 and allows data replication between IBM Spectrum Virtualize family members. IP replication uses IP-based ports of the cluster node canisters. This function is transparent to servers and applications in the same way that traditional FC-based mirroring is. All remote mirroring modes (Metro Mirror, Global Mirror, and Global Mirror with changed volumes) are supported.

The configuration of the system is straightforward, and IBM Storwize family systems normally find each other in the network and can be selected from the GUI.

IP replication includes Bridgeworks SANSlide network optimization technology and is available at no extra charge. Remember, remote mirror is a chargeable option, but the price does not change with IP replication. Remote mirror users have access to the function at no extra charge.

IP connections that are used for replication can have long latency (the time to transmit a signal from one end to the other). This latency can be caused by distance or by many “hops” between switches and other appliances in the network. Traditional replication solutions transmit data, wait for a response, and then transmit more data, which can result in network utilization as low as 20% (based on IBM measurements).

Bridgeworks SANSlide technology, which is integrated with the IBM Storwize family, requires no separate appliances and so requires no extra cost and no configuration steps. It uses artificial intelligence (AI) technology to transmit multiple data streams in parallel, adjusting automatically to changing network environments and workloads.

SANSlide improves network bandwidth utilization up to 3x. Therefore, customers can deploy a less costly network infrastructure, or take advantage of faster data transfer to speed replication cycles, improve remote data currency, and enjoy faster recovery.

2.5.24 IBM Storwize V7000 copy services

IBM Spectrum Virtualize supports the following copy services:

- Synchronous remote copy (Metro Mirror)
- Asynchronous remote copy (Global Mirror)
- FlashCopy (Point-in-Time Copy)
- Transparent Cloud Tiering

Copy services functions are implemented within a single IBM Storwize V7000 or between multiple members of the IBM Spectrum Virtualize family. The Copy Services layer sits above and operates independently of the function or characteristics of the underlying disk subsystems that are used to provide storage resources to an IBM Storwize V7000.

2.5.25 Synchronous or asynchronous remote copy

The general application of remote copy seeks to maintain two copies of data. Often, the two copies are separated by distance, but not always. The remote copy can be maintained in synchronous or asynchronous modes.

With IBM Spectrum Virtualize, *Metro Mirror* and *Global Mirror* are the IBM branded terms for the functions that are synchronous remote copy and asynchronous remote copy.

Synchronous remote copy ensures that updates are committed at the primary and secondary volumes before the application considers the updates complete. Therefore, the secondary volume is fully up-to-date if it is needed in a failover.

However, the application is fully exposed to the latency and bandwidth limitations of the communication link to the secondary volume. In a truly remote situation, this extra latency can have a significant, adverse effect on application performance.

Special configuration guidelines exist for SAN fabrics and IP networks that are used for data replication. Consider the distance and available bandwidth of the intersite links when planning for synchronous remote copy.

A function of Global Mirror that is designed for low bandwidth was introduced in IBM Spectrum Virtualize. It uses change volumes that are associated with the primary and secondary volumes. These volumes are used to record changes to the remote copy volume, the FlashCopy relationship that exists between the secondary volume and the change volume, and between the primary volume and the change volume.

This function is called *Global Mirror cycling mode*. Figure 2-15 shows an example of this function where you can see the relationship between volumes and change volumes.

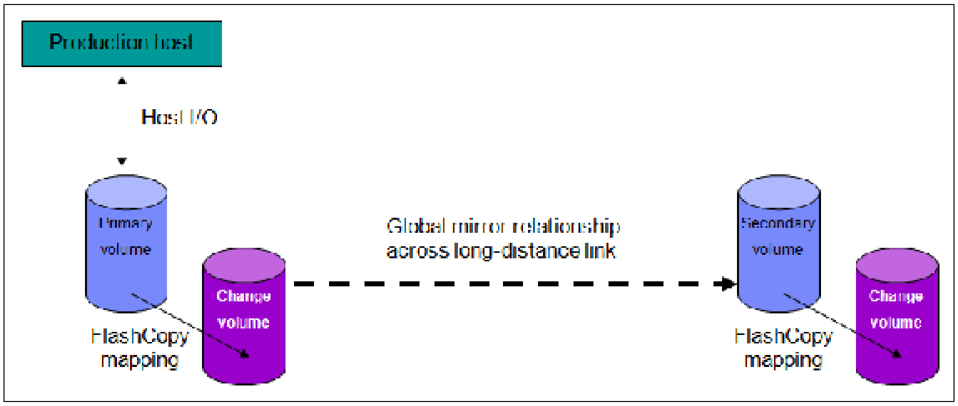


Figure 2-15 Global Mirror with change volumes

In asynchronous remote copy, the application acknowledges that the write is complete before the write is committed at the secondary volume. Therefore, on a failover, certain updates (data) might be missing at the secondary volume. The application must have an external mechanism for recovering the missing updates, if possible. This mechanism can involve user intervention. Recovery on the secondary site involves starting the application on this recent backup, and then rolling forward or backward to the most recent commit point.

2.5.26 FlashCopy and Transparent Cloud Tiering

FlashCopy and Cloud Backup are used to make a copy of a source volume on a target volume. After the copy operation starts, the original content of the target volume is lost. The target volume has the contents of the source volume as they existed at a single point in time. Although the copy operation takes time, the resulting data at the target appears as though the copy was made instantaneously.

FlashCopy

FlashCopy is sometimes described as an instance of a time-zero (T0) copy or a point-in-time (PiT) copy technology.

FlashCopy can be performed on multiple source and target volumes. FlashCopy permits the management operations to be coordinated so that a common single point in time is chosen for copying target volumes from their respective source volumes.

With IBM Spectrum Virtualize, multiple target volumes can undergo FlashCopy from the same source volume. This capability can be used to create images from separate points in time for the source volume, and to create multiple images from a source volume at a common point in time. Source and target volumes can be thin-provisioned volumes.

Reverse FlashCopy enables target volumes to become restore points for the source volume without breaking the FlashCopy relationship, and without waiting for the original copy operation to complete. IBM Spectrum Virtualize supports multiple targets and therefore, multiple rollback points.

Most clients aim to integrate the FlashCopy feature for point-in-time copies and quick recovery of their applications and databases. An IBM solution to this configuration is provided by IBM Spectrum Protect™, which is available at this website.

Transparent Cloud Tiering

IBM Spectrum Transparent Cloud Tiering is a new function that was introduced in IBM Spectrum Virtualize V7.8. Transparent Cloud Tiering is an alternative solution for data protection, backup, and restore that interfaces to Cloud Service Providers, such as IBM Cloud™, Amazon S3, and OpenStack Swift. Transparent Cloud Tiering is charged as additional priced software per IBM Storwize V7000 Controller.

The Transparent Cloud Tiering function helps organizations to reduce costs related to power and cooling when off-site data protection is required to send sensitive data out of the main site.

Transparent Cloud Tiering uses IBM FlashCopy techniques that provide full and incremental snapshots of one or more volumes. Snapshots are encrypted and compressed before being uploaded to the cloud. Reverse operations are also supported within that function. When a set of data is transferred out to cloud, the volume snapshot is stored as an Object Storage.

Cloud Object Storage uses innovative approach and cost-effective solution to store large amount of unstructured data and delivers mechanisms to provide security services, high availability, and reliability.

The management GUI provides an easy-to-use initial setup, advanced security settings, and audit logs that records all backup and restore to cloud.

For more information about cloud Object Storage, see this website.

2.6 Business continuity

Business continuity and continuous application availability are among the top requirements for many organizations. Advances in virtualization, storage, and networking have made enhanced business continuity possible.

Information technology solutions can now manage planned and unplanned outages, and provide the flexibility and cost efficiencies that are available from cloud-computing models.

2.6.1 Business Continuity with HyperSwap

The HyperSwap high availability feature in the IBM Spectrum Virtualize software allows business continuity in the event of hardware failure, power failure, connectivity failure, or disasters, such as fire or flooding. The HyperSwap feature is available on the IBM SAN Volume Controller, IBM Storwize V7000, IBM Storwize V7000 Unified, and IBM Storwize V5000 products.

The HyperSwap feature provides highly available volumes accessible through two sites at up to 300 km (186.4 miles) 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 transmitted between sites and to minimize host read and write latency.

HyperSwap includes the following key features:

- Works with IBM SAN Volume Controller and IBM Storwize V7000, V5000, and V7000 Unified hardware.
- Uses intra-cluster synchronous remote copy (Metro Mirror) capabilities along with existing change volume and access I/O group technologies.

- Makes a host’s volumes accessible across two IBM V7000 or V5000 Storwize or IBM SAN Volume Controller I/O groups in a clustered system using the Metro Mirror relationship in the background. The volumes appear a single volume to the host.
- Works with the standard multipathing drivers that are available on a various host types, with no extra host support required to access the highly available volume.

2.7 Management and support tools

The IBM Spectrum Virtualize system can be managed through the included management software that runs on the IBM Storwize V7000 hardware.

2.7.1 IBM Assist On-site and Remote Support Assistance

The IBM Assist On-site tool is a remote desktop-sharing solution that is offered through the IBM website. With it, the IBM service representative can remotely view your system to troubleshoot a problem.

You can maintain a chat session with the IBM service representative so that you can monitor this activity and either understand how to fix the problem yourself or allow the representative to fix it for you.

To use the IBM Assist On-site tool, the master console must access the internet. For more information about this tool, see this website.

When you access the website, you sign in and enter a code that the IBM service representative provides to you. This code is unique to each IBM Assist On-site session. A plug-in is downloaded onto your master console to connect you and your IBM service representative to the remote service session. The IBM Assist On-site tool contains several layers of security to protect your applications and your computers. The plug-in is removed after the next restart.

You can also use security features to restrict access by the IBM service representative. Your IBM service representative can provide you with more detailed instructions for using the tool.

The embedded part of the IBM Spectrum Virtualize V8.1 code is a software toolset that is called Remote Support Client. It establishes a network connection over a secured channel with Remote Support Server in the IBM network. The Remote Support Server provides predictive analysis of IBM SAN Volume Controller status and assists administrators for troubleshooting and fix activities. *Remote Support Assistance* is available for no extra fee, and no additional license is needed.

2.7.2 Event notifications

IBM Storwize V7000 can use Simple Network Management Protocol (SNMP) traps, syslog messages, and a 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.

Each event that IBM Storwize V7000 detects is assigned a notification type of `Error`, `Warning`, or `Information`. You can configure the IBM Storwize V7000 to send each type of notification to specific recipients.

2.7.3 Simple Network Management Protocol traps

SNMP is a standard protocol for managing networks and exchanging messages. The IBM Spectrum Virtualize can send SNMP messages that notify personnel about an event. You can use an SNMP manager to view the SNMP messages that IBM Spectrum Virtualize sends. You can use the management GUI or the IBM Storwize V7000 CLI to configure and modify your SNMP settings.

You can use the Management Information Base (MIB) file for SNMP to configure a network management program to receive SNMP messages that are sent by the IBM Spectrum Virtualize.

2.7.4 Syslog messages

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.

IBM Storwize V7000 can send syslog messages that notify personnel about an event. The event messages can be sent in expanded or concise format. You can use a syslog manager to view the syslog messages that IBM Storwize V7000 sends.

IBM Spectrum Virtualize uses the User Datagram Protocol (UDP) to transmit the syslog message. You can use the management GUI or the IBM Storwize V7000 CLI to configure and modify your syslog settings.

2.7.5 Call Home email

The Call Home feature transmits operational and error-related data to you and IBM through an SMTP server connection in the form of an event notification email. When configured, this function alerts IBM service personnel about hardware failures and potentially serious configuration or environmental issues. You can use the Call Home function if you have a maintenance contract with IBM or if the IBM Storwize V7000 is within the warranty period.

To send email, at least one SMTP server must be configured. The system support as many as five more SMTP servers for backup purposes. The SMTP server must accept the relaying of email from the IBM Storwize V7000 clustered system IP address.

Use the management GUI or the IBM Storwize V7000 CLI to configure the email settings, including contact information and email recipients. Set the reply address to a valid email address.

Send a test email to check that all connections and infrastructure are set up correctly. The Call Home function can be disabled at any time by using the management GUI or the IBM Storwize V7000 CLI.

2.8 Useful IBM Storwize V7000 websites

See the following IBM Storwize V7000 web pages for more information:

- IBM Support page:

  https://www.ibm.com/support/home/product/5402112/IBM_Storwize_V7000_(2076)
- IBM Storwize V7000 Unified and IBM Storwize V7000 Systems:

  https://www.ibm.com/support/home/product/5421300/IBM_Storwize_V7000_Unified
- IBM Storwize V7000 page support

  http://www-01.ibm.com/support/docview.wss?uid=ssg1S1003741
- Direct attachment of IBM Storwize V7000

  https://www-01.ibm.com/support/docview.wss?uid=ssg1S1005776
- IBM Knowledge Center:

  https://www.ibm.com/support/knowledgecenter/en/ST3FR7_8.2.1/com.ibm.storwize.v7000.821.doc/v7000_ichome.html

3

Planning

This chapter describes steps that are required to plan the installation of an IBM System Storage SAN Volume Controller in your storage network.

This chapter includes the following topics:

3.1 General planning rules

> **Important:** At the time of this writing, the statements that are provided in this book are accurate but can change. Always verify any statements that are made in this book with the IBM Storwize V7000 supported hardware list, device driver, firmware, and recommended software levels information that are available at the following websites:
>
> - Support Information for Storwize V7000
> - IBM System Storage Interoperation Center (SSIC)

To maximize the benefit that is realized from the Storwize V7000, pre-installation planning must include several important steps. These steps ensure that the Storwize V7000 provides the best possible performance, reliability, and ease of management for your application needs. The correct configuration also helps minimize downtime by avoiding changes to the Storwize V7000 and the storage area network (SAN) environment to meet future growth needs.

This book is *not* intended to provide in-depth information about the described topics. For an enhanced analysis of advanced topics, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

3.1.1 Basic planning flow

The general rule of planning is to define your goals, and then, plan a solution that can be shown to meet these goals. Always remember to verify that each element of your configuration is supported.

Consider the following points when planning for the Storwize V7000:

- Collect and document the number of hosts (application servers) to attach to the Storwize V7000. Identify the traffic profile activity (read or write, sequential, or random), and the performance requirements (bandwidth and input/output [I/O] operations per second [IOPS]) for each host.
- Decide whether you are going to use Storwize V7000 to virtualize external storage. If you do, collect and document the following items:
  - Information on the back-end storage that exists in the environment and is intended to be virtualized by the Storwize V7000.
  - Whether you must configure image mode volumes. If you want to use image mode volumes, decide whether and how you plan to migrate them into managed mode volumes.
  - Information about the planned new back-end storage to be virtualized by the Storwize V7000.
  - The required virtual storage capacity for fully provisioned and space-efficient (SE) volumes.
  - The required storage capacity for:
    - Local mirror copy (volume mirroring)
    - Point-in-time copy (IBM FlashCopy)
    - Remote copy (Metro Mirror and Global Mirror)
    - Compressed volumes
    - Encrypted volumes

  – Shared storage (volumes presented to more than one host) that is required in your environment.

- Define the amount of internal storage that you plan to provision from internal arrays. Consider the following information:
  – Amount of storage in each drive tier that you plan to deploy.
  – The number of drives that you must provision the required storage. Consider RAID levels, number of drives per array, and hot spares for each tier. Define the number of drives and their form factor.

  > **Note:** Use DRAID 6 as the default RAID level for your arrays.

  – The number of expansion enclosures you must hold the drives required to provision the required storage.
- Per host:
  – Volume capacity.
  – Logical unit number (LUN) quantity.
  – Volume sizes.

> **Note:** When planning the capacities, note if the numbers state the net storage capacity (that is, available to be used by applications running on any host), or gross capacity, which includes overhead for spare drives (both because of RAID redundancy and planned hot spare drives) and for file system metadata.
>
> For file system metadata, include overhead incurred by all layers of storage virtualization. In particular, if you plan storage for virtual machines whose drives are actualized as files on a parallel file system, include metadata overhead for the storage virtualization technology that is used by your hypervisor software.

- Decide whether you must plan for more than one site. For multi-site deployment, review the extra configuration requirements that are imposed.
- Define the number of clusters and the number of enclosures (1 - 4) for each cluster. The number of necessary I/O Groups depends on the overall performance requirements and the number of hosts you plan to attach.
- Decide whether you are going to use N_Port ID Virtualization (NPIV). If you plan to use NPIV, review the extra configuration requirements that are imposed.
- Design the SAN according to the requirement for high availability (HA) and best performance. Consider the total number of ports and the bandwidth that is needed at each link, especially Inter-Switch Links (ISLs). Consider ISL trunking for improved performance. Separately collect requirements for Fibre Channel and IP-based storage network.

> **Note:** Check and carefully count the required ports. Separately note the ports that are dedicated for extended links. You might need more long-wave gigabit interface converters (GBICs), especially in an enhanced stretched cluster (ESC) or HyperSwap environment.

- Define a naming convention for the Storwize V7000 clusters, nodes, hosts, and storage objects.
- Define the Storwize V7000 service Internet Protocol (IP) addresses and the system's management IP addresses.

- Define subnets for the Storwize V7000 system and for the hosts for internet Small Computer System Interface (iSCSI) connectivity.
- Define the IP addresses for IP replication (if required).
- Define back-end storage that will be used by the system.
- Define the managed disks (MDisks) in the back-end storage to be used by Storwize V7000.
- Define the storage pools, specify MDisks for each pool and document mapping of MDisks to back-end storage. Parameters of the back-end storage determine the characteristics of the volumes in the pool. Make sure that each pool contains MDisks of similar (ideally, identical) performance characteristics.
- Plan allocation of hosts and volumes to I/O Groups to optimize the I/O load distribution between the hosts and the Storwize V7000. Allowing a host to access more than one I/O group might better distribute the load between system nodes. However, doing so reduces the maximum number of hosts that are attached to the Storwize V7000.
- Plan queue depths for the attached hosts. For more information, see this website.
- Plan for the physical location of the equipment in the rack.
- Verify that your planned environment is a supported configuration.
- Verify that your planned environment does not exceed system configuration limits.

Planning activities required for Storwize V7000 deployment are described next.

3.2 Planning for availability

When planning deployment of IBM Storwize V7000, avoid creating single points of failure. Plan system availability according to the requirements specified for your solution. Consider the following aspects, depending on your availability needs:

- Single site or multi-site configuration

  Multi-site configurations increase solution resiliency and can be the basis of disaster recovery solutions. Storwize V7000 allows configuration of multi-site solutions, with sites working in active-active or active-standby mode. Synchronous and asynchronous data replication is supported by multiple inter-site link options.

  For more information about cross-site configurations, see *IBM Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317.

- Physical separation of system building blocks

  Dual rack deployment might increase availability of your system if your back-end storage, SAN, and LAN infrastructure also do not use single rack placement scheme. You can further increase system availability by ensuring that cluster nodes are powered from different power circuits and are in different fire protection zones.

- Quorum availability

  The Storwize V7000 uses three drives that are configured as array members or hot spares, as quorum disks for the clustered system. If no external MDisks can be used as quorum disks, the preferred practice is to distribute quorum disks over SAS channels to increase system resiliency. The current locations of the quorum disks can be displayed by using the **lsquorum** command, and relocated by using the **chquorum** command.

For clustered Storwize V7000 (two or more I/O Groups) systems that do not virtualize external storage (and therefore cannot place a quorum disk on an external MDisk), generally use IP quorum device to reduce the risk of outage in case of a split-brain cluster.

For HyperSwap systems, it is recommended to have at least two independent IP quorum devices defined to prevent the active quorum device placement in one of sites that hold Storwize V7000 HyperSwap storage components. Also, make sure that at least one quorum drive is in each site.

The IP quorum device acts only as a tie breaker in case of a split-brain cluster scenario. Quorum disks are still required to store cluster state.

- Failure domain sizes

  Failure of an MDisks takes offline the entire storage pool that contains this MDisk. To reduce the effects of an MDisk failure, consider reducing the number of back-end storage systems per storage pool, and increasing the number of storage pools and reducing their size. This configuration limits the maximum performance of the pool (fewer back-end systems to share the load), increases storage management effort, can lead to less efficient storage capacity consumption. It also might be subject to limitation by system configuration maximums.

  For internal storage, you might be able to define arrays in such a way that they can survive an expansion drive failure. However, such array design might restrict performance, available array sizes and RAID levels, depending on your hardware configuration.

- Consistency

  Strive to achieve consistent availability levels of all system building blocks.

3.3 Connectivity planning

IBM Storwize V7000 offers a wide range of connectivity options, both to back-end storage and to hosts. They include Fibre Channel (FC) SAN (8 and 16 Gbps, including direct attachment for some purposes), iSCSI (with 1 Gbps and 10 Gbps ports, depending on hardware configuration), and FCoE connectivity on 10 Gbps ports.

Storwize V7000 supports SAN routing technologies between Storwize V7000 and storage systems if the routing stays entirely within Fibre Channel connectivity and does not use other transport technologies, such as IP. However, SAN routing technologies (including FCIP links) are supported for connections between the Storwize V7000 and hosts. The use of long-distance FCIP connections might degrade the storage performance for any servers that are attached through this technology.

Table 3-1 lists the fabric type that can be used for communicating between hosts, nodes, and back-end storage systems. All fabric types can be used at the same time.

Table 3-1 Storwize V7000 communication options

Communication type	Host to Storwize V7000	Storwize V7000 to storage	Storwize V7000 to Storwize V7000
Fibre Channel (FC) SAN	Yes	Yes	Yes
iSCSI (1 GbE or 10 GbE)	Yes	Yes	No
iSCSI (25 GbE) (RDMA iSER)	Yes	No	Yes
FCoE (10 GbE)	Yes	Yes	Yes

When you plan deployment of Storwize V7000, identify networking technologies that you will use.

Note: With Spectrum Virtualize V8.1.1.1 and later, RDMA (iSER) is supported by 25 Gb Ethernet iSCSI adapter cards with V7000 Gen2+ only. For more information, see 3.7.4, “iSCSI Extensions for RDMA (iSER)” on page 62.

3.4 Physical planning

You must consider several key factors when you are planning the physical site of a Storwize V7000 installation. The physical site must have the following characteristics:

- Meets power, cooling, and location requirements of the Storwize V7000 nodes.
- Has two separate power sources.
- Sufficient rack space exists for the installation of controller and disk expansion enclosures.
- Has sufficient maximum power rating of the rack. Plan your rack placement carefully to not exceed maximum power rating of the rack. For more information about the power and environmental requirements, see this website.

Your Storwize V7000 2076-524 and Storwize V7000 2076-624 order includes a printed copy of the IBM Storwize V7000 Gen2 and Gen2+ Quick Installation Guide, which also provides information about environmental and power requirements.

3.4.1 Cabling

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

- Power
- Ethernet
- SAS
- iSCSI or Fibre Channel over Ethernet (FCoE) connections
- Switch ports (FC, Ethernet, and FCoE)

Distribute your disk expansion enclosures evenly between control enclosures, nodes within control enclosures, and SAS channels within nodes. For more information, search for “SAS cabling guidelines” at this IBM Knowledge Center page.

When planning SAN cabling make sure that your physical topology allows you to observe zoning rules and recommendations.

If the data center provides more than one power source, make sure that you use that capacity when planning power cabling for your system.

3.5 Planning IP connectivity

System management is performed through an embedded graphical user interface (GUI) that is running on the nodes. To access the management GUI, direct a web browser to the system management IP address.

Storwize V7000 Gen2 and Storwize V7000 Gen2+ nodes have a new feature called a *Technician port*. It is an Ethernet marked with a T. All initial configuration for each node is performed by using the Technician port. The port runs a Dynamic Host Configuration Protocol (DHCP) service so that any notebook or computer connected to the port is automatically assigned an IP address.

After the cluster configuration is complete, the Technician port automatically routes the connected user directly to the service GUI.

> **Note:** The default IP address for the Technician port is `192.168.0.1`. If the Technician port is connected to a switch, it is disabled and an error is logged.

Each Storwize V7000 node requires one Ethernet cable to connect it to an Ethernet switch or hub. The cable must be connected to port 1. A 10/100/1000 Mb Ethernet connection is supported on the port. Both Internet Protocol Version 4 (IPv4) and Internet Protocol Version 6 (IPv6) are supported.

> **Note:** For increased availability, an optional second Ethernet connection is supported for each Storwize V7000 node.

Ethernet port 1 on every node must be connected to the same set of subnets. The same rule applies to Ethernet port 2 if it is used. However, the subnets available for Ethernet port 1 do not have to be the same as configured for interfaces on Ethernet port 2.

Each Storwize V7000 cluster has a Cluster Management IP address, in addition to a Service IP address for each node in the cluster (see Example 3-1).

Example 3-1 System addressing example

```
management IP add. 10.11.12.120
node 1 service IP add. 10.11.12.121
node 2 service IP add. 10.11.12.122
```

Each node in an Storwize V7000 clustered system must have at least one Ethernet connection. Both IPv4 and IPv6 addresses are supported. Storwize V7000 can operate with either Internet Protocol or with both internet protocols concurrently.

For configuration and management, you must allocate an IP address to the system, which is referred to as the management IP address. For more fault tolerance, you can also configure a second IP address for the second Ethernet port on the node. The addresses must be fixed addresses. If both IPv4 and IPv6 are operating concurrently, an address is required for each protocol.

> **Note:** The management IP address cannot be the same as any of the defined service IPs.

Support for iSCSI enables one extra IPv4 address, IPv6 address, or both for each Ethernet port on every node. These IP addresses are independent of the system’s management and service IP addresses.

If you configure management IP on both Ethernet ports, choose one of the IP addresses to connect to GUI or CLI. The system cannot automatically fail over the management IP address to a different port. If one management IP address is unavailable, use an IP address on the alternative network.

Clients might be able to use the intelligence in domain name servers (DNSs) to provide partial failover. Figure 3-1 shows a simple IP addressing scheme that uses a single subnet for iSCSI and management.

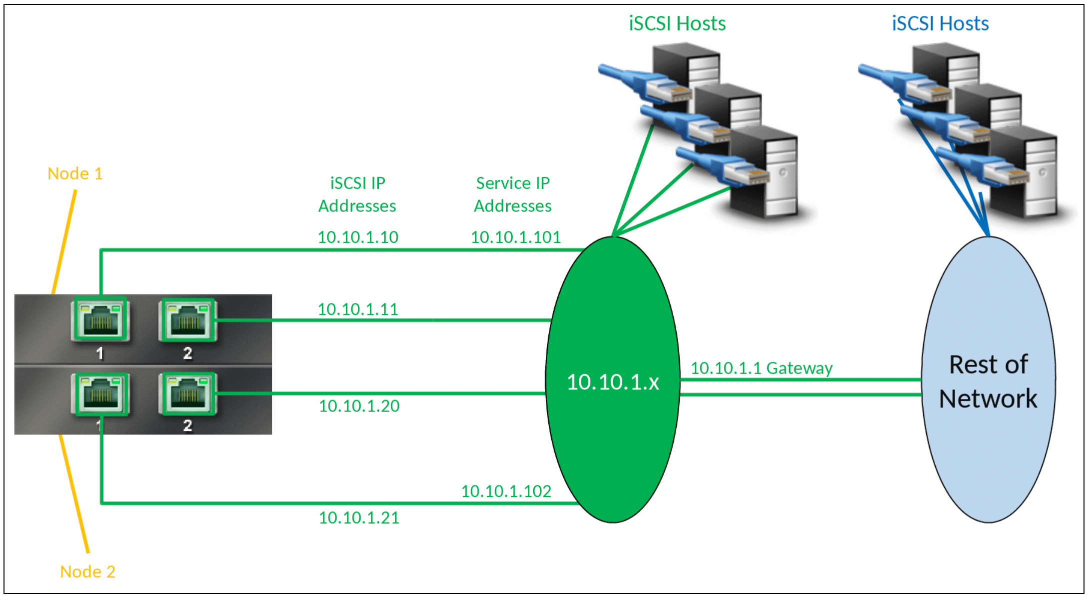


Figure 3-1 IP addressing scheme: Use of a single subnet

3.5.1 Firewall planning

After you have your IP network planned, identify network flows that are required for the correct functioning of the environment. The list must specify source IP address, destination IP addresses, and required protocols or ports for each flow. Present the list to the firewall administrators and request the set up of appropriate firewall rules.

For a list of mandatory and optional network flows that are required for operating IBM SAN Volume Controller, search for “TCP/IP requirements for the system” at IBM Knowledge Center.

3.6 SAN configuration planning

Storwize V7000 cluster can be configured with a minimum of two (and up to eight) Storwize V7000 nodes. These nodes can use SAN fabric to communicate with back-end storage subsystems and hosts.

3.6.1 Physical topology

The switch configuration in an Storwize V7000 fabric must comply with the switch manufacturer’s configuration rules, which can impose restrictions on the switch configuration. For example, a switch manufacturer might limit the number of supported switches in a SAN. Operation outside of the switch manufacturer’s rules is not supported.

The hardware compatible with V8.2.1 supports 8 Gbps and 16 Gbps FC fabric, depending on the hardware platform and on the switch to which the Storwize V7000 is connected. In an environment where you have a fabric with multiple-speed switches, the preferred practice is to connect the Storwize V7000 and back-end storage systems to the switch operating at the highest speed.

You can use the **lsfabric** command to generate a report that displays the connectivity between nodes and other controllers and hosts. This report is helpful for diagnosing SAN problems.

Storwize V7000 control enclosure contains a pair of nodes (I/O Groups). Odd number of nodes in a cluster is a valid configuration only if a node fails or is removed from the configuration. In such case, the remaining node operates in a degraded mode.

For clustered systems avoid, if possible, communication between nodes that route across ISLs. Connect all nodes to the same Fibre Channel or FCF switches.

Direct connection of the system Fibre Channel ports without the use of a Fibre Channel switch is supported. Such direct connections between the system nodes might be useful in small configurations where there is no Fibre Channel switch. It can also be used to connect nodes in the same I/O group to provide a dedicated connection for mirroring the fast write cache data.

No more than three ISL hops are permitted among nodes that are in the same system though 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 support center.

Avoid ISL on the path between nodes and back-end storage. If possible, connect all storage systems to the same Fibre Channel or FCF switches as the nodes. One ISL hop between the nodes and the storage systems is permitted. If your configuration requires more than one ISL, contact your support center.

In larger configurations, it is common to have ISLs between host systems and the nodes.

To verify the supported connection speed for FC links to the Storwize V7000, use the hIBM System Storage Interoperation Center (SSIC) website.

For more information about the configuration of Enhanced Stretched Cluster or HyperSwap setup, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

3.6.2 Zoning

In Storwize V7000 deployments, the SAN fabric must have two distinct zone classes:

- Host zones: Allows communication between Storwize V7000Storwize V7000 and hosts.
- Storage zone: Allows communication between Storwize V7000 and back-end storage.

In clustered configurations, a third zone is required that allows communication between storage system nodes (intra-cluster traffic).

Figure 3-2 shows the Storwize V7000 zoning classes.

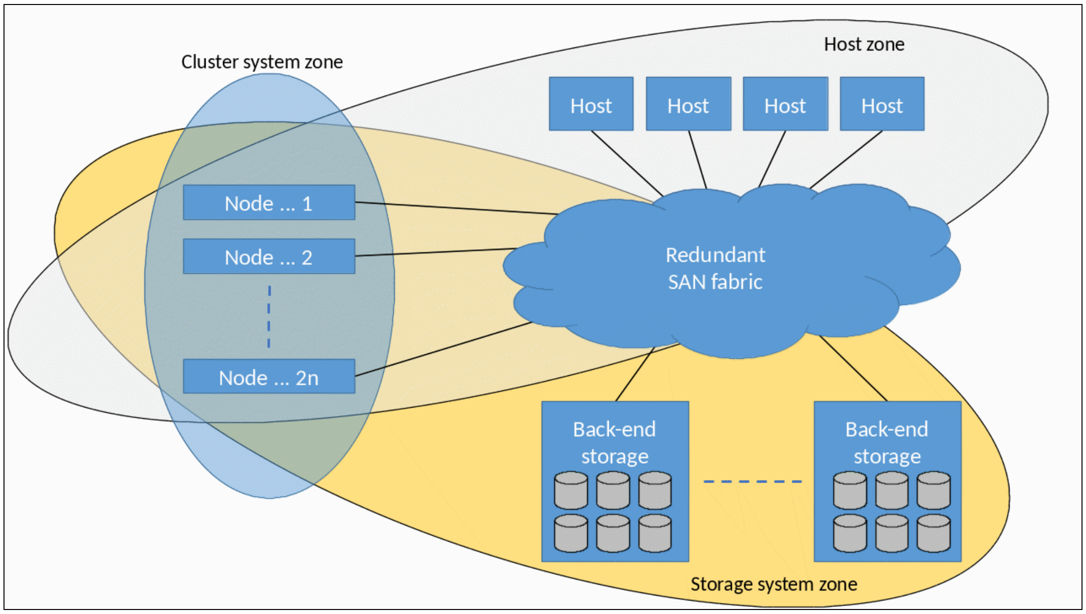


Figure 3-2 Storwize V7000 zoning classes

The fundamental rules of Storwize V7000 zoning are described next. However, also review the latest zoning guidelines and requirements when designing zoning for the planned solution by searching for "SAN configuration and zoning rules summary" at IBM Knowledge Center.

> **Note:** Configurations that use Metro Mirror, Global Mirror, N_Port ID Virtualization, or long-distance links have extra zoning requirements. Do not follow only the general zoning rules if you plan to use any of these.
>
> The FCoE fabric uses the same set of zoning rules as the Fibre Channel fabric.

3.6.3 Storwize V7000 cluster system zone

The Storwize V7000 cluster system zone is required only if you deploy solution with more than one control enclosure. The purpose of cluster system zone is to enable traffic between all Storwize V7000 nodes within the clustered system. This traffic consists of heartbeats, cache synchronization, and other data that nodes must exchange to maintain a healthy cluster state.

Each Storwize V7000 port must be zoned so that it can be used for internode communications. A system node cannot have more than 16 paths to another node in the same system.

Mixed port speeds are not possible for intracluster communication. All node ports within a clustered system must be running at the same speed.

Storwize V7000 supports the use of mixed fabrics for communication between nodes. The 10 GbE FCoE ports of one Storwize V7000 can be zoned to the FC ports of another node that is part of the same clustered system.

Note: You can use more than four fabric ports per node to improve peak load I/O performance. However, if a node receives more than 16 logins from another node, it causes node error 860. To avoid this error, you must use zoning, port masking, or combination of the two. For more information, see 3.6.7, “Port designation recommendations” on page 58, 3.6.8, “Port masking” on page 59, and the IBM Storwize V7000 documentation.

3.6.4 Back-end storage zones

Create one Storwize V7000 *storage zone* for each back-end storage subsystem that is virtualized by the Storwize V7000.

A storage controller can present LUNs to the Storwize V7000 (as MDisks) and to other hosts in the SAN. However, if this is the case, it is better to allocate different ports on the back-end storage for communication with Storwize V7000 and for hosts traffic.

All nodes in a system must connect to the same set of storage system ports on each device. A system that contains any two nodes that cannot connect to the same set of storage-system ports is considered degraded. In this situation, a system error is logged that requires a repair action. This rule can have important effects on a storage system. For example, an IBM DS4000® series controller can have exclusion rules that determine to which host bus adapter (HBA) worldwide node names (WWNNs) to which a storage partition can be mapped.

In the example that is shown in Figure 3-3, all back-end storage ports are connected to all Storwize V7000 ports. Therefore, each port on back-end storage is zoned to all ports of the Storwize V7000 that are connected to the same fabric. For clarity, the figure shows only paths for one back-end storage port per controller: Port P1 in Fabric 0 and port P4 in Fabric 1.

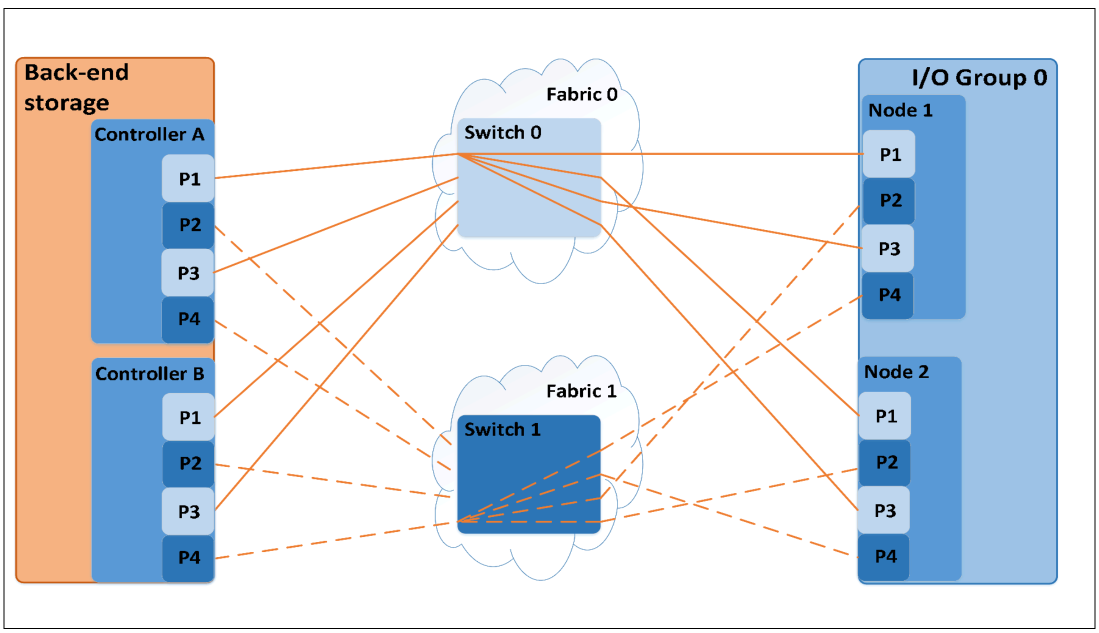


Figure 3-3 Example of back-end storage subsystem zoning

There might be particular zoning rules governing attachment of specific back-end storage systems. Review guidelines at this website to verify whether you must consider other policies when planning zoning for your back-end systems.

3.6.5 Host zones

Host must be zoned to the I/O Group to access volumes that are presented by this I/O Group.

The preferred zoning policy is to create a separate zone for each host 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 defined in the system, this host zoning scheme is mandatory.

If you plan to use NPIV, see the other host zoning requirements at that available at this website.

When a dual-core SAN design is used, it is a requirement that no internode communications use the ISL link. When you create host zones in this type of configuration, ensure that each system port in the host zone is attached to the same Fibre Channel switch.

Consider the following rules for zoning hosts with the Storwize V7000:

- HBA to Storwize V7000 port zones

  Place each host's HBA in a separate zone with exactly one port from each node in each I/O group that the host accesses.

  It is not prohibited to zone host's HBA to one port from every node in the cluster, but it does reduce the maximum number of hosts that can be attached to the system.

> **Number of paths:** For $n + 1$ redundancy, use the following number of paths:
>
> - With two HBA ports, zone HBA ports to Storwize V7000 ports 1:2 for a total of four paths.
> - With four HBA ports, zone HBA ports to Storwize V7000 ports 1:1 for a total of four paths.
> - Optional (n+2 redundancy): With four HBA ports, zone HBA ports to Storwize V7000 ports 1:2 for a total of eight paths.
>
> Here, the term *HBA port* is used to describe the SCSI initiator and *Storwize V7000 port* to describe the SCSI target.

- Maximum host paths per logical unit (LU)

  For any volume, the number of paths through the SAN from the Storwize V7000 nodes to a host must not exceed eight. For most configurations, four paths to an I/O Group are sufficient.

> **Important:** The maximum number of host paths per LUN must not exceed eight.

Another way to control the number of paths between hosts and the Storwize V7000 is to use *port mask*. The port mask is an optional parameter of the **`mkhost`** and **`chhost`** commands. The port-mask configuration has no effect on iSCSI connections.

For each login between a host Fibre Channel port and node Fibre Channel port, the node examines the port mask for the associated host object. It then determines whether access is allowed (port mask bit for specific port is set) or denied (port mask bit is cleared).

If access is denied, the node responds to SCSI commands as though the HBA WWPN is unknown. The port mask is 64 bits. Valid mask values range from all zeroes (no ports enabled) to all ones (all ports enabled). For example, a mask of 0011 enables port 1 and port 2. The default value is all 1s.

- Balanced 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 Storwize V7000 ports.

- Balanced host load across Storwize V7000 ports

  To obtain the best overall performance of the subsystem and to prevent overloading, the load of each Storwize V7000 port should be equal. Assuming a similar load is generated by each host, you can achieve this balance by zoning approximately the same number of host ports to each Storwize V7000 port.

Figure 3-4 on page 56 shows an example of a balanced zoning configuration that was created by completing the following steps:

1. Divide ports on the I/O Group into two disjoint sets, such that each set contains two ports from each I/O Group node, each connected to a different fabric.

   For consistency, use the same port number on each I/O Group node. The example that is shown in Figure 3-4 on page 56 assigns ports 1 and 4 to one port set, and ports 2 and 3 to the second set.

   Because the I/O Group nodes have four FC ports each, two port sets are created.

2. Divide hosts attached to the I/O Group into two equally numerous groups.

   In general, for I/O Group nodes with more than four ports, divide the hosts into as many groups as you created sets in step 1.

3. Map each host group to exactly one port set.

4. Zone all hosts from each group to the corresponding set of I/O Group node ports.

   The host connections in the example on Figure 3-4 on page 56 are defined in the following manner:

   – Hosts in group one are always zoned to ports 1 and 4 on both nodes.
   – Hosts in group two are always zoned to ports 2 and 3 on both nodes of the I/O Group.

Tip: Create an alias for the I/O Group port set. This step makes it easier to correctly zone hosts to the correct set of I/O Group ports. It also makes host group membership visible in the FC switch configuration.

The use of this schema provides four paths to one I/O Group for each host, and helps to maintain an equal distribution of host connections on Storwize V7000 ports.

Tip: To maximize performance from the host point of view, distribute volumes that are mapped to each host between both I/O Group nodes.

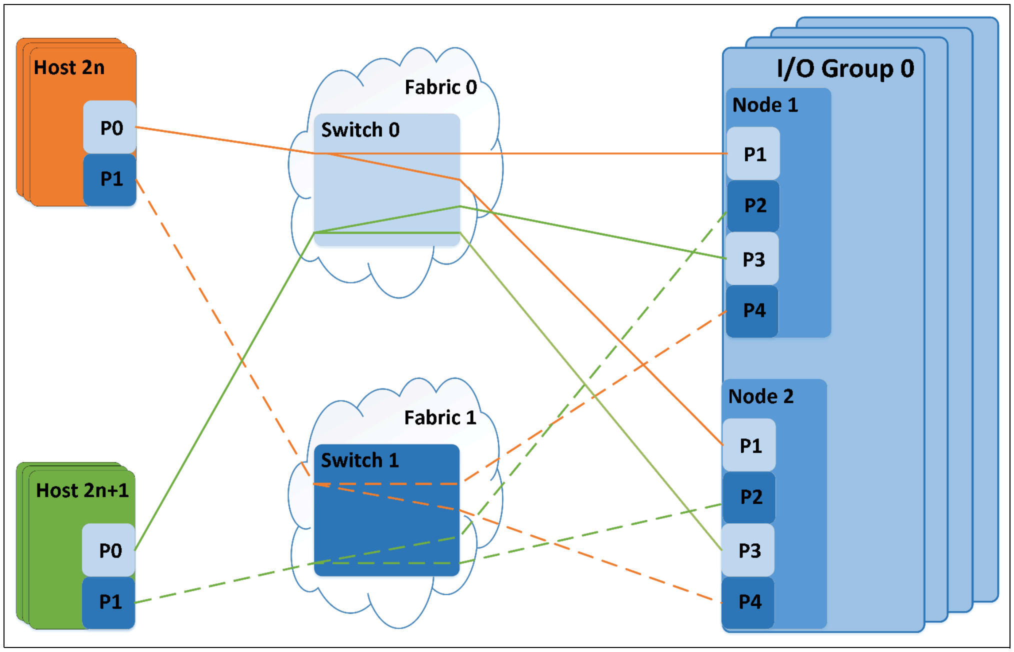


Figure 3-4 Overview of four-path host zoning

When possible, use the minimum number of paths that are necessary to achieve a sufficient level of redundancy. For the Storwize V7000 environment, no more than four paths per I/O Group are required to accomplish this layout.

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 required to access all volumes mapped to the host.

For hosts that use four HBAs/ports with eight connections to an I/O Group, use the zoning schema that is shown in Figure 3-5 on page 57. You can combine this schema with the previous four-path zoning schema.

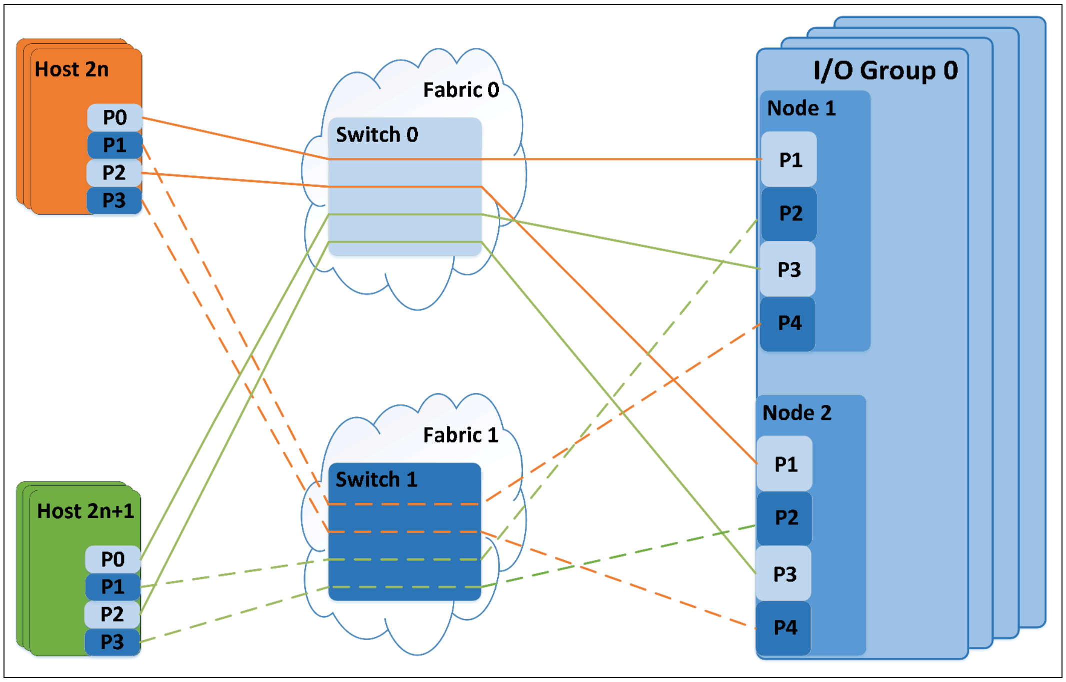


Figure 3-5 Overview of eight-path host zoning

For more information, see Chapter 8, “Hosts” on page 317.

3.6.6 Zoning considerations for Metro Mirror and Global Mirror

The SAN configurations that use intercluster Metro Mirror and Global Mirror relationships require the following other switch zoning considerations:

- Review the latest requirements and recommendations at this website.
- 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.

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

Important: Be careful when you perform the zoning so that ports that are dedicated for intra-cluster communication are *not* used for Host/Storage traffic in the 8-port and 12-port configurations.

The use of 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.

Important: If you zone two Fibre Channel ports on each node in the local system to two Fibre Channel ports on each node in the remote system, you can limit the effect of severe and abrupt overload of the intercluster link on system operations.

If you zone all node ports 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 Storwize V7000 nodes can perform local intracluster heartbeat communication. This situation can, in turn, result in the nodes experiencing lease expiry events.

In a lease expiry event, a node restarts to attempt to reestablish communication with the other nodes in the clustered system. If the leases for all nodes expire simultaneously, a loss of host access to volumes can occur during the restart events.

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

3.6.7 Port designation recommendations

The intracluster communication is used for mirroring write cache and metadata exchange between nodes, and is critical to the stable operation of the cluster. It is possible to upgrade nodes beyond the standard 4 FC port configuration. Such upgrade provides an opportunity to dedicate ports to local node traffic, separating them from other cluster traffic on the remaining ports. This configuration provides a level of protection against malfunctioning devices and workload spikes that might otherwise impact the intracluster traffic.

Additionally, a benefit is realized in isolating remote replication traffic to dedicated ports, and ensuring that any problems that affect the cluster-to-cluster interconnect do not affect all ports on the local cluster.

IBM suggests the port designations for isolating both port-to-local and port-to-remote canister traffic for Storwize V7000 Gen2 and Gen2+ canisters, as shown in Figure 3-6.

Slot/Port	Port#	SAN	4-Port Canister	8-Port Canister
S2P1	1	A	Host/Storage/Inter-node	Host/Storage
S2P2	2	B	Host/Storage/Inter-node	Host/Storage
S2P3	3	A	Host/Storage/Replication*	Inter-node
S2P4	4	B	Host/Storage/Replication*	Host/Storage/Replication**
S3P1	5	A	--	Host/Storage
S3P2	6	B	--	Host/Storage
S3P3	7	A	--	Host/Storage/Replication**
S3P4	8	B	--	Inter-node
localportmask			With Rep 0011/ No Rep 1111	10000100
remoteportmask			With Rep 1100	01001000
* Inter-node if no replication planned				
** Use for Host/Storage in case no replication is in place.				

Figure 3-6 Port masking configuration on Storwize V7000

Note: With 12 or more ports per node, four ports should be dedicated for node-to-node traffic. Doing so is especially important when high write data rates are expected because all writes are mirrored between I/O Group nodes over these ports.

The port designation patterns that are shown in 3.6, “SAN configuration planning” on page 50 provide the required traffic isolation and simplify migrations to configurations with greater number of ports. More complicated port mapping configurations that spread the port traffic across the adapters are supported and can be considered. However, these approaches do not appreciably increase availability of the solution.

Alternative port mappings that spread traffic across HBAs might allow adapters to come back online following 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 presented approach takes all of these considerations into account with a view that increased complexity can lead to migration challenges in the future, and simpler approach is usually better.

3.6.8 Port masking

You can use a port mask to control the node target ports that a host can access. By using local FC port masking, you can set which ports can be used for node-to-node/intracluster communication. By using remote FC port masking, you can set which ports can be used for replication communication. Port masking, combined with zoning, enables you to dedicate ports to particular type of traffic. Setting up Fibre Channel port masks is particularly useful when you have more than four Fibre Channel ports on any node in the system, as it saves setting up many SAN zones.

Two Fibre Channel port masks are on a system. The local port mask controls connectivity to other nodes in the same system; the partner port mask controls connectivity to nodes in remote, partnered systems. By default, all ports are enabled for both local and partner connectivity.

The port masks apply to all nodes on a system; a different port mask cannot be set on nodes in the same system. You do not have to have the same port mask on partnered systems.

A mixed traffic of host, back-end, intracluster, and replication might cause congestion and buffer to buffer credit exhaustion. This type of traffic can result in heavy degradation of performance in your storage environment.

Fibre Channel I/O ports are logical ports, which can be on Fibre Channel platform ports or FCoE platform ports.

The port mask is a 64-bit field that applies to all nodes in the cluster. In the local FC port masking, you can set a port to be dedicated to node-to-node/intracluster traffic by setting a 1 to that port. Remote FC port masking allows you to set which ports can be used for replication traffic by setting 1 to that port.

If a port has a 0 in the specific mask, it means no traffic of that type is allowed. Therefore, in a local FC port map, a 0 means no node-to-node traffic occurs. A 0 on the remote FC port masking means that no replication traffic occurs on that port. Therefore, if a port has a 0 on local and remote FC port masking, only host and back-end storage traffic is allowed on it.

Setting port mask by using the CLI and GUI

The command to apply a local FC port mask on CLI is **chsystem -localfcportmask mask**. The command to apply a remote FC port mask is **chsystem -partnerfcportmask mask**.

If you are using the GUI, click **Settings** → **Network** → **Fibre Channel Ports**. Then, you can select the use of a port. Setting *none* means no node-to-node and no replication traffic is allowed, and only host and storage traffic is allowed. Setting `local` means only node-to-node traffic is allowed, and `remote` means that only replication traffic is allowed. Figure 3-7 shows an example of setting a port mask on port 1 to "Local".

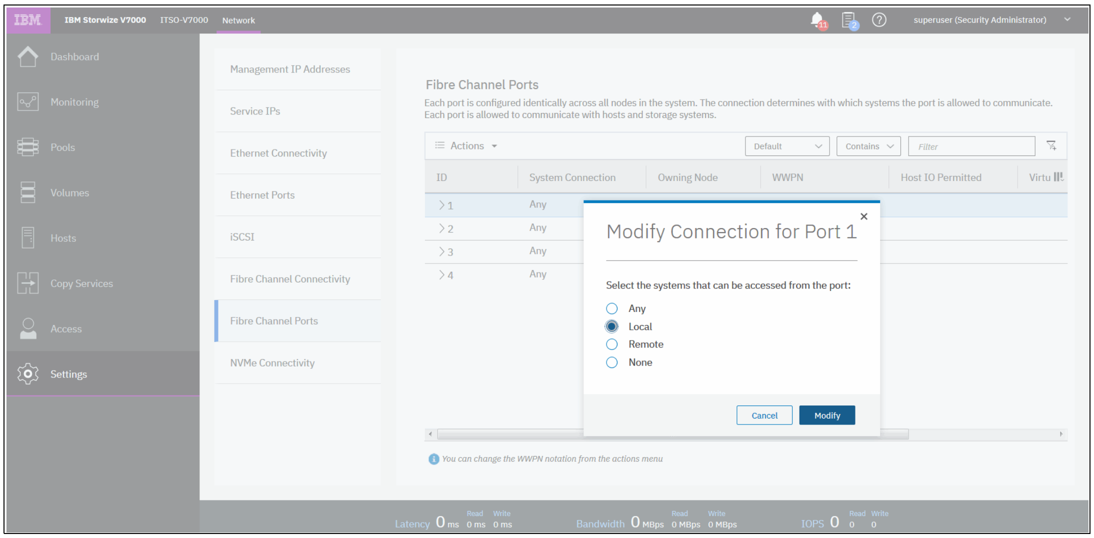


Figure 3-7 Fibre Channel port mask setting from GUI

3.7 iSCSI configuration planning

Each Storwize V7000 node is equipped with up to three onboard Ethernet network interface cards (NICs), which can operate at a link speed of 10 Mbps, 100 Mbps, or 1000 Mbps. All NICs can be used to carry iSCSI traffic. For optimal performance, use a 1 Gbps links between Storwize V7000 and iSCSI-attached hosts when the Storwize V7000 node's onboard NICs are used.

An optional 10 Gbps 2-port Ethernet adapter is available. One such adapter can be installed in each node and can be used to add 10 Gb iSCSI/FCoE connectivity to the Storwize V7000.

Figure 3-8 on page 61 shows an overview of the iSCSI implementation in the Storwize V7000.

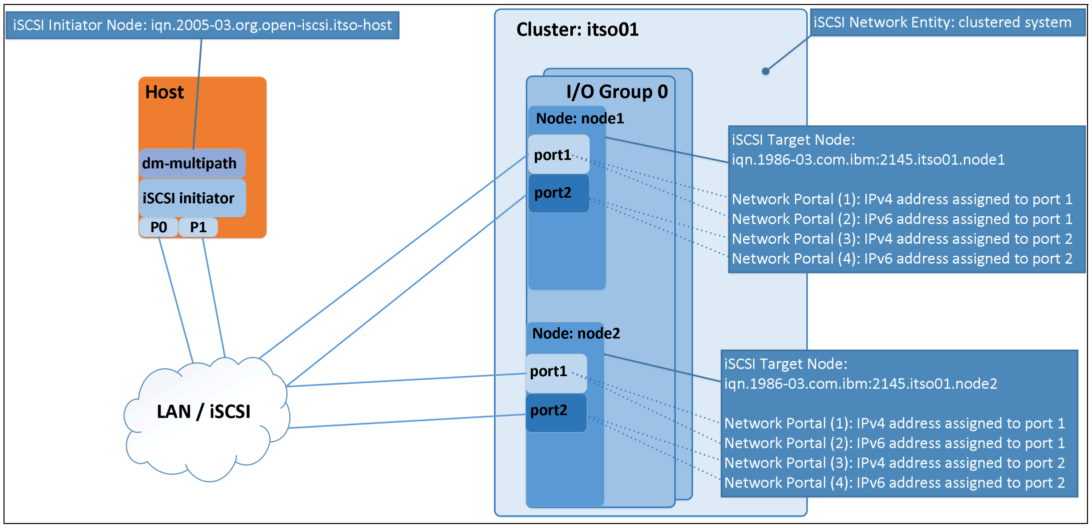


Figure 3-8   Storwize V7000 iSCSI overview

Both onboard Ethernet ports of a Storwize V7000 node can be configured for iSCSI. For each instance of an iSCSI target node (that is, each Storwize V7000 node), you can define two IPv4 and two IPv6 addresses or iSCSI network portals:

- If the optional 10 Gbps Ethernet feature is installed, you can use them for iSCSI traffic.
- Generally, enable jumbo frames in your iSCSI storage network.
- iSCSI IP addresses can be configured for one or more nodes.
- Decide whether you implement authentication for the host to Storwize V7000 iSCSI communication. The Storwize V7000 supports the Challenge Handshake Authentication Protocol (CHAP) authentication methods for iSCSI.

3.7.1 iSCSI protocol

The iSCSI connectivity is a software feature that is provided by the Storwize V7000 code. The iSCSI protocol is a block-level protocol that encapsulates SCSI commands into Transmission Control Protocol/Internet Protocol (TCP/IP) packets. Therefore, iSCSI uses IP network rather than requiring the Fibre Channel infrastructure. The iSCSI standard is defined by Request For Comments (RFC) 3720.

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

3.7.2 Topology and IP addressing

For more information about examples of topology and addressing schemes that can be used for iSCSI connectivity, see 3.5, “Planning IP connectivity” on page 48.

If you plan to use node’s 1 Gbps Ethernet ports for iSCSI host attachment, dedicate Ethernet port one for the Storwize V7000 management and port two for iSCSI use. This way, port two can be connected to a separate network segment or virtual local area network (VLAN) dedicated to iSCSI traffic.

Note: Ethernet link aggregation (port trunking) or *channel bonding* for the Storwize V7000 nodes' Ethernet ports is not supported for the 1 Gbps ports.

3.7.3 General recommendation

This section covers general preferences that are related to iSCSI.

Planning for host attachments

An iSCSI client, which is known as an iSCSI *initiator,* sends SCSI commands over an IP network to an iSCSI target. A single iSCSI initiator or iSCSI target is called an *iSCSI node*.

You can use the following types of iSCSI initiators in host systems:

- Software initiator: Available for most operating systems (OS), including AIX, Linux, and Windows.
- Hardware initiator: Implemented as a network adapter with an integrated iSCSI processing unit, which is also known as an *iSCSI HBA*.

Make sure that iSCSI initiators, targets, or both that you plan to use are supported. Use the following websites for reference:

- IBM Storwize V7000 V8.1 Support page
- IBM Knowledge Center for IBM Storwize V7000
- hIBM System Storage Interoperation Center (SSIC)

iSCSI qualified name

A Storwize V7000 cluster can provide up to eight iSCSI targets, one per node. Each Storwize V7000 node has its own IQN, which, by default, is in the following form:

```
iqn.1986-03.com.ibm:2145.<clustername>.<nodename>
```

An alias string can also be associated with an iSCSI node. The alias enables an organization to associate a string with the iSCSI name. However, the alias string is not a substitute for the iSCSI name.

Note: The cluster name and node name form part of the IQN. Changing any of them might require reconfiguration of all iSCSI nodes that communicate with the Storwize V7000.

3.7.4 iSCSI Extensions for RDMA (iSER)

IBM Spectrum Virtualize V8.2.1 introduced support for iSER host attachment for the Storwize V7000 Gen2+ by using RoCE or iWARP transport protocol, depending on HBA hardware and host platforms that provide the following components:

- A fully Ethernet based infrastructure (no Fibre Channel) in the Data center
- IBM SAN Volume Controller or Storwize inter-node communication
- HyperSwap (SVC/Storwize)

The system supports node-to-node connections that use Ethernet protocols that support remote direct memory access (RDMA) technology, such as RDMA over Converged Ethernet (RoCE) or iWARP.

To use these protocols, the system requires that a 25 Gbps Ethernet adapter is installed on each node and dedicated RDMA-based ports are only configured for node-to-node communication.

RDMA technologies, like RoCE and iWARP, enable the 25 Gbps Ethernet adapter to transfer data directly between nodes, bypassing CPU, and caches, making transfers faster.

RDMA technologies provide faster connection and processing time than traditional iSCSI connections and are a lower-cost option than Fibre Channel fabrics.

Prerequisites

The following prerequisites must be met for all RDMA-based connections between nodes:

- Installation of the node hardware is complete.
- A 25 Gbps Ethernet adapter is installed on each node.
- Ethernet cables between each node are connected correctly.
- Protocols on the source and destination adapters are the same.
- Local and remote IP addresses can be reached.
- Each IP address is unique.
- The negotiated speeds on the local and remote adapters are the same.
- The local and remote port virtual LAN identifiers are the same.
- A minimum of two dedicated 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.
- A maximum of four ports per node are allowed for node-to-node RDMA connections. Configure a local port mask that limits node-to-node connections to maximum of four ports per node. For more information, see this IBM Knowledge Center web page.

3.7.5 iSCSI back-end storage attachment

IBM Spectrum Virtualize V7.7 introduced support for external storage controllers that are attached through iSCSI.

For more information about back-end storage supported for iSCSI connectivity, see these websites:

- IBM Support Information for Storwize V7000
- IBM System Storage Interoperation Center (SSIC)

3.8 Back-end storage subsystem configuration

Back-end storage subsystem configuration must be planned for all storage controllers that are attached to the Storwize V7000.

For more information about supported storage subsystems, see these websites:

- IBM Support Information for Storwize V7000
- IBM System Storage Interoperation Center (SSIC)

Apply the following general guidelines for back-end storage subsystem configuration planning:

- In the SAN, storage controllers that are used by the Storwize V7000 clustered system must be connected through SAN switches. Direct connection between the Storwize V7000 and the storage controller is not supported.
- Enhanced Stretched Cluster configurations have additional requirements and configuration guidelines. For more information about performance and preferred practices for the Storwize V7000, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

MDisks within storage pools: Software versions 6.1 and later provide for better load distribution across paths within storage pools.

In previous code levels, the path to MDisk assignment was made in a round-robin fashion across all MDisks that are configured to the clustered system. With that method, no attention is paid to how MDisks within storage pools are distributed across paths. Therefore, it was possible and even likely that certain paths were more heavily loaded than others.

Starting with software version 6.1, the code contains logic that considers which MDisks are provided by which back-end storage systems. Therefore, the code more effectively distributes active paths based on the storage controller ports that are available.

The `Detect MDisk` commands must be run following the creation or modification (addition of or removal of MDisk) of storage pools for paths to be redistributed.

If your back-end storage system does not support the Storwize V7000 round-robin algorithm, 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.

In general, configure disk subsystems as though Storwize V7000 was not used. However, there might be specific requirements or limitations as to the features usable in the specific back-end storage system when it is attached to Storwize V7000. Review the appropriate section of documentation to verify that your back-end storage is supported and to check for any special requirements at this website.

Observe the following general rules:

- Disk drives

  Exercise caution with the use of large hard disk drives so that you do not have too few spindles to handle the load.
- Array sizes:
  - Storwize V7000 does not queue more than 60 I/O operations per MDisk. Therefore, make sure that the MDisks presented to Storwize V7000 can handle about this many requests, which corresponds to about 8 HDDs. If your array can handle higher load, split it in several LUNs of equal size to better match back-end storage capabilities with the load that Storwize V7000 can generate.

    For more information about back-end storage LUN presentation to Storwize V7000, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.
  - Since version 7.3, the system uses autobalancing to restripe volume extents evenly across all MDisks in the storage pools.

– The cluster can be connected to a maximum of 1024 WWNNs. The general practice is that:

- EMC DMX/SYMM, all HDS, and SUN/HP HDS clones use one WWNN per port. Each port appears as a separate controller to the Storwize V7000.
- IBM, EMC CLARiiON, and HP use one WWNN per subsystem. Each port appears as a part of a subsystem with multiple ports, up to a maximum of 16 ports (WWPNs) per WWNN.

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

3.9 Internal storage configuration

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

Use DRAID 6 as the default RAID level for internal arrays. DRAID 6 outperforms RAID 5. Workloads that used to be placed on RAID 10 arrays now are usually on SSD drives.

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

3.10 Storage pool configuration

The storage pool is at the center of the many-to-many relationship between the MDisks and the volumes. It acts as a container of physical disk capacity from which chunks of MDisk space, known as *extents,* are allocated to form volumes presented to hosts.

MDisks in the Storwize V7000 are LUNs that are assigned from the back-end storage subsystems to the Storwize V7000. There are two classes of MDisks: managed and unmanaged. An unmanaged MDisk is a LUN that is presented to IBM SAN Volume Controller by back-end storage, but is not assigned to any storage pool. A managed MDisk is an MDisk that is assigned to a storage pool. An MDisk can be assigned only to a single storage pool.

Storwize V7000 clustered system must have exclusive access to every LUN (MDisk) it is using. Any specific LUN cannot be presented to more than one Storwize V7000 cluster. Also, presenting the same LUN to a Storwize V7000 and a host is not allowed.

One of basic storage pool parameters is the extent size. All MDisks in the storage pool have the same extent size, and all volumes that are allocated from the storage pool inherit its extent size.

There are two implications of a storage pool extent size:

- Maximum volume, MDisk, and managed storage capacity depend on extent size (for more information, see this IBM Support web page). The bigger 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 created in the pool.
- Volume sizes must be a multiple of the extent size of the pool in which the volume is defined. Therefore, the smaller the extend size, the better control over volume size.

The Storwize V7000 supports extent sizes from 16 mebibytes (MiB) to 8192 MiB. The extent size is a property of the storage pool and is set when the storage pool is created.

The extent size of a storage pool cannot be changed. If you need to change extent size, the storage pool must be deleted and a new storage pool configured.

Table 3-2 lists all of the available extent sizes in a Storwize V7000 and the maximum managed storage capacity for each extent size.

Table 3-2 Extent size and total storage capacities per system

Extent size (MiB)	Total storage capacity manageable per system
16	64 tebibytes (TiB)
32	128 TiB
64	256 TiB
128	512 TiB
256	1 pebibytes (PiB)
512	2 PiB
1024	4 PiB
2048	8 PiB
4096	16 PiB
8192	32 PiB

When planning storage pool layout, consider the following aspects:

- Pool extent size:
  - Generally, use 128 MiB or 256 MiB. The IBM Storage Performance Council (SPC) benchmarks use a 256 MiB extent.
  - Pick the extent size and then use that size for all storage pools.
  - You cannot migrate volumes between storage pools with different extent sizes. However, you can use volume mirroring to create copies between storage pools with different extent sizes.
- Storage pool reliability, availability, and serviceability (RAS) considerations:
  - The number and size of storage pools affects system availability. The use of larger number of smaller pools reduces the failure domain in case one of pools goes offline. However, increased number of storage pools introduces management overhead, effects storage space use efficiency, and is subject to the configuration maximum limit.
  - An alternative approach is to create few large storage pools. All MDisks that constitute each of the pools should have the same performance characteristics.
  - The storage pool goes offline if an MDisk is unavailable, even if the MDisk has no data on it. Do not put MDisks into a storage pool until they are needed.
  - Put image mode volumes in a dedicated storage pool or pools.
- Storage pool performance considerations:
  - It might make sense to create multiple storage pools if you are attempting to isolate workloads to separate disk drives.

– Create storage pools out of MDisks with similar performance. This technique is the only way to ensure consistent performance characteristics of volumes created from the pool.

  This rule does not apply when you intentionally place MDisks from different storage tiers in the pool with the intent to use Easy Tier to dynamically manage workload placement on drives with appropriate performance characteristics.

3.10.1 The storage pool and Storwize V7000 cache relationship

The Storwize V7000 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 individual overloaded back-end storage system from filling system write cache and degrading the performance of the other storage pools. For more information, see Chapter 2, “System overview” on page 9.

Table 3-3 lists the limit of the write-cache data that can be used by a single storage pool.

Table 3-3 Limit 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 Storwize V7000 starts to limit incoming write I/Os for volumes that are created from the storage pool. That is, 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 Storwize V7000 is destaging data at a 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.

3.11 Volume configuration

When planning a volume, consider the required performance, availability, and cost of storage backing that volume. Volume characteristics are defined by the storage pool in which it is created.

Every volume is assigned to an I/O Group that defines which pair of Storwize V7000 nodes will service I/O requests to the volume.

Important: No fixed relationship exists between I/O Groups and storage pools.

Strive to distribute volumes evenly across available I/O Groups and nodes within the clustered system. Although volume characteristics depend on the storage pool from which it is created, any volume can be assigned to any node.

When you create a volume, it is associated with one node of an I/O Group, the *preferred access node*. By default, when you create a volume, it is associated with the I/O Group node by using a round-robin algorithm. However, you can manually specify the preferred access node if needed.

No matter how many paths are defined between the host and the volume, all I/O traffic is serviced by only one node (the preferred access node).

If you plan to use volume mirroring, put each copy in a different storage pool backed by different back-end storage subsystems for maximum availability. However, depending on your needs, it might be sufficient to use a different set of physical drives, a different storage controller, or a different back-end storage for each volume copy. Strive to place all volume copies in storage pools with similar performance characteristics. Otherwise, the volume performance as perceived by the host might be limited by the performance of the slowest storage pool.

3.11.1 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. Therefore, volume logical block address (LBA) *x* = MDisk LBA *x*. The capacity of image mode volumes is equal to the capacity of the MDisk from which it is created.

Image mode volumes are extremely useful tool in storage migration and when introducing IBM SAN Virtualization Controller to a storage environment.

3.11.2 Planning for thin-provisioned volumes

A thin-provisioned volume has a virtual capacity and a real capacity. Virtual capacity is the volume storage capacity that a host sees as available. Real capacity is the actual storage capacity that is allocated to a volume copy from a storage pool. Real capacity limits the amount of data that can be written to a thin-provisioned volume.

When planning use of thin-provisioned volumes, consider expected usage patterns for the volume. In particular, the actual size of the data and the rate of data change.

Thin-provisioned volumes require more I/Os because of directory accesses. For fully random access, and a workload with 70% reads and 30% writes, a thin-provisioned volume requires approximately one directory I/O for every user I/O. Additionally, thin-provisioned volumes require more processor processing, so the performance per I/O Group can also be reduced.

However, the directory is two-way write-back-cached (as with the Storwize V7000 fastwrite cache), so certain applications perform better.

Also, the ability to thin-provision volumes can be a worthwhile tool that allows hosts to see storage space significantly larger than what is allocated within the storage pool. Thin provisioning can also simplify storage allocation management. You can define virtual capacity of a thinly provisioned volume to an application based on the future requirements, but allocate real storage based on today's use.

The following types of thin-provisioned volumes are available:

- *Autoexpand volumes* allocate real capacity from a storage pool on demand, which minimizes required user intervention. However, a malfunctioning application can cause a volume to expand until its real capacity is equal to the virtual capacity, which potentially can starve other thin provisioned volumes in the pool.
- *Non-autoexpand volumes* have a fixed amount of assigned real capacity. In this case, the user must monitor the volume and assign more capacity when required. Although it prevents starving other thin provisioned volumes, it introduces a risk of an unplanned outage. A thin-provisioned volume goes offline if host tries to write more data than what can fit into the allocated real capacity.

The main risk that is associated with using thin-provisioned volumes is running out of real capacity in the storage volumes, pool, or both, and the resulting unplanned outage. Therefore, strict monitoring of the used capacity on all non-autoexpand volumes and monitoring of the free space in the storage pool is required.

When you configure a thin-provisioned volume, you can define a warning level attribute to generate a warning event when the used real capacity exceeds a specified amount or percentage of the total virtual capacity. You can also use the warning event to trigger other actions, such as taking low-priority applications offline or migrating data into other storage pools.

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 (error code 1865, event ID 060001). Access to the thin-provisioned volume is restored by increasing the real capacity of the volume, which might require increasing the size of the storage pool from which it is allocated. Until this time, the data is held in the Storwize V7000 cache. Although in principle this situation is not a data integrity or data loss issue, you must not rely on the Storwize V7000 cache as a backup storage mechanism.

Space is not allocated on a thin-provisioned volume if an incoming host write operation contains all zeros.

Important: Set and monitor a warning level on the used capacity so that you have adequate time to respond and provision more physical capacity.

Warnings must not be ignored by an administrator.

Consider the use of the autoexpand feature of the thin-provisioned volumes to reduce human intervention required to maintain access to thin-provisioned volumes.

When you create a thin-provisioned volume, you can choose the grain size for allocating space in 32 kibibytes (KiB), 64 KiB, 128 KiB, or 256 KiB chunks. The grain size that you select affects the maximum virtual capacity for the thin-provisioned volume. The default grain size is 256 KiB, which is the preferred option. If you select 32 KiB for the grain size, the volume size cannot exceed 260,000 gibibytes (GiB). The grain size cannot be changed after the thin-provisioned volume is created.

Generally, smaller grain sizes save space, but require more metadata access, which can adversely affect performance. 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. In this situation, ideally grain size is equal to the typical I/O size from the host.

A thin-provisioned volume feature that is called *zero detect* provides clients with the ability to reclaim unused allocated disk space (zeros) when they are converting a fully allocated volume to a thin-provisioned volume by using volume mirroring.

3.12 Host attachment planning

The typical FC host attachment to the Storwize V7000 is done through SAN fabric. However, the system allows direct attachment connectivity between its 8 Gb or 16 Gb Fibre Channel ports and host ports. No special configuration is required for host systems that are using this configuration. However, the maximum number of directly attached hosts is severely limited by the number of FC ports on Storwize V7000's nodes.

The Storwize V7000 imposes no particular limit on the distance between the Storwize V7000 nodes and host servers. However, for host attachment, the Storwize V7000 supports up to three ISL hops in the fabric. This capacity means that the server to the Storwize V7000 can be separated by up to five FC links, four of which can be 10 km long (6.2 miles) if long wave Small Form-factor Pluggables (SFPs) are used.

Figure 3-9 shows an example of a supported configuration with Storwize V7000 nodes using shortwave SFPs.

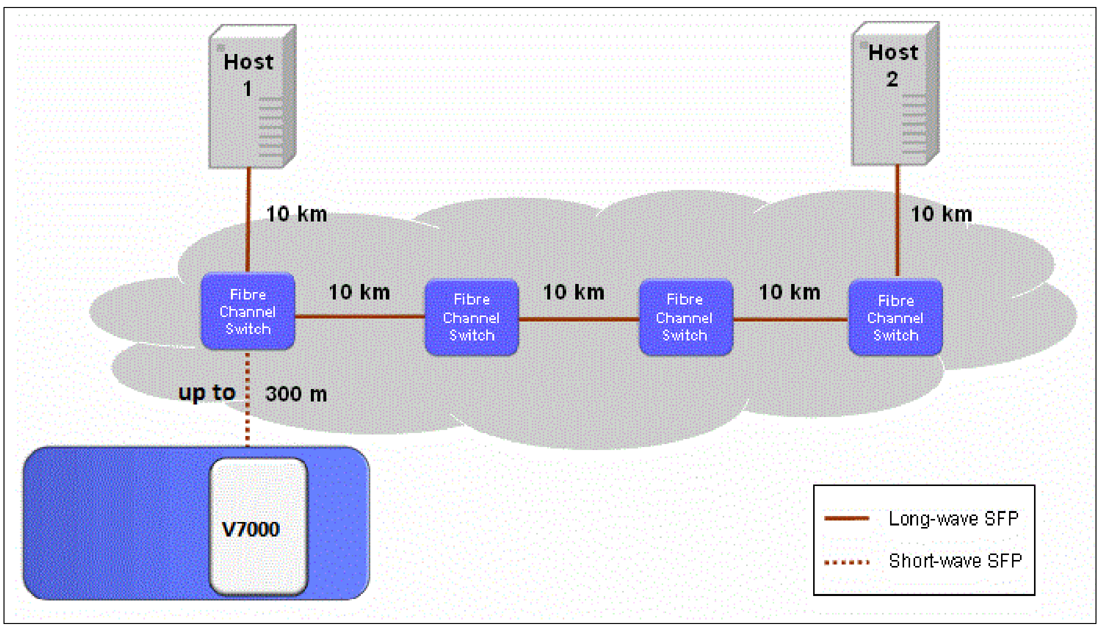


Figure 3-9 Example of host connectivity

In Figure 3-9, the optical distance between Storwize V7000 Node 1 and Host 2 is slightly over 40 km (24.85 miles).

To avoid latencies that lead to degraded performance, avoid ISL hops whenever possible. In an optimal setup, the servers connect to the same SAN switch as the Storwize V7000 nodes.

Note: Before attaching host systems to Storwize V7000, review the Configuration Limits and Restrictions for the IBM System Storage Storwize V7000 at this IBM Support web page.

3.12.1 Queue depth

Typically, hosts issue subsequent I/O requests to storage systems without waiting for 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.

Therefore, for large storage networks, plan for setting correct SCSI commands queue depth on your hosts. For this purpose, a large storage network is defined as one that contains at least 1000 volume mappings. For example, a deployment with 50 hosts with 20 volumes mapped to each of them is considered a large storage network. For more information about the queue depth calculations, search for "Queue depth in large SANs" at this IBM Knowledge Center web page.

3.12.2 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 extra load on storage controllers. The benefits and effects of enabling ODX are especially prominent in Microsoft Hyper-V environments with ODX enabled.

3.13 Host mapping and LUN masking

Host mapping is similar in concept to LUN mapping or masking. LUN mapping is the process of controlling which hosts have access to specific LUs within the disk controllers. LUN mapping is typically done at the storage system level. Host mapping is done at the software level.

LUN masking is usually implemented in the device driver software on each host. The host has visibility of more LUNs than it is intended to use. The device driver software masks the LUNs that are not to be used by this host. After the masking is complete, only some disks are visible to the operating system. The system can support this type of configuration by mapping all volumes to every host object and by using operating system-specific LUN masking technology. However, the default (and preferred) system behavior is to map only those volumes that the host is required to access.

The act of mapping a volume to a host makes the volume accessible to the WWPNs or iSCSI names such as iSCSI qualified names (IQNs) or extended-unique identifiers (EUIs) that are configured in the host object.

3.13.1 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 single I/O Group only. This approach allows you to 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 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.

3.14 NPIV planning

For more information, see N-Port Virtualization ID (NPIV) Support in Chapter 8, “Hosts” on page 317.

3.15 Advanced Copy Services

The Storwize V7000 offers the following Advanced Copy Services:

- FlashCopy
- Metro Mirror
- Global Mirror

Layers: A property called *layer* for the clustered system is used when a copy services partnership exists between a Storwize V7000 and an IBM Storwize V7000. There are two layers: *replication* and *storage*. By default, the IBM Storwize V7000 is configured as storage layer. This configuration must be changed by using the **chsystem** CLI command before you use it to make any copy services partnership with the Storwize V7000.

Figure 3-10 shows an example of replication and storage layers.

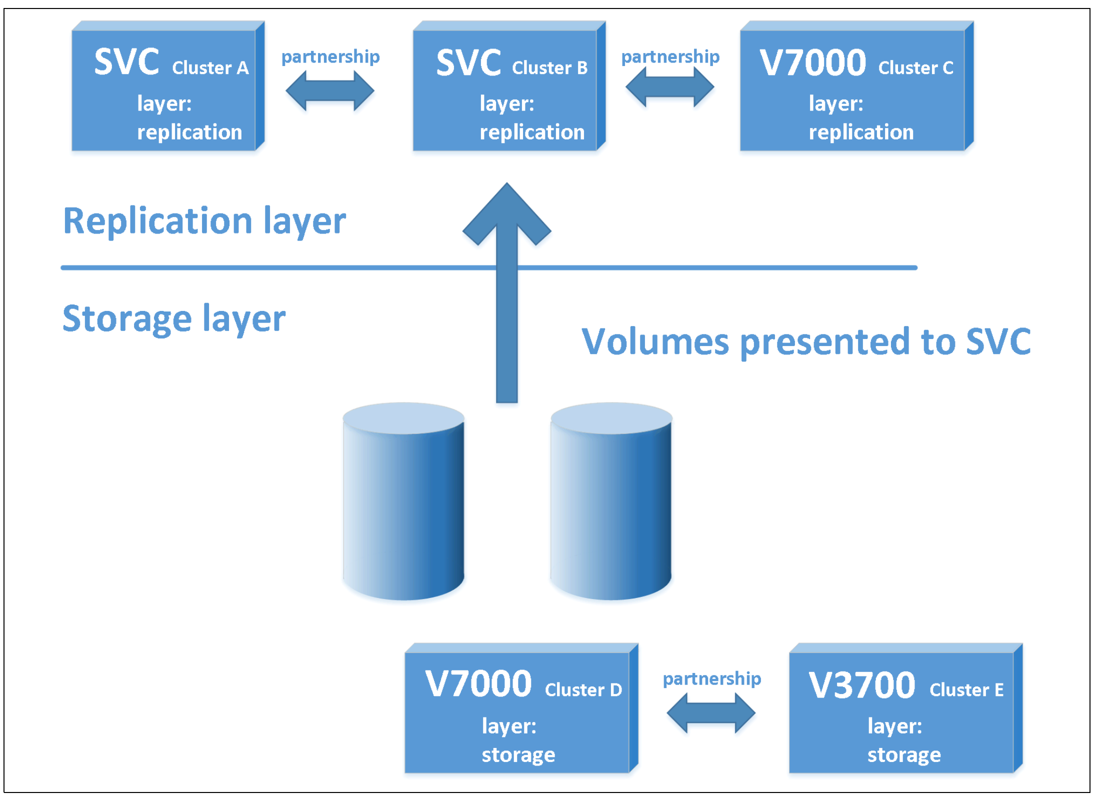


Figure 3-10 Replication and storage layer

The system's layer configuration cannot be changed if there are defined host objects with WWPNs. Therefore, ensure that the system layer is correctly planned and configured early during the deployment.

3.15.1 FlashCopy guidelines

When planning to use FlashCopy, observe the following guidelines:

- Identify each application that must have a FlashCopy function implemented for its volume.
- Identify storage pool or pools that will be used by FlashCopy volumes.
- Define which volumes need to use FlashCopy
- For each volume, define which FlashCopy type best fits your requirements:
  - No copy
  - Full copy
  - Thin-Provisioned
  - Incremental
- Define how many copies you need and the lifetime of each copy.
- Estimate the expected data change rate for FlashCopy types other than full copy.
- Consider memory allocation for copy services. If you plan to define multiple FlashCopy relationships, you might need to modify the default memory setting. For more information, see 11.2.18, "Memory allocation for FlashCopy" on page 496.
- Define the grain size that you want to use. When data is copied between volumes, it is copied in units of address space known as *grains*. The grain size is 64 KB or 256 KB. The FlashCopy bitmap contains 1 bit for each grain. The bit records whether the associated grain was split by copying the grain from the source to the target. Larger grain sizes can cause a longer FlashCopy time and a higher space usage in the FlashCopy target volume. The data structure and the source data location can modify those effects.

  If the grain is larger than most host writes, this issue can lead to write amplification on the target system. This increase is because for every write IO to an unsplit grain, the whole grain must be read from the FlashCopy source and copied to the target. Such a situation might result in performance degradation.

  If a thin-provisioned volume is used in a FlashCopy map, use the same grain size as the map grain size for best performance. Also, if a thin-provisioned volume is used directly with a host system, use a grain size that more closely matches the host IO size.
- Define which FlashCopy rate best fits your requirement in terms of the storage performance and the amount of time required to complete the FlashCopy. Table 3-4 lists the relationship of the background copy rate value to the number of grain split attempts per second.

  For performance-sensitive configurations, test the performance observed for different settings of grain size and FlashCopy rate in your actual environment before committing a solution to production use. See Table 3-4 for some baseline data.

Table 3-4 Grain splits per second

User percentage	Data copied per second	256 KiB grain per second	64 KiB grain per second
1 - 10	128 KiB	0.5	2
11 - 20	256 KiB	1	4
21 - 30	512 KiB	2	8

User percentage	Data copied per second	256 KiB grain per second	64 KiB grain per second
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	2948
111 - 120	256 MiB	1024	4096
121 - 130	512 MiB	2048	8192
131 - 140	1 GiB	4096	16384
141 - 150	2 GiB	8192	32768

3.15.2 Combining FlashCopy and Metro Mirror or Global Mirror

Use of FlashCopy in combination with Metro Mirror or Global Mirror is allowed if the following conditions are met:

- A FlashCopy mapping must be in the `idle_copied` state when its target volume is the secondary volume of a Metro Mirror or Global Mirror relationship.
- A FlashCopy mapping cannot be manipulated to change the contents of the target volume of that mapping when the target volume is the primary volume of a Metro Mirror or Global Mirror relationship that is actively mirroring.
- The I/O group for the FlashCopy mappings must be the same as the I/O group for the FlashCopy target volume.

3.15.3 Planning for Metro Mirror and Global Mirror

Metro Mirror is a copy service that provides a continuous, synchronous mirror of one volume to a second volume. The systems can be up to 300 kilometers (186.4 miles) apart. Because the mirror is updated synchronously, no data is lost if the primary system becomes unavailable. Metro Mirror is typically used for disaster-recovery purposes, where it is important to avoid any data loss.

Global Mirror is a copy service that is similar to Metro Mirror, 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 Metro Mirror. However, if a failure occurs, you might lose data.

Global Mirror uses one of two methods to replicate data. Multicycling Global Mirror 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 Global Mirror can be used so that less than a second of data is lost if a failure occurs. Global Mirror also works well when sites are more than 300 kilometers (186.4 miles) apart.

When Storwize V7000 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.

Starting with software version 7.6, you can use the **chsystem** command to set the maximum replication delay for the system. This value ensures that the single slow write operation does not affect the entire primary site.

You can configure this delay for all relationships or consistency groups that exist on the system by using the **maxreplicationdelay** parameter on the **chsystem** command. This value indicates the amount of time (in seconds) that a host write operation can be outstanding before replication is stopped for a relationship on the system. If the system detects a delay in replication on a particular relationship or consistency group, only that relationship or consistency group is stopped.

In systems with many relationships, a single slow relationship can cause delays for the remaining relationships on the system. This setting isolates the potential relationship with delays so that you can investigate the cause of these issues. When the maximum replication delay is reached, the system generates an error message that identifies the relationship that exceeded the maximum replication delay.

To avoid such incidents, consider deploying a SAN performance monitoring tool, such as IBM Tivoli Storage Productivity Center, to continuously monitor the SAN components for error conditions and performance problems. The use of such tool helps you detect potential issues before they affect your environment.

When planning for use of the data replication services, consider the following aspects of the solution:

- Volumes and consistency groups for copy services
- Copy services topology
- Choice between Metro Mirror and Global Mirror
- Connection type between clusters (FC, FCoE, IP)
- Cluster configuration for copy services, including zoning

IBM explicitly tests products for interoperability with the Storwize V7000. For more information about the current list of supported devices, see the IBM System Storage Interoperation Center (SSIC) website.

Volumes and consistency groups

Determine whether volumes can be replicated independently. Some applications use multiple volumes and require order of writes to these volumes to be preserved in the remote site. Notable examples of such applications are databases.

If application requires write order to be preserved for the set of volumes that it uses, create a consistency group for these volumes.

Copy services topology

One or more clusters can participate in a copy services relationship. One typical and simple use case is disaster recovery, where one site is active and another performs only a disaster recovery 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 allow you to design a solution that optimally fits your set of requirements. For more information about examples of valid relationships between systems, search for “Metro Mirror and Global Mirror partnerships” at this IBM Knowledge Center web page.

Global Mirror versus Metro Mirror

Decide which type of copy service you are going use. This decision should be requirements driven. Metro Mirror allows you to prevent any data loss during a system failure, but has more stringent requirements, especially regarding intercluster link bandwidth and latency, and remote site storage performance. Also, it incurs performance penalty because writes are not confirmed to host until data reception confirmation is received from the remote site.

Because of finite data transfer speeds, this remote write penalty grows with the distance between the sites. A point-to-point dark fiber-based link typically incurs a round-trip latency of 1 ms per 100 km (62.13 miles). Other technologies provide longer round-trip latencies. Inter-site link latency defines the maximum distance for any performance level.

Global Mirror allows you to relax constraints on system requirements at the cost of using asynchronous replication, which allows the remote site to lag behind the local site. Choice of the replication type has major effects on all other aspects of the copy services planning.

The use of Global Mirror and Metro Mirror between the same two clustered systems is supported.

If you plan to use copy services to realize some application function (for example, disaster recovery orchestration software), review the requirements of the application you plan to use. Verify that the complete solution is going to fulfill supportability criteria of both IBM and the application vendor.

Intercluster link

The local and remote clusters can be connected by FC, FCoE, or IP network. The IP network can be used as a carrier for FCoIP solution or as a native data carrier.

Each of the technologies has its own requirements concerning supported distance, link speeds, bandwidth, and vulnerability to frame or packet loss. For the most current information regarding requirements and limitations of each of supported technologies, see this website.

The two major parameters of a link are its bandwidth and latency. Latency might limit maximum bandwidth available over IP links, depending on the details of the technology that is used.

When planning the Intercluster link, consider the peak performance that is required. This consideration is especially important for Metro Mirror configurations.

When Metro Mirror or Global Mirror is used, a certain amount of bandwidth is required for the IBM Storwize V7000 intercluster heartbeat traffic. The amount of traffic depends on how many nodes are in each of the two clustered systems.

Table 3-5 lists the amount of heartbeat traffic (in megabits per second) that is generated by various sizes of clustered systems.

Table 3-5 Intersystem heartbeat traffic in Mbps

Storwize V7000 System 1	**Storwize V7000 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

These numbers estimate the amount of traffic between the two clustered systems when no I/O is taking place to mirrored volumes. Half of the data is sent by each of the systems. The traffic is divided evenly over all available intercluster links. Therefore, if you have two redundant links, half of this traffic is sent over each link.

The bandwidth between sites must be sized to meet the peak workload requirements. You can estimate the peak workload requirement by measuring the maximum write workload averaged over a period of 1 minute or less, and adding the heartbeat bandwidth. Statistics must be gathered over a typical application I/O workload cycle, which might be days, weeks, or months, depending on the environment in which the Storwize V7000 is used.

When planning the inter-site link, consider also the initial sync and any future resync workloads. It might be worthwhile to secure additional link bandwidth for the initial data synchronization.

If the link between the sites is configured with redundancy so that it 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, make a careful note whether it is dedicated to the inter-cluster traffic or is going to be used to carry any other data. Sharing link with other traffic (for example, cross-site IP traffic) might reduce the cost of creating the inter-site connection and improve link utilization. However, doing so might affect the links' ability to provide the required bandwidth for data replication.

Verify carefully that the devices that you plan to use to implement the intercluster link are supported.

Cluster configuration

If you configure replication services, you might decide to dedicate ports for intercluster communication, intracluster traffic, or both. In that case, make sure that your cabling and zoning reflects that decision. Also, such dedicated ports are inaccessible for host or back-end storage traffic, so plan your volume mappings and hosts and back-end storage connections accordingly.

Global Mirror volumes should have their preferred access nodes evenly distributed between the nodes of the clustered systems. Figure 3-11 shows an example of a correct relationship between volumes in a Metro Mirror or Global Mirror solution.

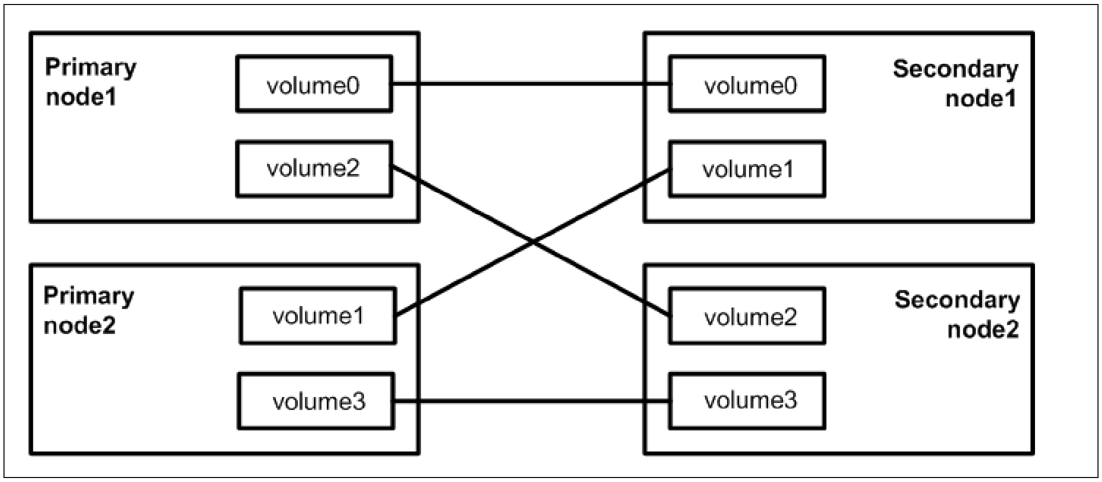


Figure 3-11 Correct volume relationship

The back-end storage systems at the replication target site must be capable to handle the peak application workload to the replicated volumes, plus the client-defined level of background copy, plus any other I/O being performed at the remote site. The performance of applications at the local clustered system can be limited by the performance of the back-end storage controllers at the remote site. This consideration is especially important for Metro Mirror replication.

A complete review must be performed before Serial Advanced Technology Attachment (SATA) drives are used for any Metro Mirror or Global Mirror replica volumes. If a slower disk subsystem is used as target for the remote volume replicas of high-performance primary volumes, the Storwize V7000 cache might not be able to buffer all the writes. The speed of writes to SATA drives at the remote site might limit the I/O rate at the local site.

To ensure that the back-end storage can support the data replication workload, you can dedicate back-end storage systems to only Global Mirror volumes. You can also configure the back-end storage to ensure sufficient quality of service (QoS) for the disks that are used by Global Mirror. Alternatively, you can ensure that physical disks are not shared between data replication volumes and other I/O.

3.16 SAN boot support

The IBM Storwize V7000 supports SAN boot or start-up for AIX, Microsoft Windows Server, and other operating systems. Because SAN boot support can change, check this IBM System Storage Interoperation Center (SSIC) web page regularly.

3.17 Data migration from a non-virtualized storage subsystem

Data migration is an important part of a Storwize V7000 implementation. Therefore, you must prepare a detailed data migration plan. You might need to migrate your data for one of the following reasons:

- Redistribute workload within a clustered system across back-end storage subsystems
- Move workload onto newly installed storage
- Move workload off old or failing storage, ahead of decommissioning it
- Move workload to rebalance a changed load pattern
- Migrate data from an older disk subsystem to Storwize V7000-managed storage
- Migrate data from one disk subsystem to another disk subsystem

Because multiple data migration methods are available, choose the method that best fits your environment, operating system platform, type of data, and the application's service level agreement (SLA).

Data migration methods can be divided into three classes:

- Based on operating system; for example, using system's Logical Volume Manager (LVM)
- Based on specialized data migration software
- Based on the Storwize V7000 data migration features

With data migration, apply the following guidelines:

- Choose the data migration method that best fits your operating system 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 Storwize V7000 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 Storwize V7000 cache might not be able to buffer all the writes. Flushing cache writes to slower back-end storage might impact performance of your hosts.

For more information about data migration using Storwize V7000, see 11.5, "Volume mirroring and migration options" on page 512.

3.18 Storwize V7000 configuration backup procedure

Save the configuration before and after any change to the clustered system, such as adding nodes and back-end storage. Saving the configuration is a crucial part of Storwize V7000 management, and various methods can be applied to back up your Storwize V7000 configuration. The preferred practice is to implement an automatic configuration backup using the configuration backup command. Make sure that you save the configuration to storage that is not dependent on the SAN Virtualization Controller.

For more information, see Chapter 13, “RAS, monitoring, and troubleshooting” on page 673.

3.19 Performance considerations

Storage virtualization with the Storwize V7000 improves flexibility and simplifies management of storage infrastructure, and can provide a substantial performance advantage. The Storwize V7000 caching capability and its ability to stripe volumes across multiple disk arrays are the reasons why usually significant performance improvements are observed when Storwize V7000 is used to virtualize midrange back-end storage subsystems.

> **Tip:** Technically, almost all storage controllers provide both striping (in the form of RAID 5, RAID 6, or RAID 10) and a form of caching. The real benefit of Storwize V7000 is the degree to which you can stripe the data across disks in a storage pool, even if they are installed in different back-end storage systems. This technique maximizes the number of active disks available to service I/O requests. The Storwize V7000 provides more caching, but its impact is secondary for sustained workloads.

To ensure the performance that you want and verify the capacity of your storage infrastructure, analyze performance and capacity to reveal the business requirements of your storage environment. Use the analysis results and the guidelines in this chapter to design a solution that meets the business requirements of your organization.

When considering performance for a system, always identify the bottleneck and, therefore, the limiting factor of a specific system. This is a multidimensional analysis that needs to be performed for each of your workload patterns. There can be different bottleneck components for different workloads.

When you are designing a storage infrastructure with the Storwize V7000 or implementing a Storwize V7000 in an existing storage infrastructure, you must ensure that the performance and capacity of the SAN, back-end disk subsystems, and Storwize V7000 meets requirements for the set of known or expected workloads.

3.19.1 SAN

The following Storwize V7000 models are supported for V8.2.1:

- Control enclosures:
  - 2076-524
  - 2076-624
- Expansion enclosures:
  - 2076-12F
  - 2076-24F

– 2076-92F
– 2076-A9F

The control enclosures can connect to 8 Gbps and 16 Gbps SAN switches. With an optional 10 Gbps Ethernet card to 10 Gbps Ethernet switch (for iSCSI traffic) and FCoE switches.

Model 2076-624 with 25 Gb Ethernet card can be used for RDMA (iSER). For more information, see 3.7.4, “iSCSI Extensions for RDMA (iSER)” on page 62.

3.19.2 Back-end storage subsystems

When connecting a back-end storage subsystem to IBM Storwize V7000, follow these guidelines:

- Connect all storage ports to the switch up to a maximum of 16, and zone them to all of the Storwize V7000 ports.
- Zone all ports on the disk back-end storage to all ports on the Storwize V7000 nodes in a clustered system.
- Ensure that you configure the storage subsystem LUN-masking settings to map all LUNs that are used by the Storwize V7000 to all the Storwize V7000 WWPNs in the clustered system.

The Storwize V7000 is designed to handle many paths to the back-end storage.

In most cases, the Storwize V7000 can improve performance, especially of mid-sized to low-end disk subsystems, older disk subsystems with slow controllers, or uncached disk systems, for the following reasons:

- The Storwize V7000 can stripe across disk arrays, and it can stripe across the entire set of configured physical disk resources.
- The Storwize V7000 control enclosure 2076-524 has 32 GB of cache and 2076-624 has 32 GB of cache (upgradeable to 64 GB).
- The Storwize V7000 can provide automated performance optimization of hot spots by using flash drives and Easy Tier.

The Storwize V7000 large cache and advanced cache management algorithms also allow it to improve the performance of many types of underlying disk technologies. The Storwize V7000 capability to asynchronously manage destaging operations that are incurred by writes while maintaining full data integrity can be important in achieving good database performance.

Because hits to the cache can occur in the upper (Storwize V7000) and the lower (back-end storage disk controller) level of the overall system, the system as a whole can use the larger amount of cache wherever it is located. Therefore, Storwize V7000 cache also provides more performance benefits for back-end storage systems with extensive cache banks.

Also, regardless of their relative capacities, both levels of cache tend to play an important role in enabling sequentially organized data to flow smoothly through the system.

However, Storwize V7000 cannot increase the throughput potential of the underlying disks in all cases. Performance benefits depend on the underlying storage technology and the workload characteristics, including the degree to which the workload exhibits hotspots or sensitivity to cache size or cache algorithms.

3.19.3 Storwize V7000

The Storwize V7000 clustered system is scalable up to eight nodes. Its performance grows nearly linearly when more nodes are added until it becomes limited by other components in the storage infrastructure. Although virtualization with the Storwize V7000 provides a great deal of flexibility, it does not abolish the necessity to have a SAN and back-end storage subsystems that can deliver the performance that you want.

Essentially, Storwize V7000 performance improvements are gained by using in parallel as many physical disks as possible, which creates a greater level of concurrent I/O to the back-end storage without overloading a single disk or array.

Assuming that no bottlenecks exist in the SAN or on the disk subsystem, you must follow specific guidelines when you perform the following tasks:

- Creating a storage pool
- Creating volumes
- Connecting to or configuring hosts that use storage presented by a Storwize V7000 clustered system

For more information about performance and preferred practices for the Storwize V7000, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

3.19.4 IBM Real-time Compression

IBM Rational® Team Concert® technology in storage systems is based on the Random Access Compression Engine (RACE) technology. It is implemented in IBM Storwize V7000 and the IBM Storwize family, IBM FlashSystem V840 systems, IBM FlashSystem V9000 systems, and IBM XIV (IBM Spectrum Accelerate™). This technology can play a key role in storage capacity savings and investment protection.

Although the technology is easy to implement and manage, it is helpful to understand the basics of internal processes and I/O workflow to ensure a successful implementation of any storage solution.

Consider the following suggestions:

- Best results can be achieved if the data compression ratio stays at 25% or more. Volumes can be scanned with the built-in Comprestimator utility to support decision if IBM Rational Team Concert™ is a good choice for the specific volume.
- More concurrency within the workload gives a better result than single-threaded sequential I/O streams.
- I/O is de-staged to RACE from the upper cache in 64 KiB pieces. The best results are achieved if the host I/O size does not exceed this size.
- Volumes that are used for only one purpose usually have the same work patterns. Mixing database, virtualization, and general-purpose data within the same volume might make the workload inconsistent. These workloads might have no stable I/O size and no specific work pattern, and a below-average compression ratio, which makes these volumes difficult to investigate during performance degradation. Real-time Compression development advises against mixing data types within the same volume whenever possible.
- It is best to not recompress pre-compressed data. Volumes with compressed data should stay as uncompressed volumes.

- Volumes with encrypted data have a low compression ratio, and are not good candidates for compression. This observation is true for data that is encrypted by the host. Real-time Compression can provide satisfactory results for volumes encrypted by Storwize V7000, because compression is performed before encryption.

3.19.5 Performance monitoring

Performance monitoring must be a part of the overall IT environment. For the Storwize V7000 and other IBM storage subsystems, the official IBM tool to collect performance statistics and provide a performance report is IBM Spectrum Control.

For more information about using IBM Spectrum Control to monitor your storage subsystem, see this website.

Also, see *IBM Spectrum Family: IBM Spectrum Control Standard Edition*, SG24-8321.

3.20 Storage Insights

Storage Insights is an integral part of the monitoring and ensuring continued availability of the Storwize V7000.

Available at no charge, cloud-based IBM Storage Insights provides a single dashboard that gives you a clear view of all your IBM block storage. You can make better decisions by seeing trends in performance and capacity.

Storage health information enables you to focus on areas needing attention and when IBM support is needed, Storage Insights simplifies uploading logs, speeds resolution with online configuration data, and provides an overview of open tickets all in one place.

Storage Insights includes the following features:

- A unified view of IBM systems:
  - Provides a single pane of glass to see all your systems characteristics.
  - See all of your IBM storage inventory.
  - Provides a live event feed so you know, up to the second, what is going on with your storage and enables you to act quickly.
- Storage IBM Insight® collects telemetry data and call home data and provides.

  Up to the second system reporting of capacity and performance.
- Overall storage monitoring looking at:
  - The overall health of the system.
  - Monitor the configuration to see whether it meets the best practices.
  - System resource management; that is, is the system being overly taxed and provide proactive recommendations to fix it.
- Storage Insights provides advanced customer service with an event filter to that allows:
  - The ability for you and support to view support tickets and open and close them and to track trends.
  - Auto log collection capability to allow you to collect the logs and send to IBM before support start looking into the problem. This feature can save as much as 50% of the time to resolve the case.

In addition to the free Storage Insights, Storage Insights Pro is available, which is a subscription service that provides longer historical views of data, more reporting and optimization options, and supports IBM file and block storage with EMC VNX and VMAX.

Figure 3-12 shows the comparison of Storage Insights and Storage Insights Pro.

Product Comparison

	Capability	IBM Storage Insights (Free)	IBM Storage Insights Pro (Subscription)
Monitoring	Health, Performance and Capacity	✓	✓
	Filter events to quickly isolate trouble spots	✓	✓
	Drill down performance workflows to enable deep troubleshooting		✓
	Application / server storage performance troubleshooting		✓
	Customizable multi-conditional alerting		✓
Support Services	Simplified ticketing / log workflows and ticket history	✓	✓
	Proactive notification of risks (select systems)	✓	✓
Device Analytics	Part failure prediction	✓	✓
	Configuration best practice	✓	✓
	Customized upgrade recommendation	✓	✓
TCO Analytics	Capacity planning		✓
	Performance planning		✓
	Application / server storage consumption		✓
	Capacity optimization with reclamation planning		✓
	Data optimization with tier planning		✓

Figure 3-12 Storage Insights and Storage Insights Pro comparison chart

3.20.1 Architecture, security, and data collection

Figure 3-13 shows the architecture of the Storage Insights application, the supported products, and the three main teams who can benefit from the use of the tool.

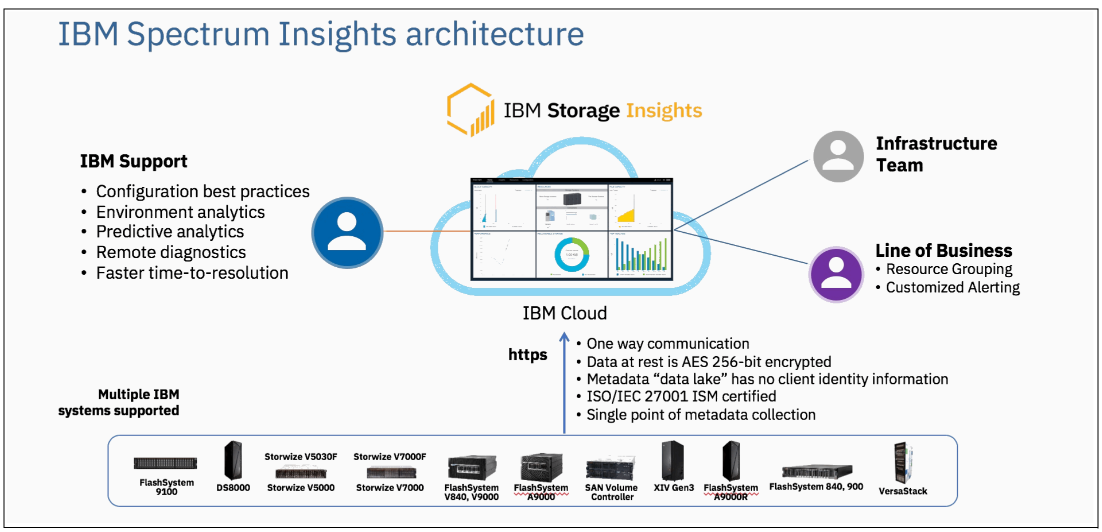


Figure 3-13 Storage Insights Architecture

Storage Insights provides a lightweight data collector that is deployed on a customer supplied server. This server can be a Linux, Windows, or AIX server or a guest in a virtual machine (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: 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.

Figure 3-14 shows the data flow from systems to the IBM Storage Insights cloud.

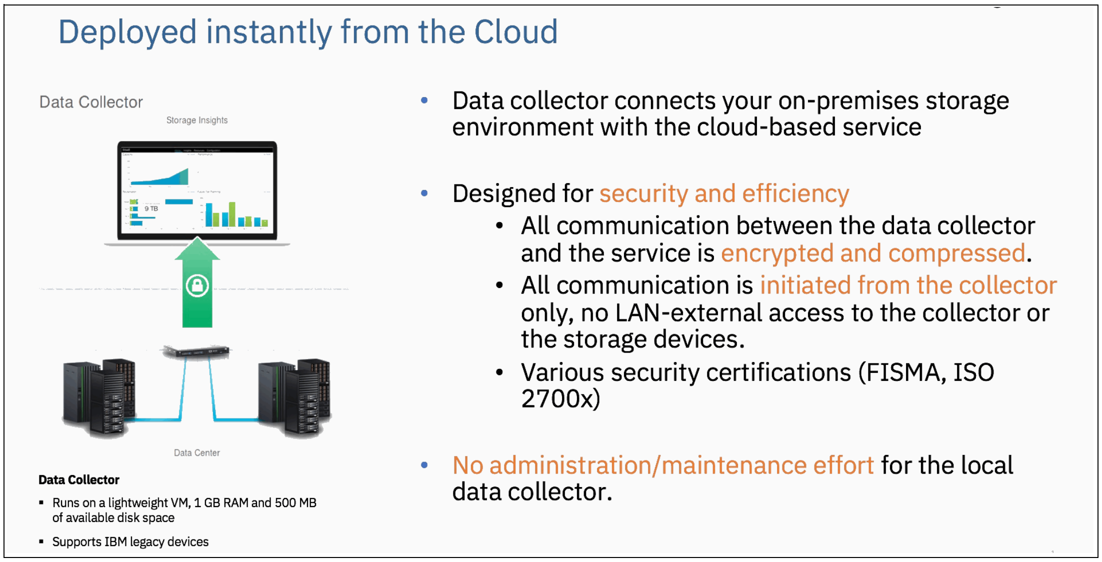


Figure 3-14 Data flow from the storage systems to the IBM Storage Insights cloud

What metadata is collected

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.

Who can access the metadata

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.

3.20.2 Customer Dashboard

Figure 3-15 shows a view of the Storage Insights main dashboard and the systems that it is monitoring.

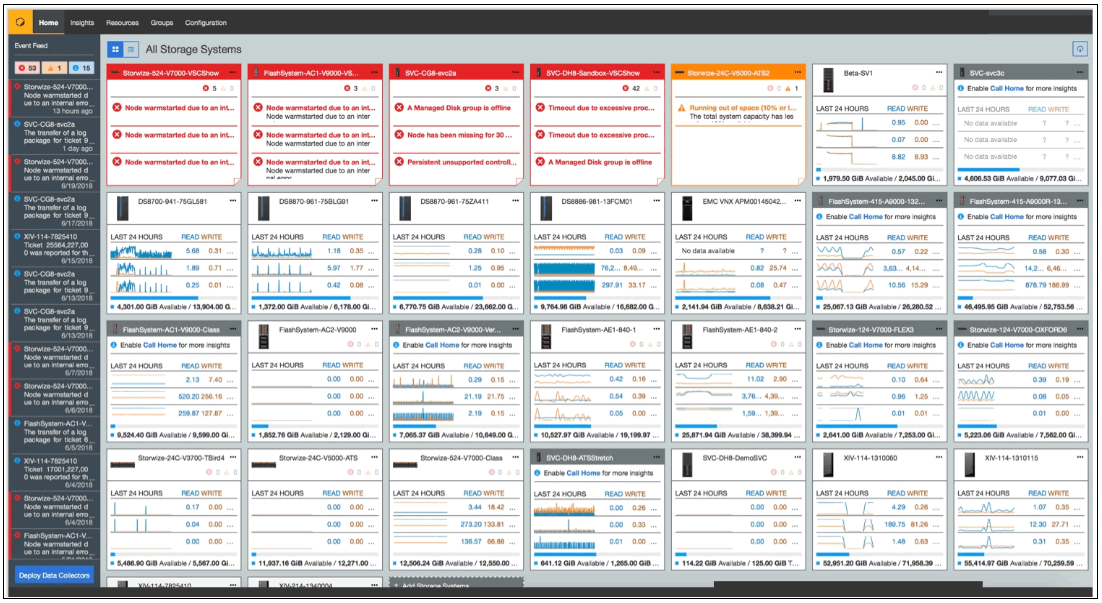


Figure 3-15 Storage Insights dashboard

More views and images of dashboard displays and drill-downs can be found in the supporting documentation.

More information and registration

For more information about Storage Insights and to sign up and register for the free service, see the following resources:

- Fact Sheet
- Demonstration
- Security Guide
- IBM Knowledge Center
- Registration

4

Initial configuration

This chapter describes the initial configuration of the IBM Storwize V7000 system. It provides step-by-step instructions about how to create the cluster, define its basic settings, and add nodes and optional expansion enclosures.

Other features, such as user authentication, secure communications, and local port masking, are also covered. These features are optional and do not need to be configured during the initial configuration.

This chapter includes the following topics:

4.1 Prerequisites

Before initializing and setting up the Storwize V7000, ensure that the following prerequisites are met:

- The installation of physical components is planned to fulfill all requirements and correctly executed, including:
  - Control enclosures are physically installed with the correct cabling.
  - The Ethernet and Fibre Channel connectivity are correctly configured.
  - Expansion enclosures, if available, are physically installed and attached to the Storwize V7000 nodes in the I/O group that is meant to use them.
  - The Storwize V7000 control enclosures and optional expansion enclosures are powered on.
- Your web browser is supported and has the appropriate settings enabled. For more information about supported browsers and settings, see IBM Knowledge Center.
- You have the required information available, including:
  - For IPv4 addressing (if used):
    - Cluster IPv4 address, which is the address that is used for the management of the system.
    - Service IPv4 addresses, which are used to access node service interfaces. You need one address for each node.
    - IPv4 subnet mask for each subnet used.
    - IPv4 gateway for each subnet used.
  - For IPv6 addressing (if used):
    - Cluster IPv6 address, which is used for the management of the system.
    - Service IPv6 addresses, which are used to access node service interfaces. You need one address for each node.
    - IPv6 prefix for each subnet used.
    - IPv6 gateway. for each subnet used.
  - The licenses that enable you to use licensed functions, which include the licenses that indicate your entitlement to use licensed functions:
    - Remote Copy
    - External Virtualization
    - Real-time Compression
    - Transparent Cloud Tiering
  - Physical location of the system.
  - The name, email address, and phone number of the storage administrator who IBM can contact if necessary.
  - The Network Time Protocol (NTP) server IP address (optional, but recommended), which is necessary only if you want to use an NTP service instead of manually entering date and time.
  - The Simple Mail Transfer Protocol (SMTP) email server IP address (optional), which is necessary only if you want to enable *call home*.
  - The IP addresses for Remote Support Proxy Servers (optional), which are necessary only if you want to enable Support Assistance.

4.2 System initialization

This section provides step-by-step instructions about how to create the Storwize V7000 cluster. The procedure is initiated by connecting to the technician port.

Attention: Do not perform the system initialization procedure on more than one node canister. After system initialization completes, use the management GUI to add control enclosures to the system.

During system initialization, you must specify an IPv4 or an IPv6 system address. This address is assigned to Ethernet port 1. After system initialization, you can specify more IP addresses for port 1 and port 2 until both ports have an IPv4 address and an IPv6 address.

Connect a personal computer (PC) or notebook to the technician port on the rear of the control enclosure. If you plan to use more than one enclosure to configure the system, choose any of the enclosures.

Figure 4-1 shows the location of the technician port.

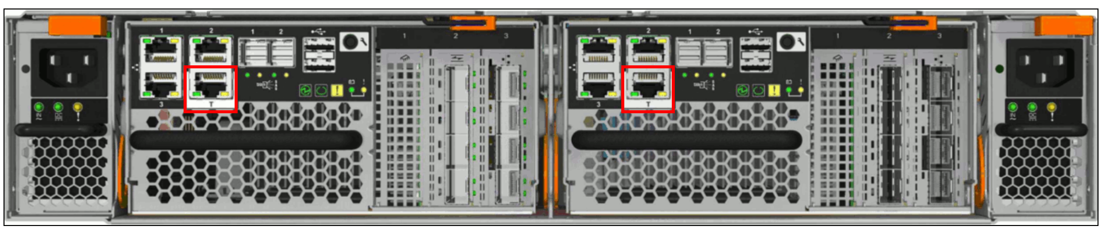


Figure 4-1 Location of the technician port

The technician port runs an IPv4 DHCP server and it can assign an address to any device connected to this port. Ensure that your PC or notebook Ethernet port is configured for DHCP 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 192.168.0.0/24 subnet; for example, `192.168.0.2/24`.

The default IP address for a new node is `192.168.0.1/24`. Do not use this IP address for your PC or notebook.

Note: The Storwize V7000 does *not* provide IPv6 IP addresses for the technician port.

Ensure that the technician port is not connected to the organization's network, or it acts as a DHCP server for other devices in the network segment.

4.2.1 System initialization wizard

After connecting your PC or notebook to the technician port, ensure that you obtained a valid IPv4 DCHP address (for example, `192.168.0.12/24`) and then complete the following steps to initialize the system:

1. Open a supported browser and browse to `http://install`. The browser is automatically redirected to the System Initialization wizard. You can use instead the URL with the IP address (`http://192.168.0.1`) if you are not automatically redirected.

   If the system cannot be initialized, you are redirected to the Service Assistant interface. Use the displayed error codes to troubleshoot the problem.

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 as to the identity of the certificate issuer, so you can safely accept the certificates.

2. The welcome dialog box opens, as shown in Figure 4-2. Click **Next** to start the procedure.

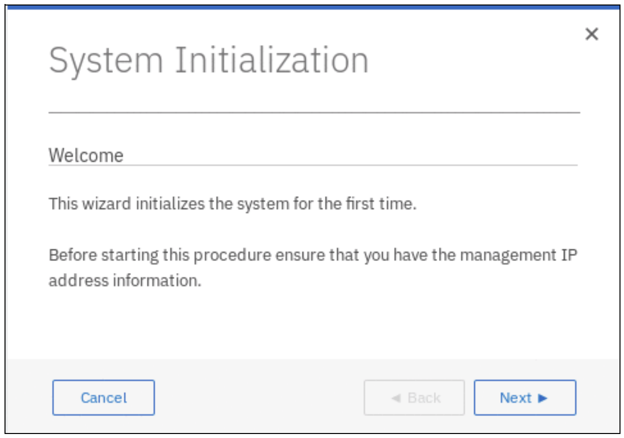


Figure 4-2 System initialization: Welcome

3. A window opens in which two options are presented, as shown in Figure 4-3. Select the first option, **As the first enclosure in a new system**. Click **Next**.

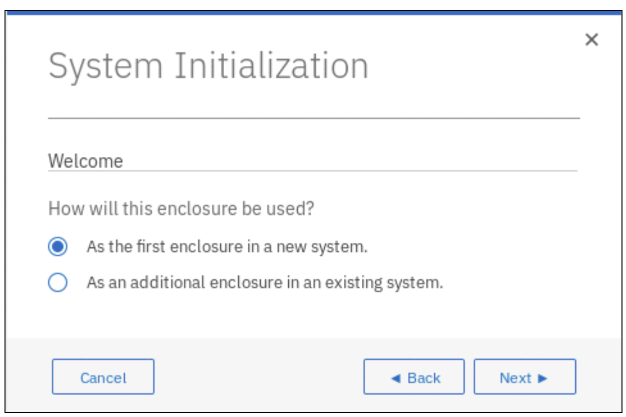


Figure 4-3 System initialization: Configuring the first enclosure in a new system

4. Enter the IP address information for the new system (see Figure 4-4 on page 91). Choose between an IPv4 and IPv6 address. Enter the wanted address and click **Next**.

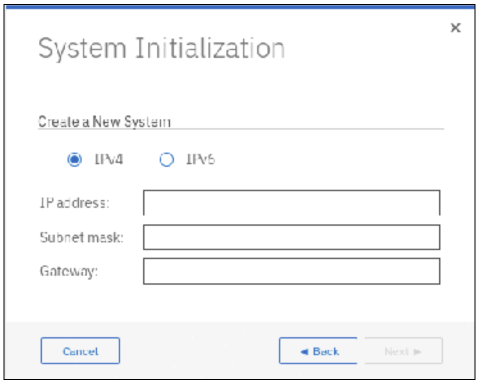


Figure 4-4 System initialization: Setting the system IP address

5. The web server restarts. Wait until the timer reaches the end, as shown in Figure 4-5, and click **Next**.

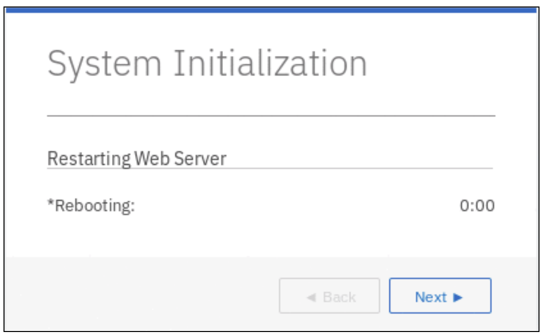


Figure 4-5 System initialization: Web server restart

6. After the system initialization is complete, follow the instructions as shown in Figure 4-6:
   a. Disconnect the Ethernet cable from the technician port and from your PC or notebook.
   b. Connect the PC or notebook to the same network as the system.
   c. Click **Finish** to be redirected to the management GUI to complete the system setup.

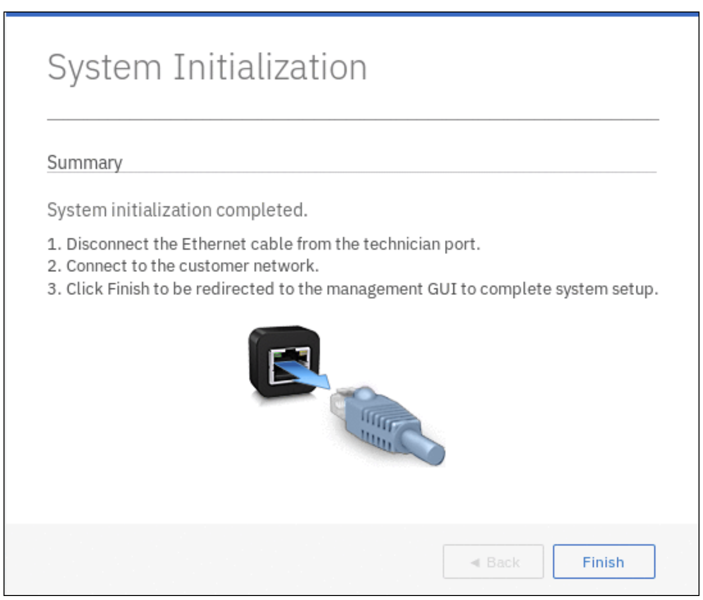


Figure 4-6 System initialization: Summary

Note: Make sure that your PC or notebook has a network route to the system IP address that you specified. In particular, you can access the management GUI from any management console that is connected to the same subnet as the system. Enter the system IP address on a supported browser to access the management GUI.

4.3 System setup

This section provides instructions about how to define the basic settings of the system with the system setup wizard, and how to add nodes and optional expansion enclosures.

4.3.1 System setup wizard

Whether you are redirected from your PC or notebook after completing system initialization or you browse to the management IP address manually, you must complete the system setup wizard to define the basic settings of the system.

Note: The first time that you connect to the management GUI, you are prompted to accept untrusted certificates because the system certificates are self-signed.

You can install certificates that are signed by a trusted certificate authority after you complete system setup. For more information about how to perform this task, see 4.5, “Configuring secure communications” on page 117.

Complete the following steps to successfully complete the system setup wizard:

1. Log in to system by using the superuser account, as shown in Figure 4-7. Click **Log in**.

> **Important:** The default password for the superuser account is `passw0rd` (with the number zero and not the capital letter o).

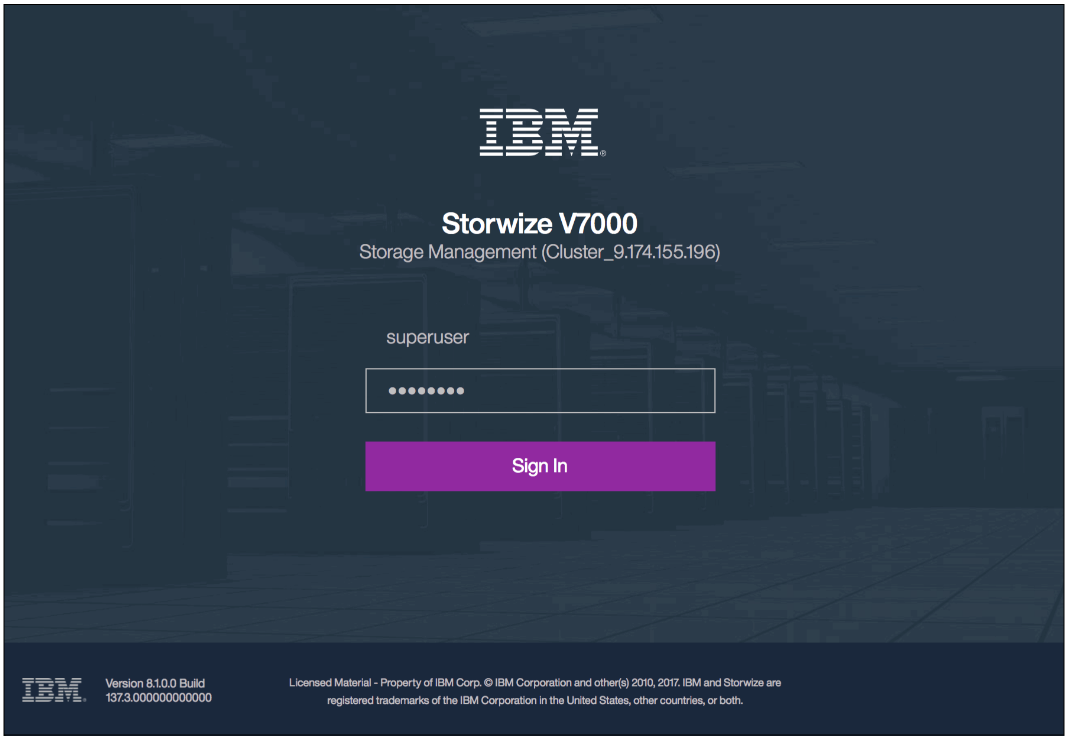


Figure 4-7 System setup: Logging in for the first time

2. The welcome window opens, as shown in Figure 4-8. Verify the prerequisites and click **Next**.

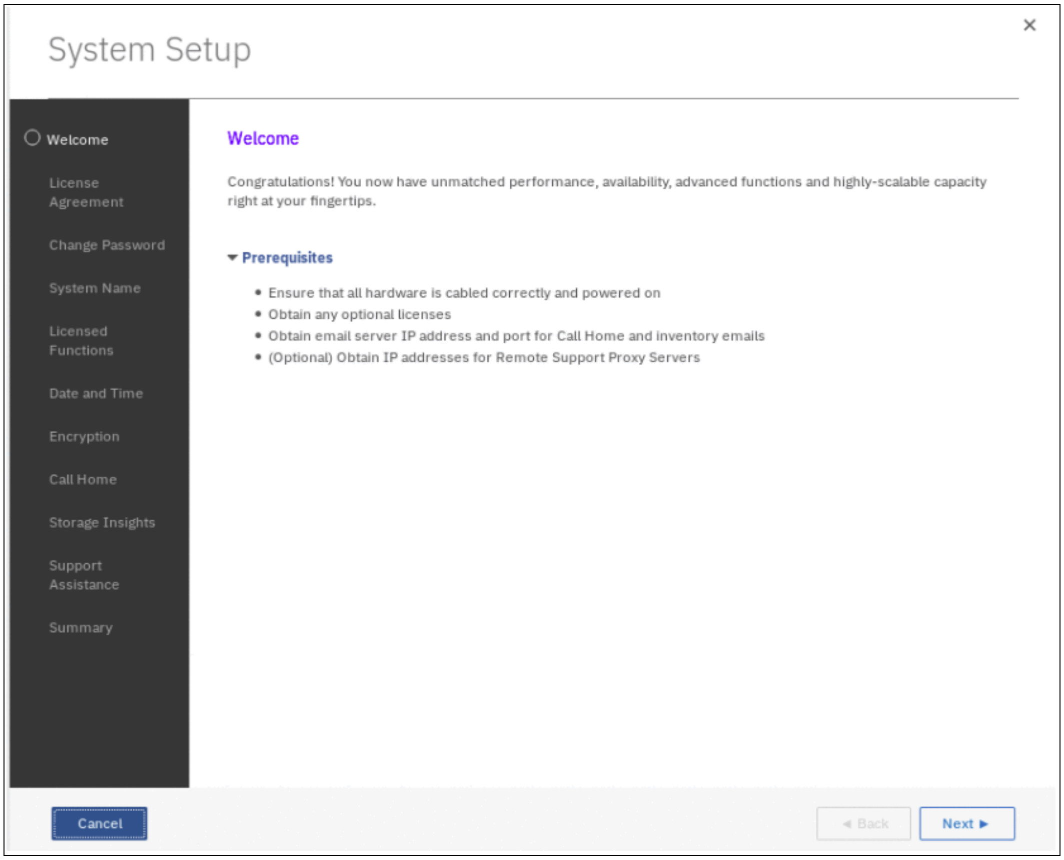


Figure 4-8 System setup: Welcome

3. Carefully read the license agreement. Select **I agree with the terms in the license agreement** when you are ready, as shown in Figure 4-9. Click **Next**.

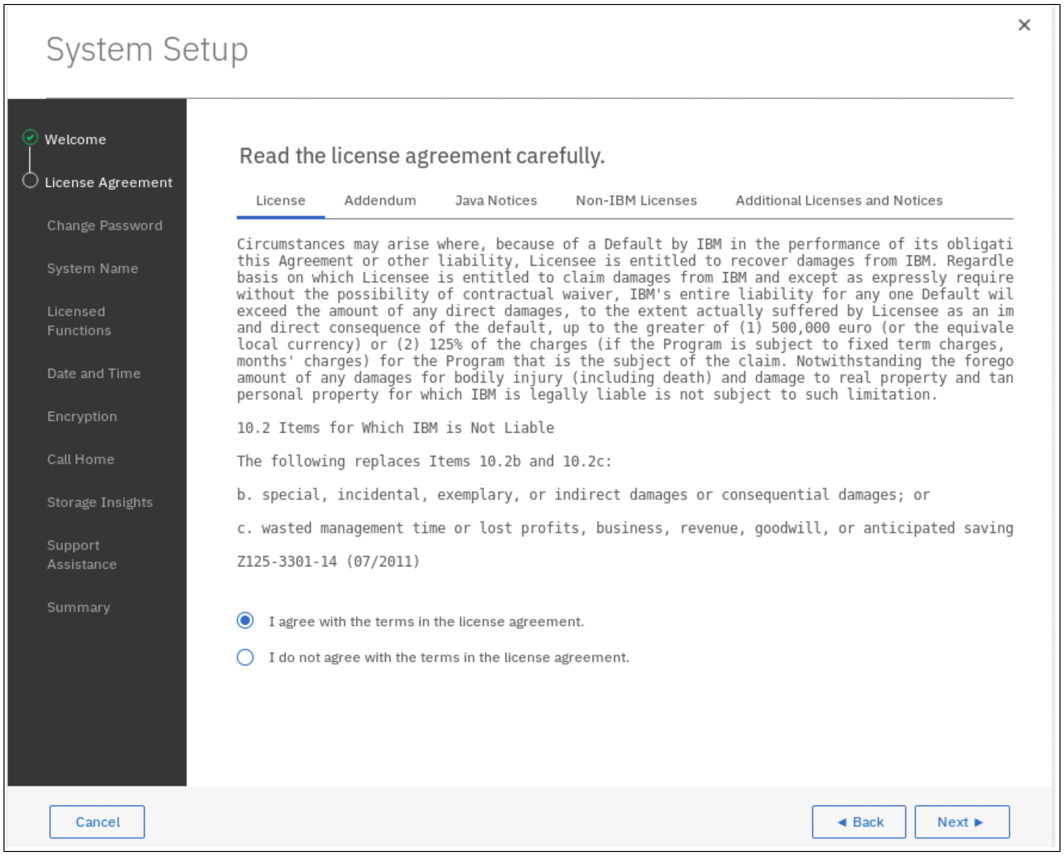


Figure 4-9 System setup: License agreement

4. Enter a new password for superuser, as shown in Figure 4-10. The valid password is 6 - 64 characters long and it cannot begin or end with a space. Click **Apply and Next**.

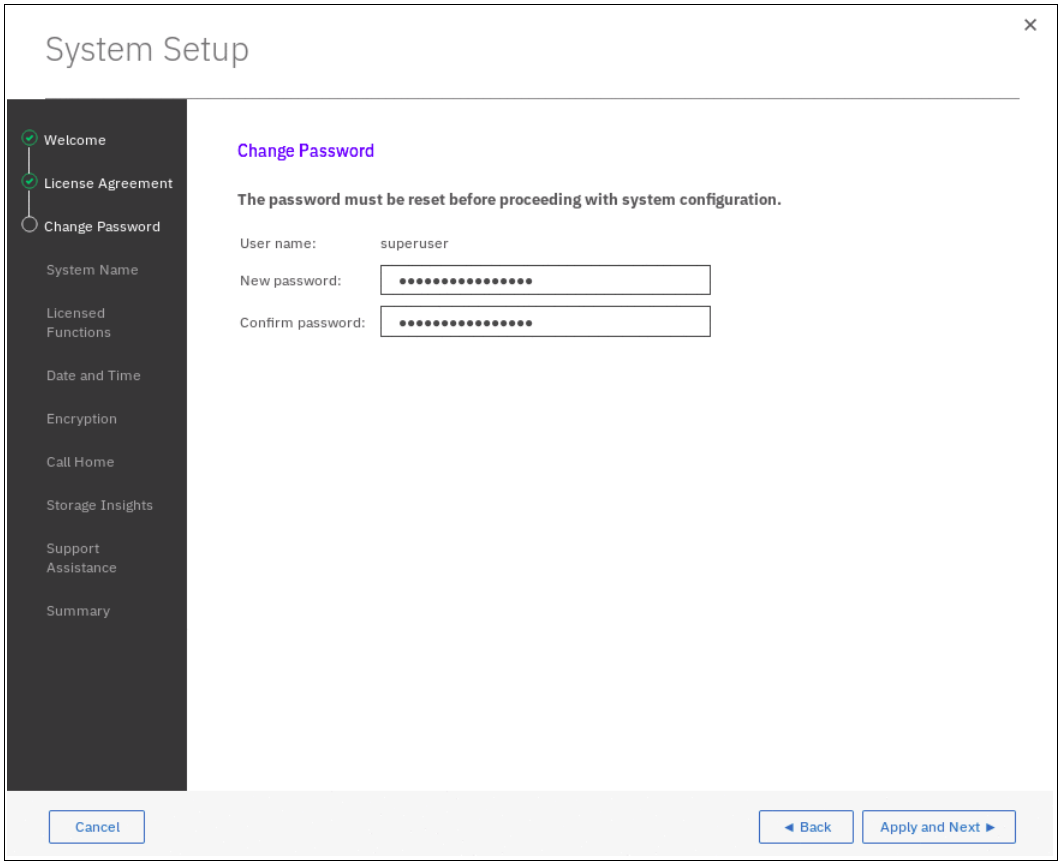


Figure 4-10 System setup: Changing the password for superuser

5. Enter the name that you want to give the new system, as shown in Figure 4-11. Click **Apply and Next**.

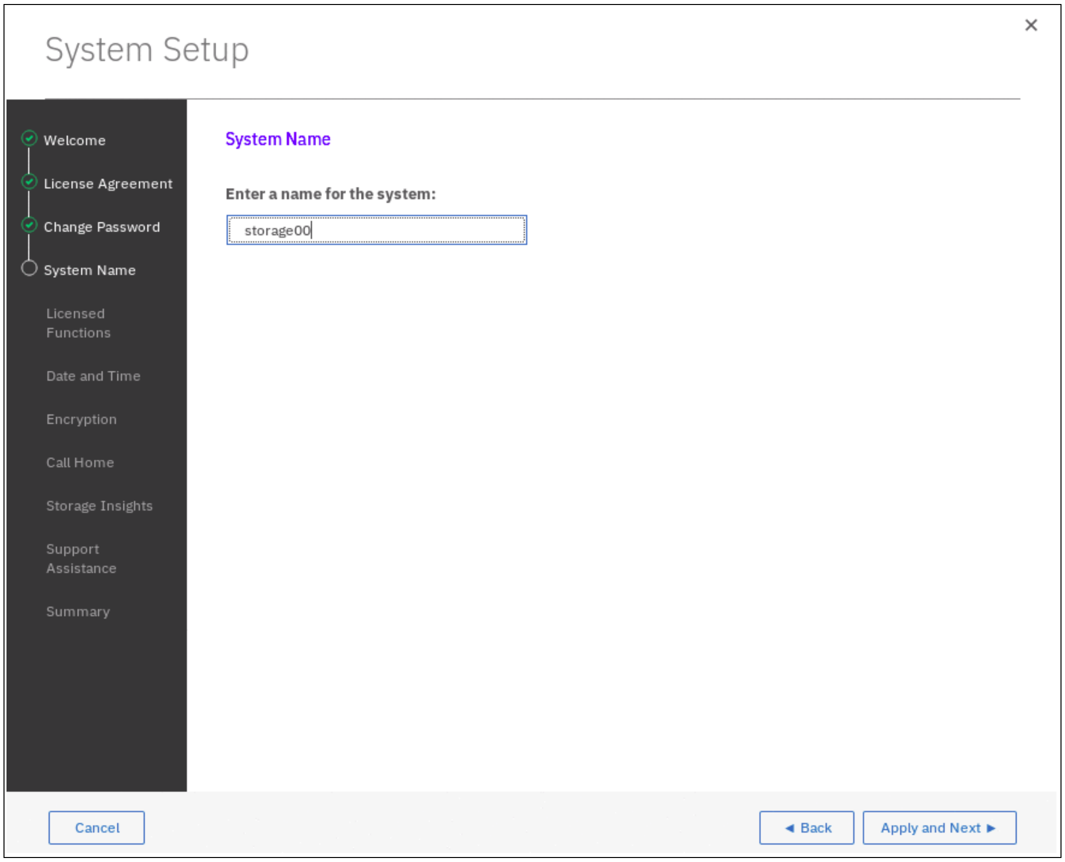


Figure 4-11 System setup: Setting the system name

6. Enter the number of enclosures licensed for each function as authorized by your license agreement (see Figure 4-12). Click **Apply and Next**.

> **Note:** Encryption uses a different licensing scheme and is activated later in the wizard.

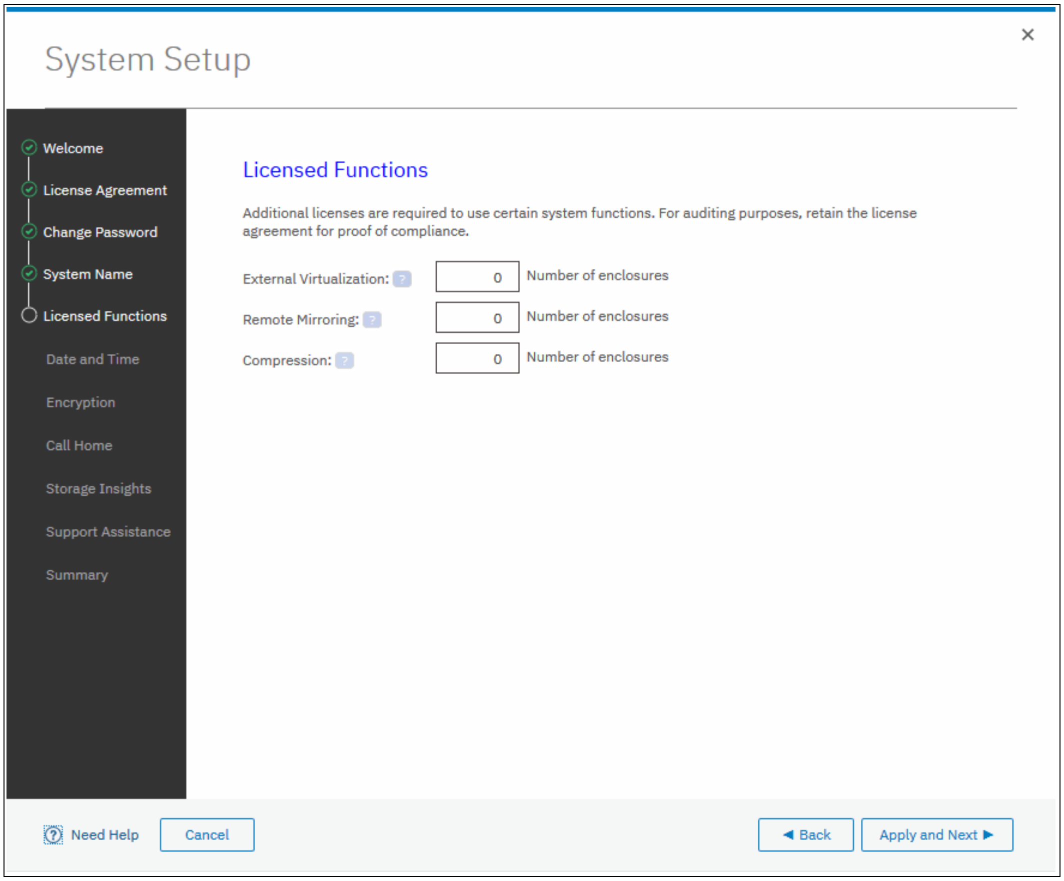


Figure 4-12 System setup: Setting the system licenses

All functions follow an enclosure-based licensing scheme. Use the following guidelines to ensure that you are licensing the correct number of enclosures:

– External Virtualization

 The system does not require a license for its own control and expansion enclosures. However, a license is required for each enclosure of any external storage systems being virtualized. For example, if the system is made up of one control enclosure, one expansion enclosure, and one virtualized storage system that is made up of two enclosures, two licenses are required. The system does not require an external virtualization license for external enclosures that are only being used to provide managed disks for a quorum disk and are not providing any capacity for volumes.

– Remote Mirroring, Real-time Compression, or Transparent Cloud Tiering

 The total number of enclosures must include the enclosures on external storage system licensed for virtualization plus the number of control and expansion enclosures that are part of your local system. For example, if the system is made up of one control enclosure, one expansion enclosure, and one virtualized storage system that is made up of two enclosures, four licenses are required. Only Storwize V7000 Gen2 and Storwize V7000 Gen2+ models support transparent cloud tiering.

7. Enter the date and time settings. In the example that is shown in Figure 4-13, the date and time are set by using an NTP server. Generally, use an NTP server so that all of your storage area network and storage devices have a common time stamp. This practice facilitates troubleshooting and prevents time stamp-related errors if you use a key server as an encryption key provider. Click **Apply and Next**.

> **Note:** If you choose to manually enter these settings, you cannot select the 24-hour clock now. However, you can select the 24-hour clock after you complete the wizard by clicking **Settings** → **System** and selecting **Date and Time**.

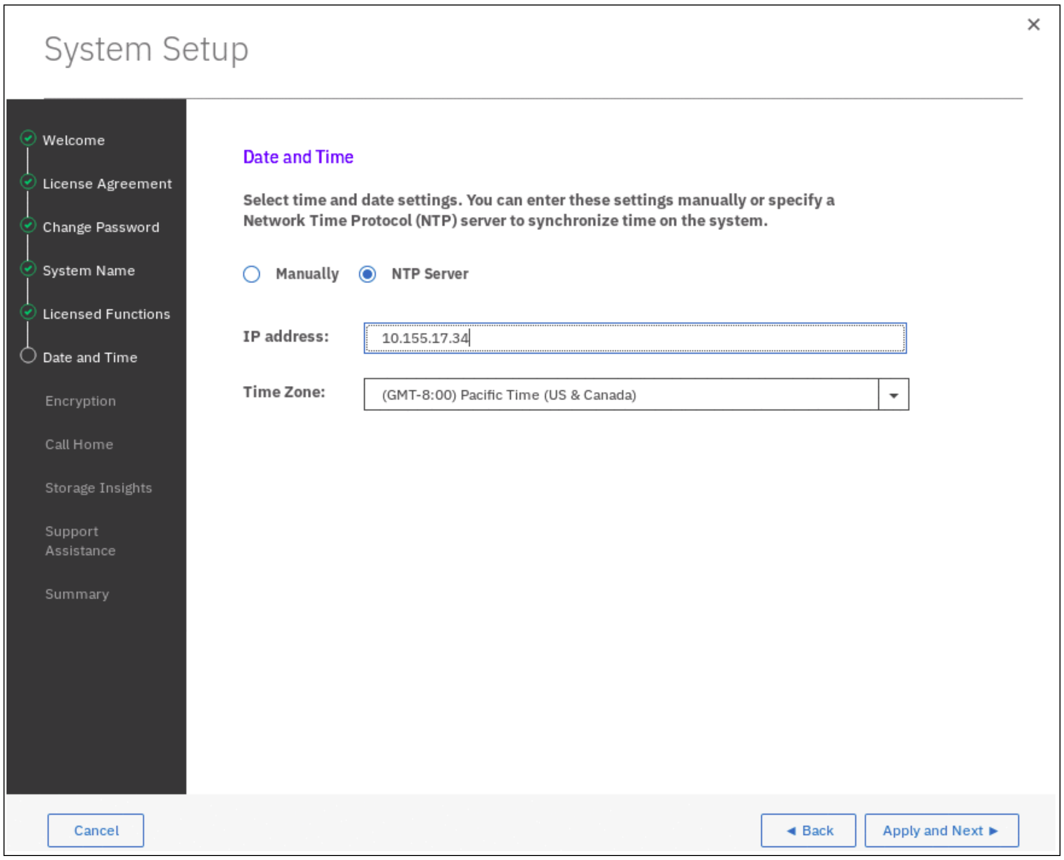


Figure 4-13 System setup: Setting date and time settings

8. Select whether the encryption feature was purchased for this system. In this example, it is assumed encryption was not purchased, as shown in Figure 4-14. Click **Next**.

Note: If you purchased the encryption feature, you are prompted to activate your encryption license manually or automatically. For more information about how to activate your encryption license during the system setup wizard, see Chapter 12, “Encryption” on page 603.

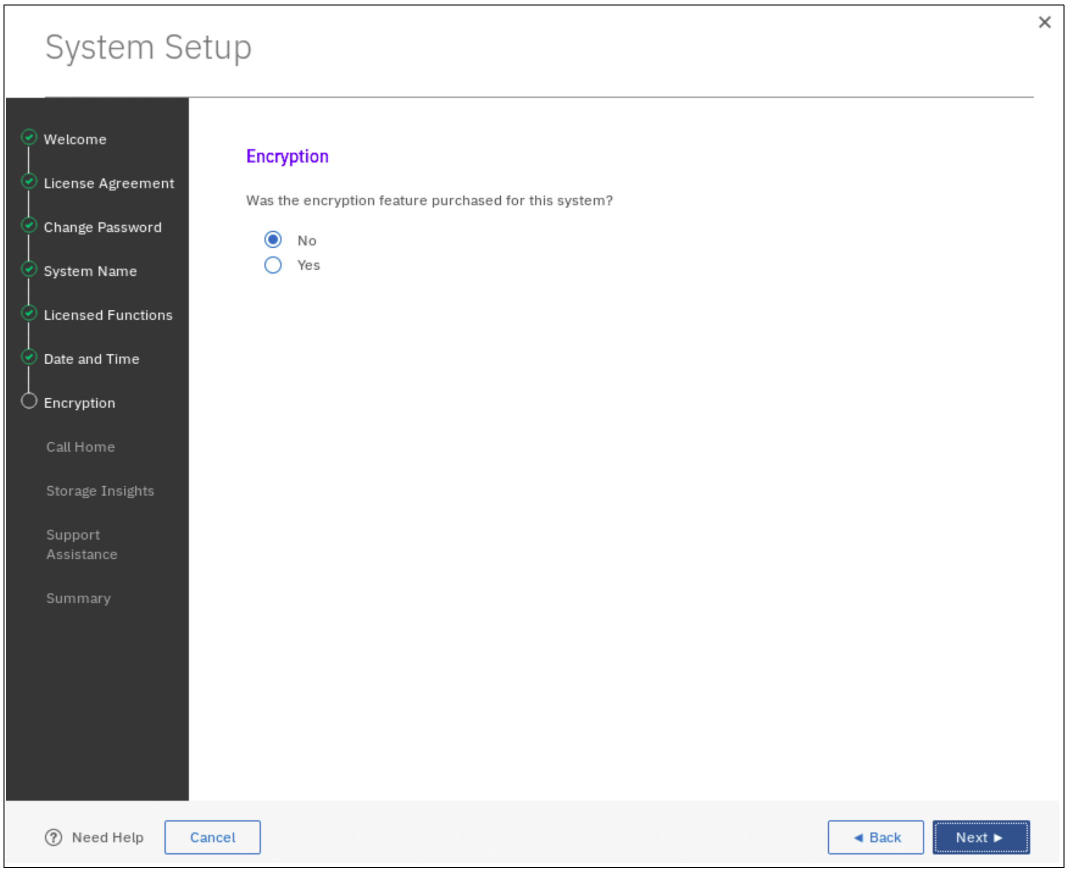


Figure 4-14 System setup: Encryption

9. To set up remote support, check the Send data to the support center and click **Next** (see Figure 4-15).

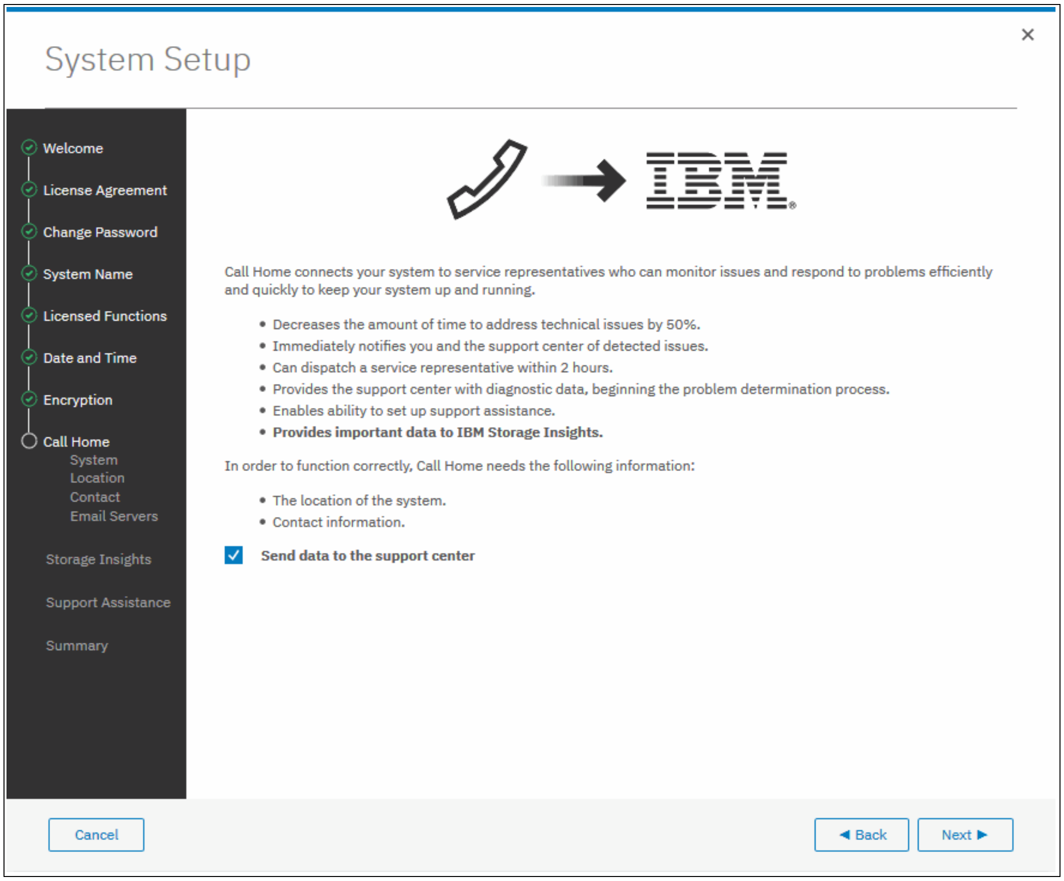


Figure 4-15 Initial Call Home setup

For more information about setting up Call Home, including Cloud Call Home see Chapter 13, "RAS, monitoring, and troubleshooting" on page 673.

10. Enter the system location details. Figure 4-16 shows some details as an example only. Click **Next**.

> **Note:** If your system is not in the US, enter `XX` into the state or province field.

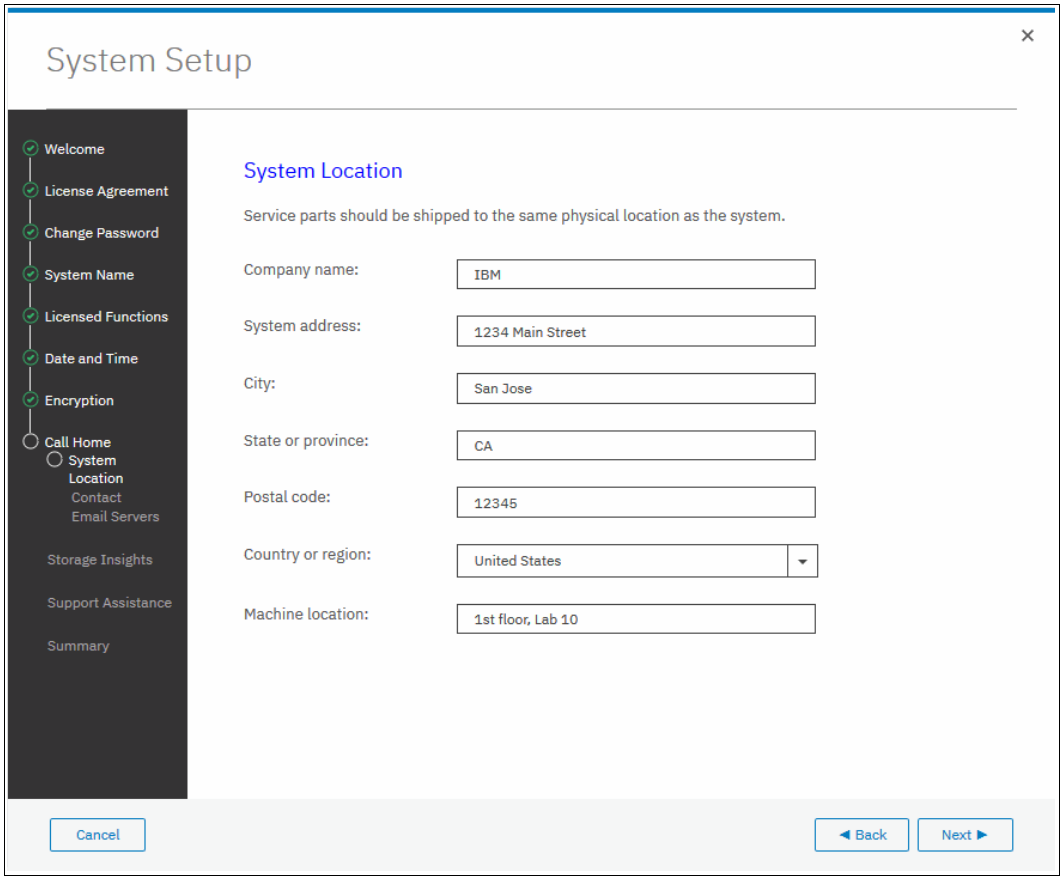


Figure 4-16 System setup: Setting the system location details

11. Enter the contact details of the person to be contacted to resolve issues on the system. You can choose to enter the contact information for a 24-hour operations desk. Figure 4-17 shows some details as an example only. Click **Apply and Next**.

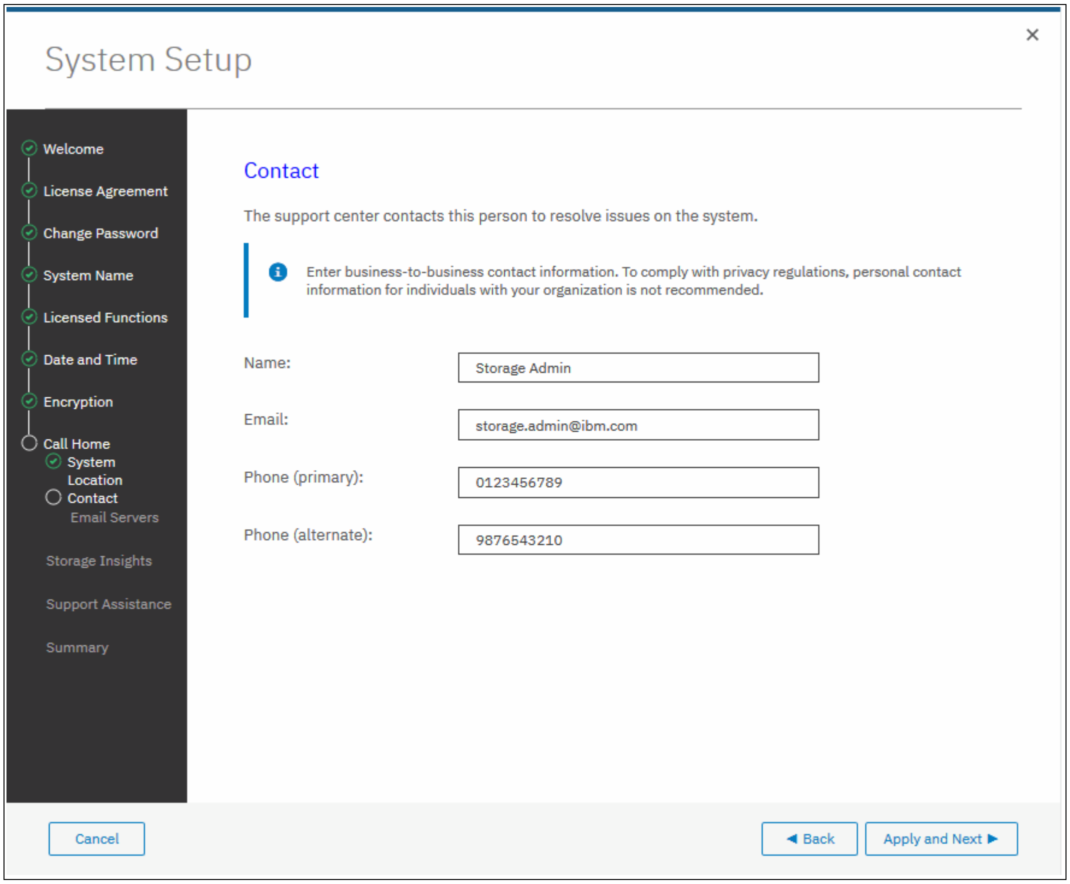


Figure 4-17 System setup: Setting contact information

12. Enter the details for the email servers to be used for Call Home. Call Home sends email reports to IBM with inventory details and event notifications. This setting allows IBM to automatically open problem reports and to contact you to verify whether replacement parts are required.

Figure 4-18 shows an example. You can click **Ping** to verify that the email server is reachable over the network. Click **Apply and Next**.

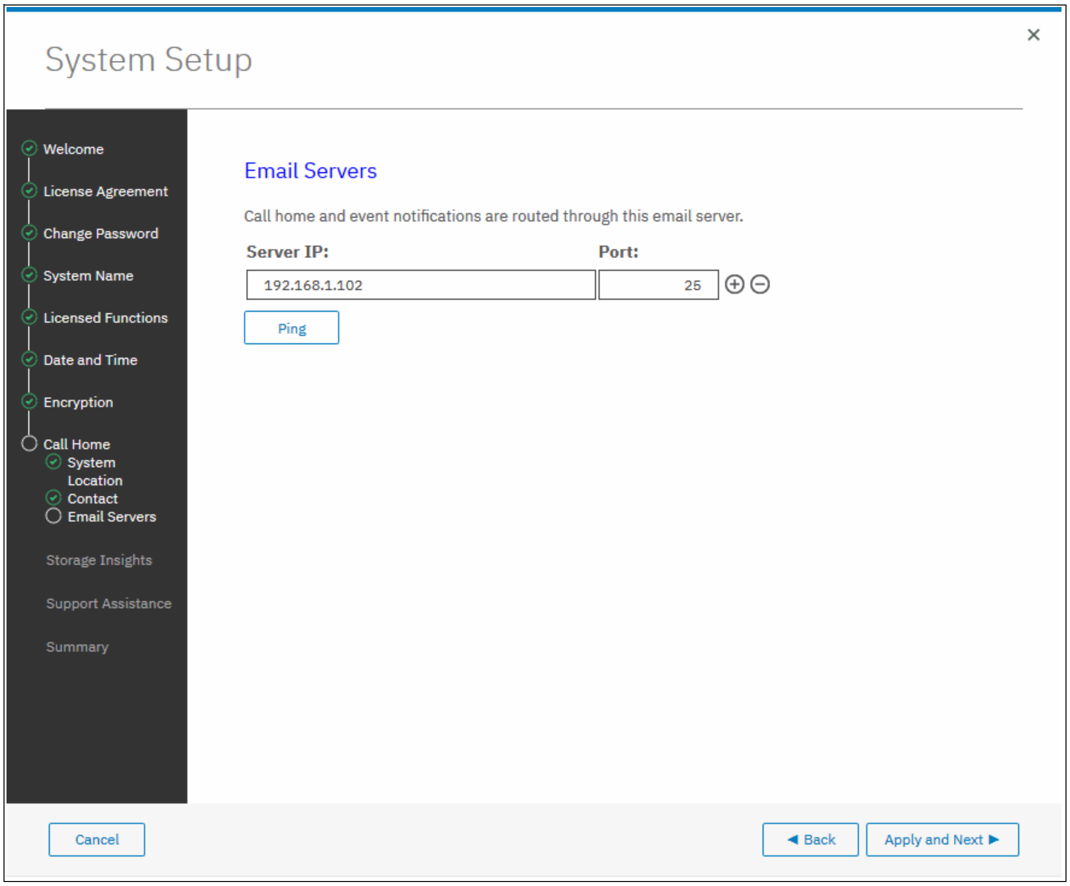


Figure 4-18 Setting email servers details

Storwize V7000 can use SNMP traps, syslog messages, and Call Home to notify you and IBM Support when significant events are detected. Any combination of these notification methods can be used simultaneously. However, only Call Home is configured during the system setup wizard. For more information about how to configure other notification methods, see Chapter 4, “Initial configuration” on page 87.

Note: When Call Home is configured, the system automatically creates a support contact with one of the following email addresses, depending on country or region of installation:

- US, Canada, Latin America, and Caribbean Islands: `callhome1@de.ibm.com`
- All other countries or regions: `callhome0@de.ibm.com`

If you do not want to configure Call Home now, it can be done later by navigating to **Settings** → **Notifications**.

Advisory note: Configure Call Home if your system is under warranty or if you have a hardware maintenance agreement.

13. Provide data that is required to use IBM Storage Insights. Although contact information for IBM Storage Insights is copied from the Call Home configuration, you must supply an IBM ID to use this feature (see Figure 4-19). Optionally, you can skip this step by selecting **I'm not interested in Storage Insights**.

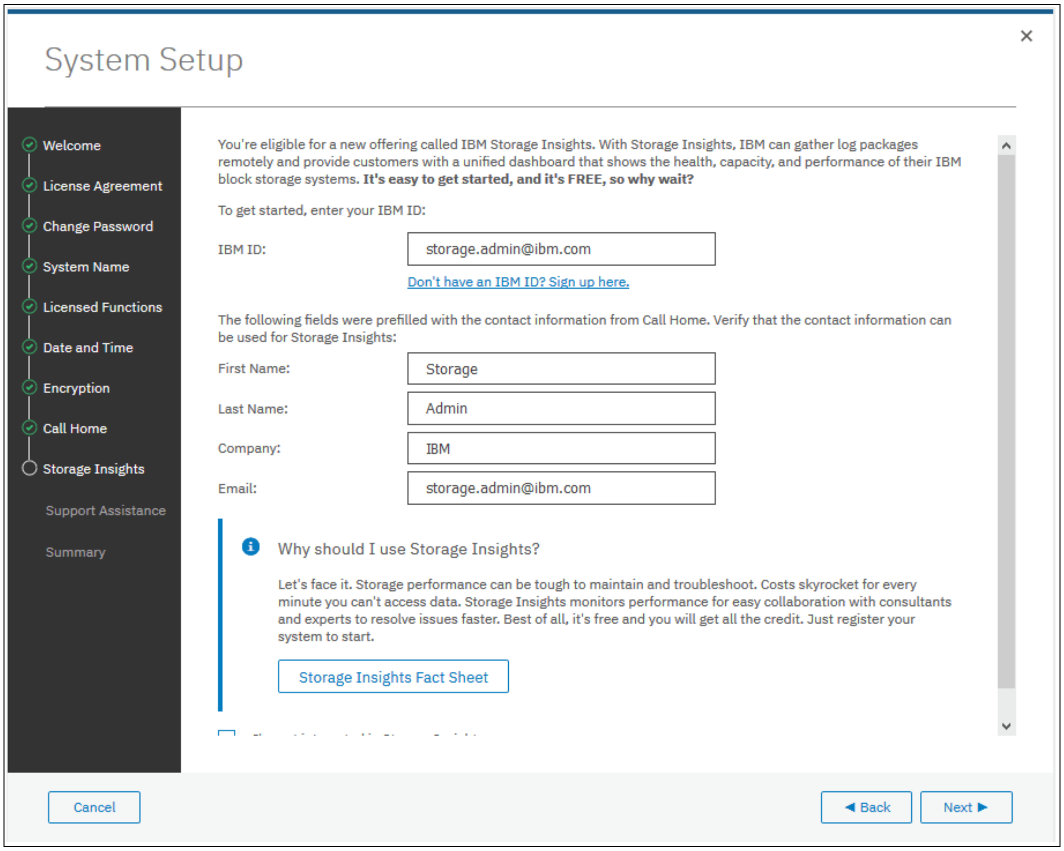


Figure 4-19 Storage Insights System Setup window

14. The Support Assistance window opens in which you are given the option to allow support personnel to work onsite only, or both remotely and onsite (see Figure 4-20). Refer to your organization's security policy to ensure that you set up a compliant environment.

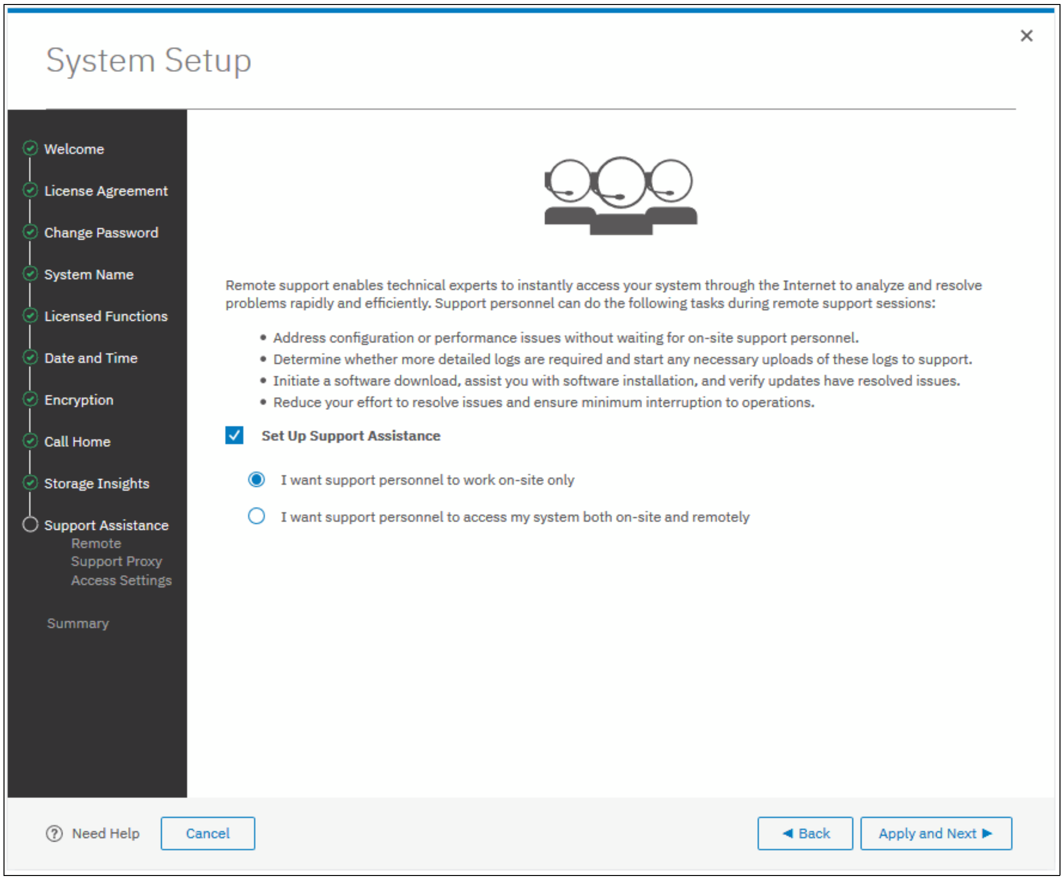


Figure 4-20 Support Assistance setup window

If you allow remote support, you are given IP addresses and ports of the remote support centers and an opportunity to provide proxy server details, if it is required to allow the connectivity. Also, you can allow remote connectivity at any time or only after obtaining a permission from the storage administrator.

15. In the summary window, you can review the system configuration. After you click **Finish**, a message is displayed that the setup is complete (Figure 4-21). After you click **Close**, you are redirected to the management GUI Dashboard, as shown in Figure 4-21.

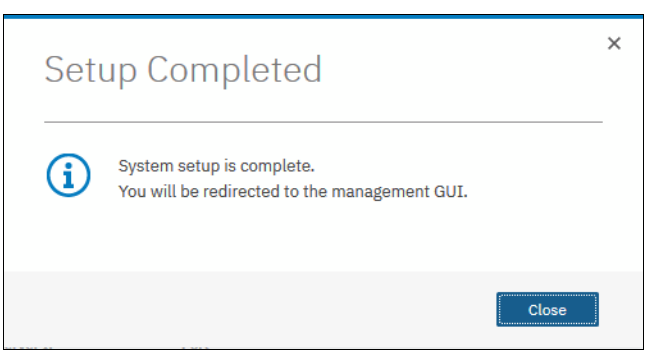


Figure 4-21 System setup: Completion

After the wizard completes, your system consists only of the control enclosure that includes the node canister that you used to initialize the system. If you have other control and expansion enclosures, you must add them to complete system setup.

For more information about how to add a control enclosure, see 4.3.2, “Adding a control enclosure” on page 107. For more information about how to add an expansion enclosure, see Chapter 5, “Graphical user interface” on page 131.

After you add the enclosures or if you have no expansion enclosures to install, the system setup process is complete.

Completing system setup means that all mandatory steps of the initial configuration are completed and you can start configuring your storage. Optionally, you can configure other features, such as user authentication, secure communications, and local port masking.

4.3.2 Adding a control enclosure

This procedure is the same whether you are configuring the system for the first time or expanding it later.

Before beginning this process, ensure that the new control enclosure is correctly installed and cabled to the existing system. Ensure that the Ethernet and Fibre Channel connectivity is correctly configured and that the enclosure is powered on.

Complete the following steps to add a control enclosure to the system:

1. In the GUI, go to **Monitoring** → **System** and click **Add Enclosure**, as shown in Figure 4-22.

Figure 4-22 System Overview: Add Enclosure button

> **Note:** If the Add Enclosure action does not appear, review the installation instructions to ensure the new enclosure is cabled and zoned correctly.

2. Choose an enclosure that you want to add from the list of available enclosures, as shown in Figure 4-23 on page 108 and click **Next**.

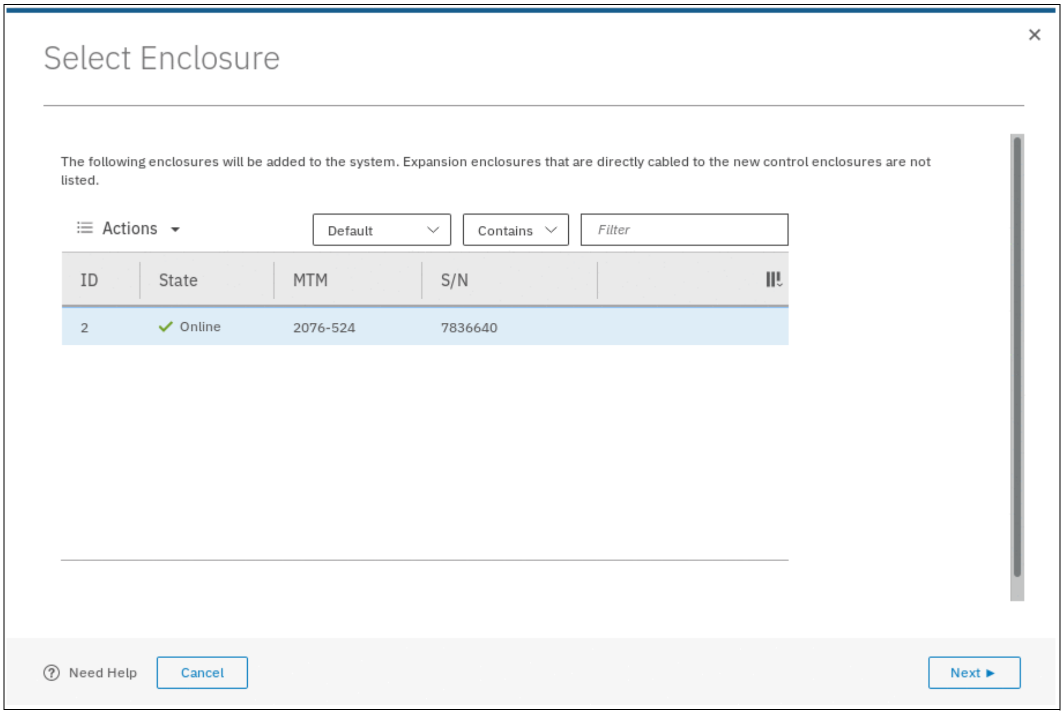


Figure 4-23 Select Enclosure window: available control enclosures

3. Review the summary in the next window and click **Finish** to add the control enclosure and all its expansions to the system.
4. After the control enclosure is successfully added to the system, a success message is displayed. Click **Close** to return to the **System Overview** view, as shown in Figure 4-24. The view displays two control enclosures configured on the system.

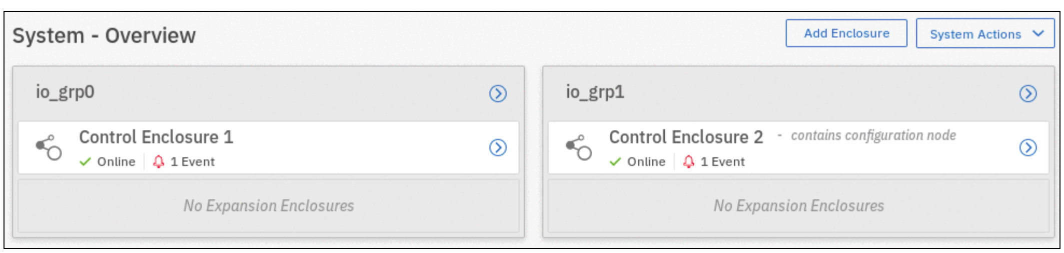


Figure 4-24 System Overview view with two control enclosures visible

4.4 Configuring user authentication

There are two methods of user authentication to control access to the GUI and to the CLI:

- *Local authentication* is performed within the Storwize V7000 system. Local GUI authentication is done with user name and password. Local CLI authentication is done either with an SSH public key or a user name and password.
- *Remote authentication* allows users to authenticate to the system using credentials stored on an external authentication service. This feature allows you to use user credentials and user groups 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.

4.4.1 Default superuser account

Every system has a default user called *superuser*. Superuser cannot be deleted or modified, except for changing the password and SSH key. Superuser is a *local* user and cannot be authenticated remotely.

> **Note:** Superuser is the only user allowed to log in to the Service Assistant Tool. It is also the only user allowed to run **`sainfo`** and **`satask`** commands through the CLI.

Superuser is a member of the SecurityAdmin user group, which is the most privileged role within the system.

The password for superuser is set during system setup. The superuser password can be reset to its default value of `passw0rd` using the technician port.

4.4.2 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 name, which must be unique across all users in one system.

User names 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 user name cannot begin or end with a blank space.

Passwords for local users can be up to 64 printable ASCII characters. There are no forbidden characters. However, passwords cannot begin or end with blanks.

When connecting to the CLI, encryption key authentication is attempted first with the user name 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.

> **Note:** If local authentication is used, user accounts must be created for each Storwize V7000 system. If you want to allow access for a user on multiple systems, you must define the user in each system.

4.4.3 Remote authentication

A *remote user* is authenticated by using identity information that is accessible by way of 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.

Configuring remote authentication with LDAP

IBM Spectrum Virtualize systems support the following types of LDAP servers:

- IBM Security Directory Server
- Microsoft Active Directory
- OpenLDAP

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, nor do they need an SSH key 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 using LDAP, start by enabling remote authentication:

1. Click **Settings** → **Security**, and select **Remote Authentication** and then, **Configure Remote Authentication**, as shown in Figure 4-25.

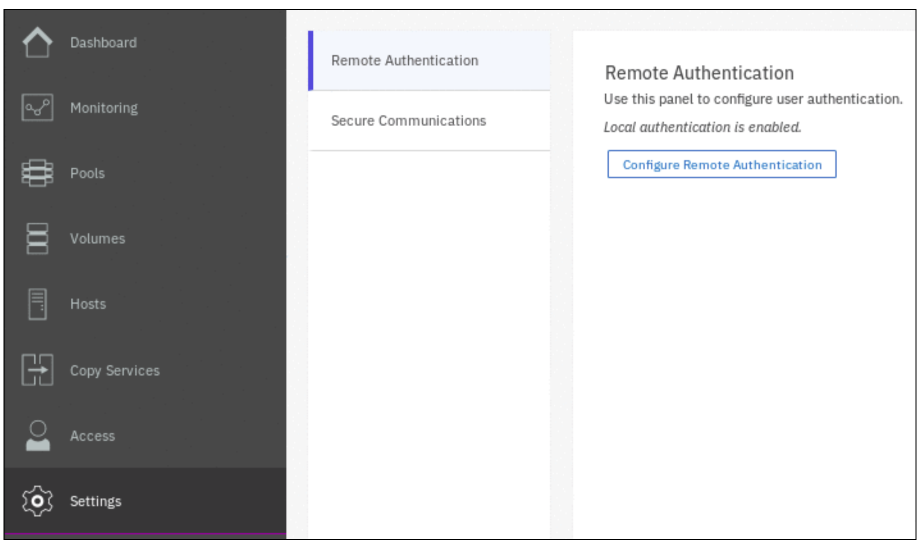


Figure 4-25 Configuring remote authentication

2. Enter the LDAP settings. These settings are not server-specific. They are applied to all LDAP servers configured in the system. Extra optional settings are available by clicking **Advanced Settings**. The following settings are available:

   – LDAP type:

      - **IBM Tivoli Directory Server** (for IBM Security Directory Server)
      - **Microsoft Active Directory**
      - **Other** (for OpenLDAP)

   In this example, we configure an OpenLDAP server, as shown in Figure 4-26 on page 111.

– Security

Choose between **None**, **SSL**, or **Transport Layer Security**. Using some form of security ensures that user credentials are encrypted before being transmitted. Select **SSL** to use LDAP over SSL (LDAPS) to establish secure connections using port 636 for negotiation and data transfer. Select **Transport Layer Security** to establish secure connections using Start TLS, allowing both encrypted and decrypted connections to be handled by the same port.

In this example, we choose **SSL**, as shown in Figure 4-26.

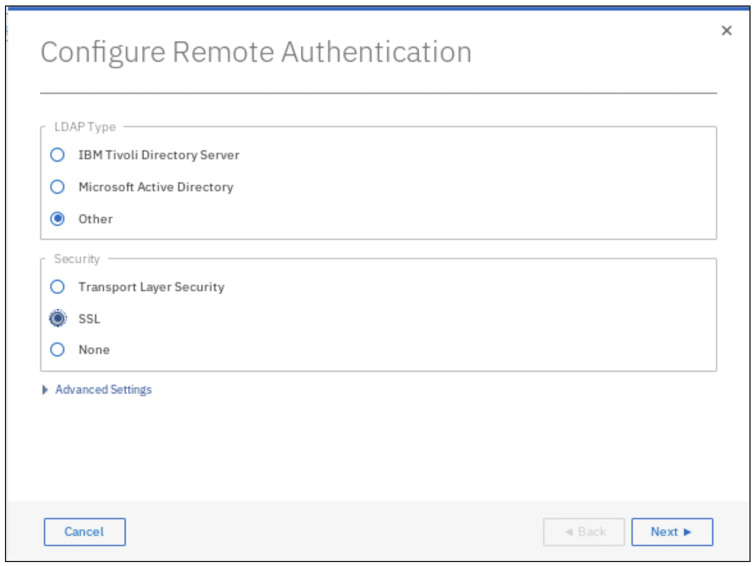


Figure 4-26 Configure remote authentication: Mandatory LDAP settings

– Service Credentials

Figure Figure 4-27 on page 112 shows advanced and optional setting. Consult the administrator of the LDAP server that you plan to use for the information to ensure the fields are filled correctly.

If your LDAP server supports anonymous bind, leave **Distinguished Name** and **Password** empty. Otherwise, enter the credentials of a user defined on the LDAP server with permission to query the LDAP directory. You can enter this information in the format of an email address (for example, `administrator@ssd.hursley.ibm.com`) or as a distinguished name (for example, `zcn=Administrator,cn=users,dc=ssd,dc=hursley,dc=ibm,dc=com`).

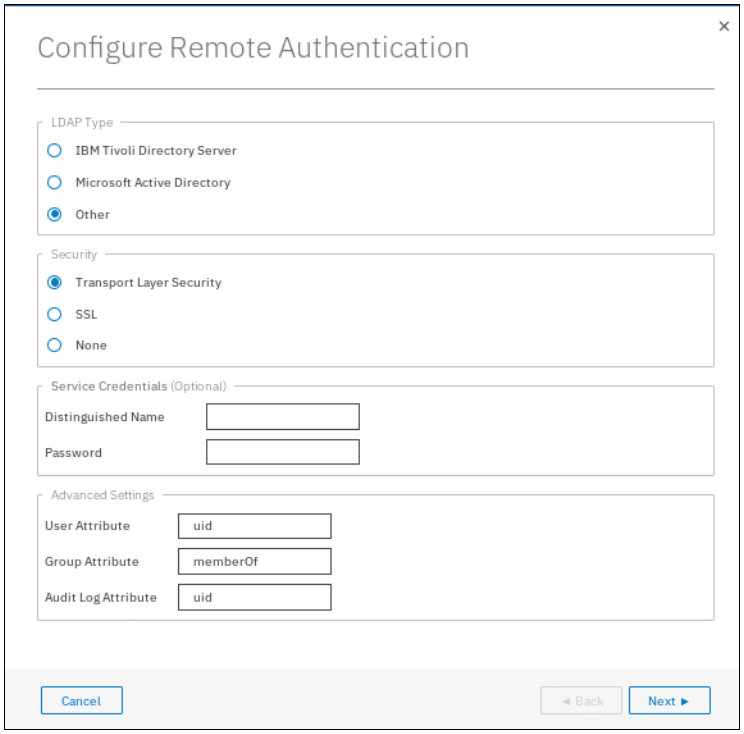


Figure 4-27 Configure remote authentication: Advanced LDAP settings

– User Attribute

  This LDAP attribute is used to determine the user name 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 `sAMAaccountName` for Microsoft Active Directory and to `uid` for IBM Security Directory Server and OpenLDAP.

– 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 Microsoft Active Directory and OpenLDAP and to `ibm-allGroups` for IBM Security Directory Server. For OpenLDAP implementations, you might need to configure the `memberOf` overlay if it is not in place.

– Audit Log Attribute

  This LDAP is attribute 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 Microsoft Active Directory and to `uid` for IBM Security Directory Server and OpenLDAP.

3. Enter the server settings for one or more LDAP servers, as shown in Figure 4-28 on page 113. To add servers, click the plus (+) icon. The following settings are available:

   – Preferred

     Authentication requests are processed by the preferred servers unless the connection fails or a user is not found. Requests are distributed across all preferred servers for load balancing. Select **Preferred** to set the server as a preferred server.

– 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=ssd,dc=hursley,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 **Finish** to save the settings.

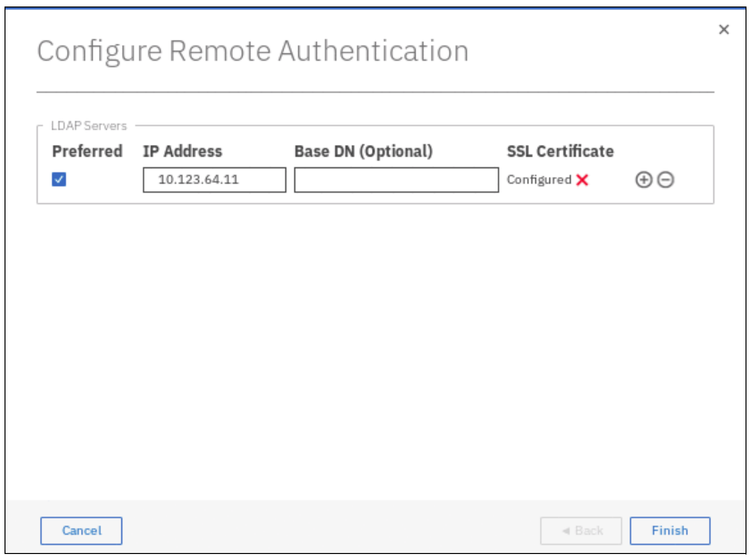


Figure 4-28 Configure remote authentication: Creating an LDAP server

Now that 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.

Complete the following steps to create a user group with remote authentication enabled:

1. Click **Access** → **Users** and select **Create User Group**, as shown in Figure 4-29.

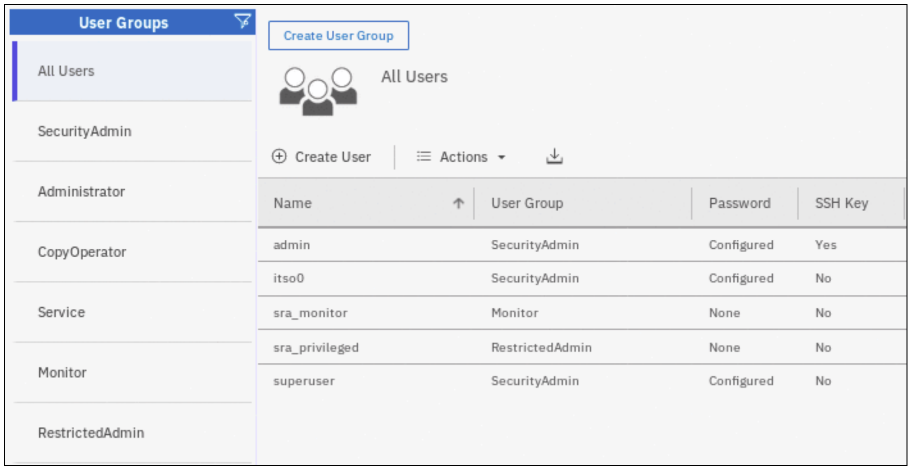


Figure 4-29 Option to create a user group

2. Enter the details for the new group. Select **Enable for this group** to enable remote authentication, as shown in Figure 4-30. Click **Create**.

> **Note:** This option is not available if LDAP authentication is not enabled.

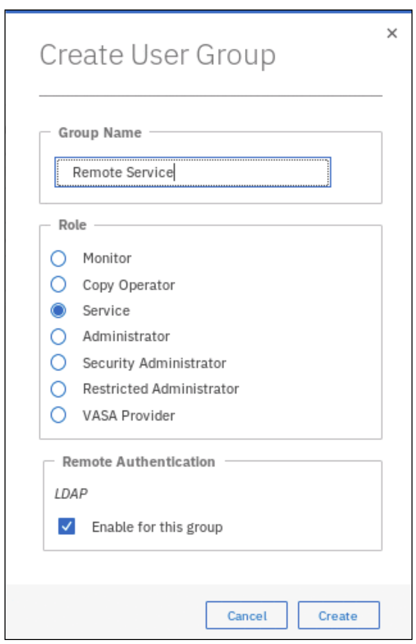


Figure 4-30 Creating a user group with remote authentication enabled

Complete the following steps to enable remote authentication for a default role:

1. Click **Access** → **Users**. Select the user group that you want to modify, click **Actions**, and then **Properties**. In this example, the Service user group is chosen, as shown in Figure 4-31.

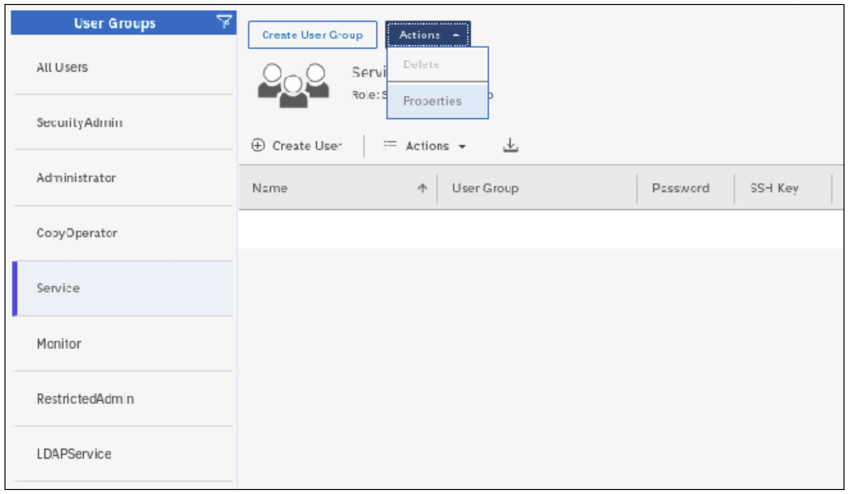


Figure 4-31 Changing the properties of a user group

2. Select **Enable for this group**, as shown in Figure 4-32.

> **Note:** This option is not available if LDAP is not enabled.

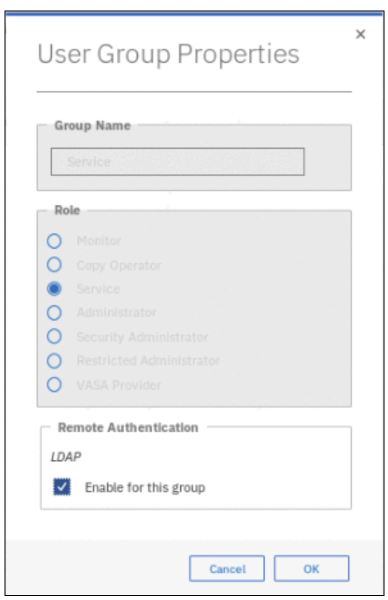


Figure 4-32 Enabling remote authentication for a default group

When you have at least one user group 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 you modified or created on the storage system.

- Each user that you want to authenticate remotely is a member of the LDAP user group configured for the system role.

The system is now ready to authenticate users using the LDAP server. To ensure that everything works correctly, test LDAP connectivity. To perform this test, click **Settings** → **Security** → **Remote Authentication**. Select **Global Actions** and then, **Test LDAP Connections**, as shown in Figure 4-33.

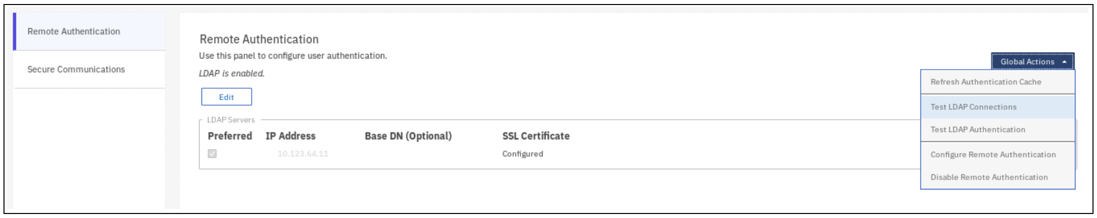


Figure 4-33 Option to test LDAP connections

If the test completes successfully, the system displays the message `CMMVC7075I The LDAP task completed successfully`. Otherwise, an error is logged in the event log.

The option to test a real user authentication attempt also is available. Click **Settings** → **Security** → **Remote Authentication**, and select **Global Actions** and then, **Test LDAP Authentication**, as shown in Figure 4-34.

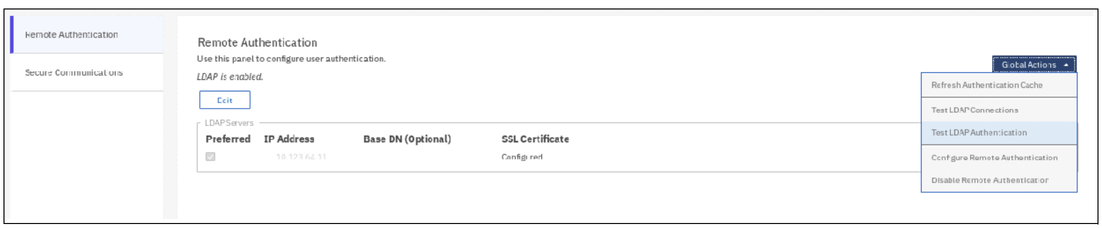


Figure 4-34 Option to test LDAP authentication

Enter the user credentials of a user defined on the LDAP server, as shown in Figure 4-35. Click **Test**.

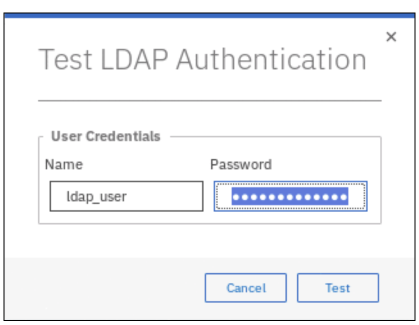


Figure 4-35 LDAP authentication test

Again, the message `CMMVC70751 The LDAP task completed successfully` is shown after a successful test.

Both the connection test and the authentication test must complete successfully to ensure that LDAP authentication works correctly. Assuming both tests succeed, users can log in to the GUI and CLI by using their network credentials.

A user can log in with their short name (that is, without the domain component) or with the fully qualified user name in the form of an email address.

4.4.4 User groups and roles

User groups are used to determine what tasks the user is authorized to perform. Each user group is associated with a single *role*. The role for a user group cannot be changed, but user groups (with one of the defined roles) can be created.

The rights of a user who belongs to a specific user group are defined by the role that is assigned to the user group. It is the role that defines what a user can or cannot do on the system.

Storwize V7000 provides six user groups and seven roles by default, as listed in Table 4-1. The VasaProvider role is not associated with a default user group.

Note: The VasaProvider role is used to allow VMware to interact with the system when implementing Virtual Volumes. Avoid using this role for users who are not controlled by VMware.

Table 4-1 Default user groups and roles

User group	Role
SecurityAdmin	SecurityAdmin
Administrator	Administrator
CopyOperator	CopyOperator
Service	Service
Monitor	Monitor
RestrictedAdmin	RestrictedAdmin
-	VasaProvider

4.5 Configuring secure communications

During system initialization, a *self-signed* 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 certificate authority. A browser maintains a list of trusted certificate authorities, identified by their *root* certificate. The root certificate must be included in this list in order for the signed certificate to be trusted. If it is not, the browser presents security warnings.

To see the details of your current system certificate, click **Settings** → **Security** and select **Secure Communications**, as shown in Figure 4-36.

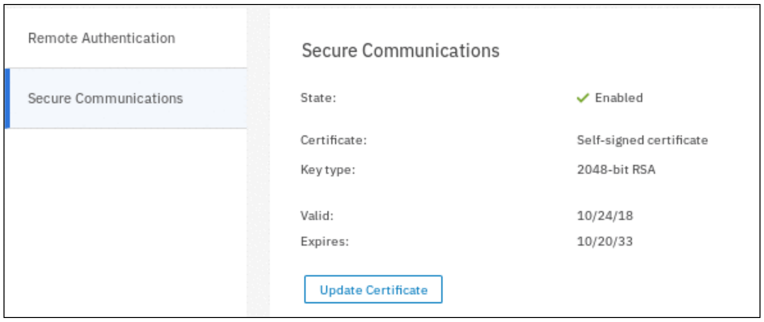


Figure 4-36 Accessing the Secure Communications window

IBM Spectrum Virtualize systems allow you to generate a self-signed certificate or to configure a signed certificate.

4.5.1 Configuring a signed certificate

Complete the following steps to configure a signed certificate:

1. Select **Update Certificate** on the Secure Communications window.
2. Select **Signed certificate** and enter the details for the new certificate signing request. All fields are mandatory, except for the email address. Figure 4-37 shows some values as an example.

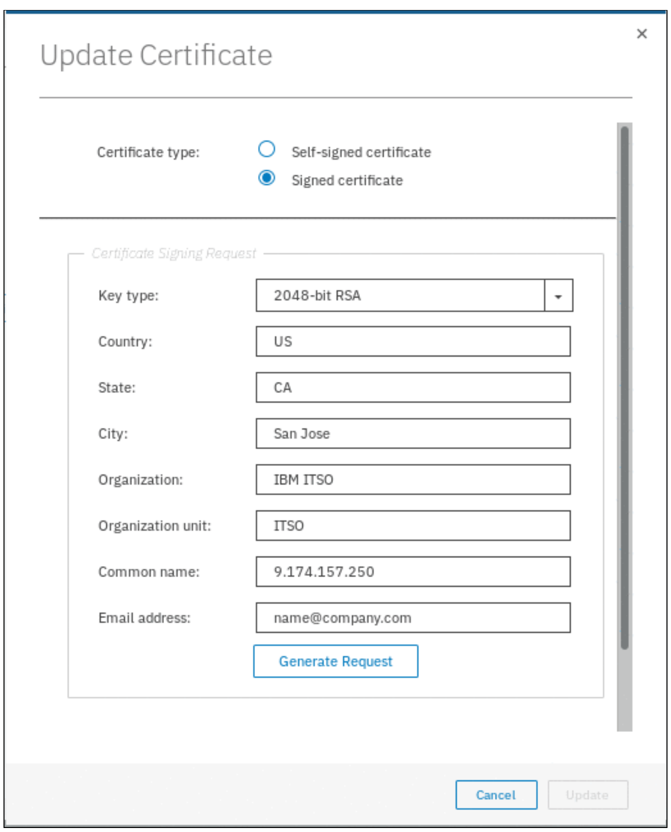


Figure 4-37 Generating a certificate request

> **Attention:** Before generating a request, ensure that your current browser does not have restrictions on the type of keys that are used for certificates. Some browsers limit the use of specific key-types for security and compatibility reasons.
>
> Also, consult your organization's security policy to ensure that the key type you are configuring is compliant.

Click **Generate Request**.

3. Save the generated request file. 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 to signing by the proper certificate authority. However, this process invalidates the previous certificate request and prevents the installation of the signed certificate associated with the original request.

4. Submit the request to the certificate authority to receive a signed certificate.
5. When you receive the signed certificate, select **Update Certificate** on the Secure Communications window again.
6. Click the folder icon to upload the signed certificate, as shown in Figure 4-38. Click **Update**.

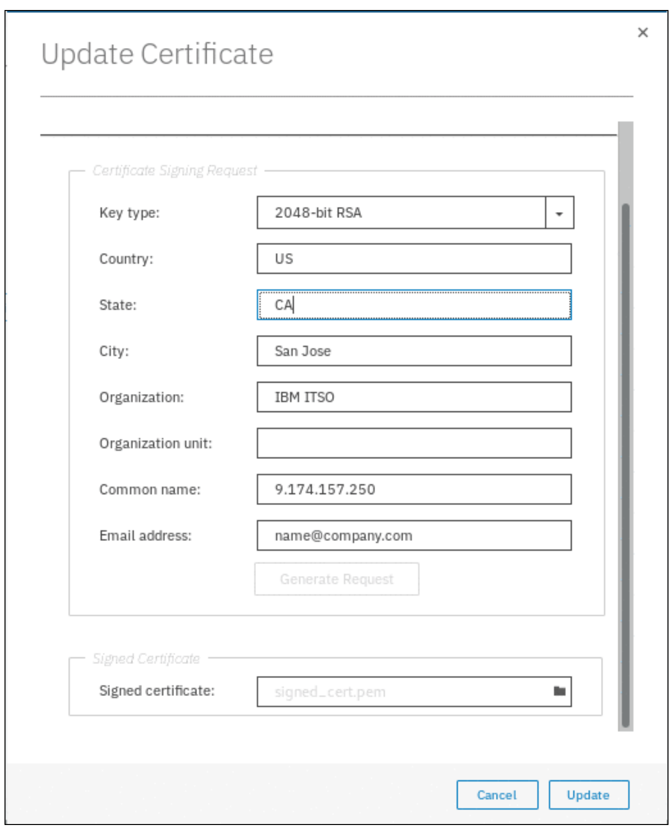


Figure 4-38 Installing a signed certificate

7. You are prompted to confirm the action. Click **Yes** to proceed. The signed certificate is installed. Close the browser and wait approximately 2 minutes before reconnecting to the management GUI.

4.5.2 Generating a self-signed certificate

Complete the following steps to generate a self-signed certificate:

1. Select **Update Certificate** on the Secure Communications window.
2. Select **Self-signed certificate** and enter the details for the new certificate. Key type and validity days are the only mandatory fields. Figure 4-39 shows some values as an example.

> **Attention:** Before creating a self-signed certificate, ensure that your browser does not have restrictions on the type of keys that are used for certificates. Some browsers limit the use of specific key-types for security and compatibility reasons.
>
> Consult your organization's security policy to ensure that the key type you are configuring is compliant.

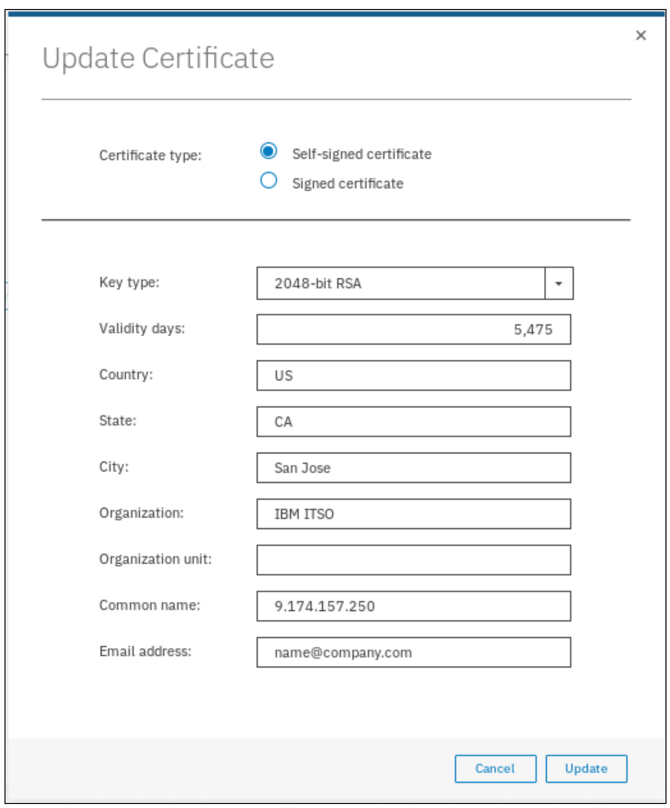


Figure 4-39 Generating a self-signed certificate

3. Click **Update**.

4. You are prompted to confirm the action, click **Yes** to proceed. The self-signed certificate is generated immediately. Close the browser and wait approximately 2 minutes before reconnecting to the management GUI.

4.6 Configuring local Fibre Channel port masking

This section provides information about configuring local port masking of Fibre Channel ports in a clustered system.

With Fibre Channel port masking, you control the use of Fibre Channel ports. You can control whether the ports are used to communicate to other nodes within the same local system, and if they are used to communicate to nodes in partnered systems. Fibre Channel port masking does not affect host or storage traffic. It is applied only to node-to-node communications within a system and replication between systems.

> **Note:** This section applies to local port masking only. For more information about configuring the partner port mask for intercluster node communications, see 11.6.4, "Remote copy intercluster communication" on page 519.

The setup of Fibre Channel port masks is useful when you have more than four Fibre Channel ports on any node in the system because it saves setting up many SAN zones on your switches. Fibre Channel I/O ports are logical ports, which can exist on Fibre Channel platform ports or on FCoE platform ports. B using a combination of port masking and fabric zoning, you can ensure that the number of logins per node is not more than the supported limit. If a canister receives more than 16 logins from another node, it causes node error 860.

4.6.1 Planning for local port masking

The system has two Fibre Channel port masks. The local port mask controls connectivity to other nodes in the same system, and the partner port mask control connectivity to nodes in remote, partnered systems. By default, all ports are enabled for both local and partner connectivity.

The port masks apply to all nodes on a system. It is not possible to have different port mask on different nodes of the same cluster. However, you do not have to have the same port mask on partnered systems.

> **Note:** The use of the **`lsfabric`** command shows all of the paths that are possible in IBM Spectrum Virtualize (as defined by zoning) independent of their usage. Therefore, the command output includes paths that are used because of port masking.

A port mask is 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, and the system attempts to send and receive traffic on that port. If it is "0", the system does not send or receive traffic on the port. If not enough digits are in the configuration string to set a value for the specific port, that port is disabled for traffic.

For example, if the local port mask is set to `101101` on a node with eight Fibre Channel ports, ports 1, 3, 4 and 6 can connect to other nodes in the system. Ports 2, 5, 7, and 8 do not have connections. On a node in the system with four Fibre Channel ports, ports 1, 3, and 4 can connect to other nodes in the system.

The Fibre Channel ports for the system can be viewed by navigating to **Settings** → **Network** and opening the **Fibre Channel Ports** menu, as shown in Figure 4-40. Port numbers refer to the Fibre Channel I/O port IDs.

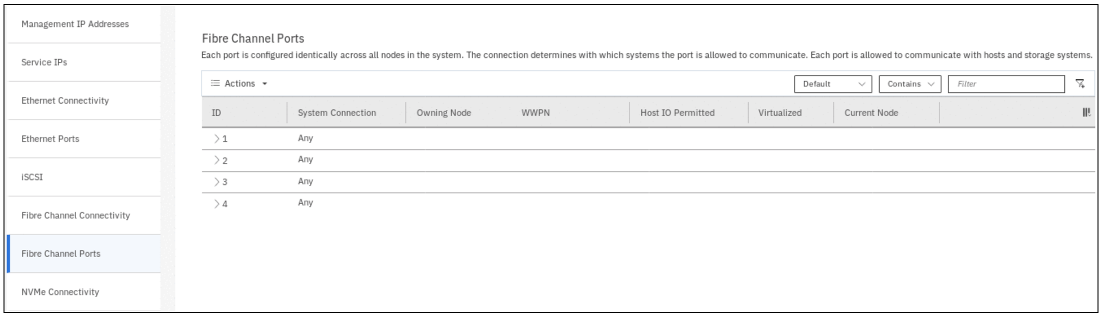


Figure 4-40 Viewing the Fibre Channel ports

To see the Fibre Channel connectivity of the system, navigate to **Settings** → **Network** and open the **Fibre Channel Connectivity** menu, as shown in Figure 4-41. The window displays the connectivity between nodes and other storage systems and hosts that are attached through the Fibre Channel network.

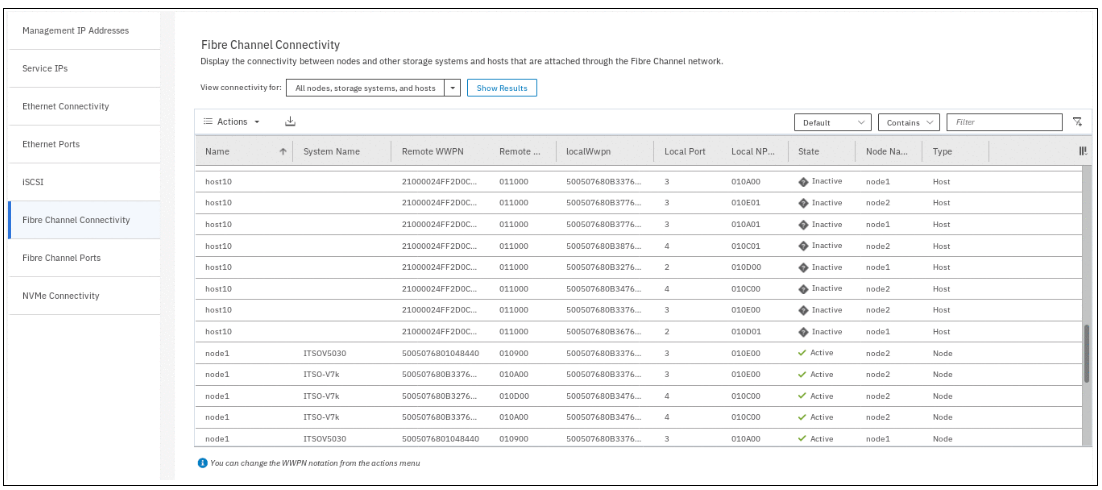


Figure 4-41 Fibre Channel connectivity of the system

When replacing or upgrading your node hardware to newer models, consider that the number of Fibre Channel ports and their arrangement might be changed. If so, make sure that any configured port masks are still valid for the new configuration.

4.6.2 Setting the local port mask

Consider the example fabric port configuration that is shown in Figure 4-42.

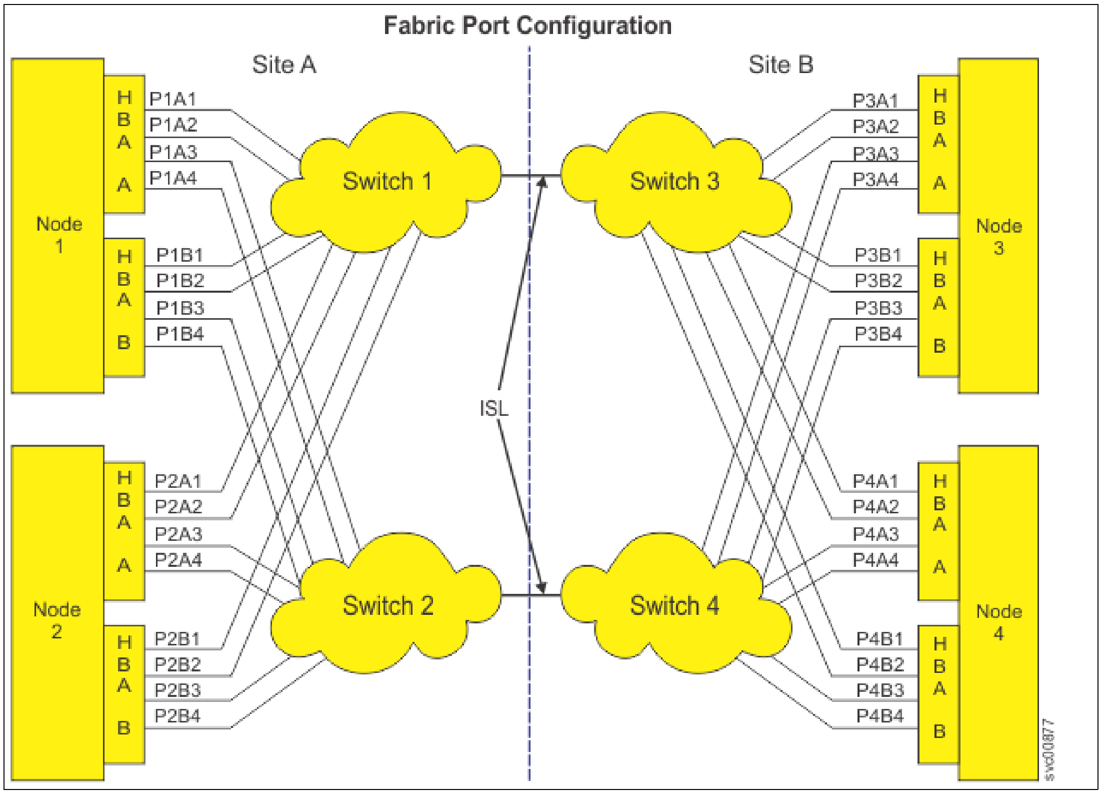


Figure 4-42 Example fabric port configuration

This configuration includes the following characteristics:

- It consists of four nodes, where nodes 1 and 2 are in a system at site A and nodes 3 and 4 are in a system at site B.
- Each site has two switches (switch 1 and 2 at site A; switch 3 and 4 at site B).
- Each node has two adapters (A and B).
- Each adapter has four ports where the port is named `P nodeid adapterid portnumber`; for example, port 3 on adapter B of node 4 is named `P4B3`.

The positions in the mask represent the Fibre Channel I/O port IDs with ID 1 in the rightmost position. In this example, ports A1, A2, A3, A4, B1, B2, B3, and B4 correspond to FC I/O port IDs 1, 2, 3, 4, 5, 6, 7 and 8.

To set the local port mask, use the **`chsystem`** command. Limit local node-to-node communication to ports A1, A2, A3, and A4 by applying a port mask of 00001111 to both systems, as shown in Example 4-1.

Example 4-1 Setting a local port mask using the chsystem command

```
IBM_Storwize:ITSO:superuser>chsystem -localfcportmask 00001111
IBM_Storwize:ITSO:superuser>
```

4.6.3 Viewing the local port mask

To view the local port mask for the system, use the **lssystem** command, as shown in Example 4-2.

Example 4-2 Viewing the local port mask

```
IBM_Storwize:ITSO:superuser>lssystem
id 000001003D600126
name ITSO
location local
partnership
...
...
local_fc_port_mask 0000000000000000000000000000000000000000000000000000000000001111
partner_fc_port_mask 1111111111111111111111111111111111111111111111111111111111111111
...
```

4.7 Other administrative procedures

In this section, we discuss other administrative procedures.

4.7.1 Removing a control enclosure from a clustered system

Removing a control enclosure from a system causes a loss of access to drives in this control enclosure and any expansions enclosures connected to this control enclosure. If not planned and executed carefully and correctly, this procedure can cause disruption in access to storage or data loss. Make sure that you have current and verified data backups before removing a control enclosure from a system. Follow carefully the enclosure removal procedure that is provided at this IBM Knowledge Center web page.

4.7.2 Shutting down the system

You can safely shut down an Storwize V7000 cluster by using the GUI or CLI.

After you shut down the entire cluster, you need to power it on manually to start the system again. Make sure that someone is available with physical access to the IBM Spectrum Virtualize hardware who can start the system after it is shut down.

> **Note:** For systems with enabled encryption, ensure that the cluster can access at least one valid encryption key provider. Access to an encryption key provider (USB key or a key server) is required at start time to unlock encrypted data.
>
> Also, never shut down your IBM Storwize V7000 system by powering off the PSUs, removing both PSUs, or removing both power cables from the enclosure. It can lead to data loss.

Before shutting down the cluster, make sure that all hosts that have volumes mapped from the system are prepared for the storage system shutdown. This can be achieved by several methods, as shown in the following examples:

- Shutting down the host. This is the safest option.

- Unmounting all file systems that were created on IBM Spectrum Virtualize volumes (for file systems created directly on IBM Spectrum Virtualize volumes) or bringing offline all logical volume groups by using volumes that are presented by the IBM Spectrum Virtualize system.
- Accepting loss of access to storage for volumes that are mirrored at the operating system level.

For volumes that are mirrored at the operating system level, loss of one of the data copies triggers errors in the operating system and causes a loss of redundancy because the data copies go out of sync.

Note: Some applications (for example, databases), can use a volume that is not mounted as a file system. Make sure that no volumes that are presented by the IBM Spectrum Virtualize are in use on a host if you want to shut down the storage system but not the host.

Before shutting down the system, ensure that you stopped all FlashCopy mappings, remote copy relationships, data migration operations, and forced deletions.

Storwize V7000 control enclosures contain batteries that provide backup power to the system to protect against unforeseen loss of power. When AC power to the enclosure is interrupted for more than approximately 10 seconds, the system starts system state dump procedure, which includes saving cached data to an internal drive.

When the process to save critical data starts, the system stops handling I/O requests from the host applications, and Metro Mirror and Global Mirror relationships go offline. The system powers off when the saving of the critical data completes. If AC power is restored after the system state dump starts, the dump continues to completion. The system then restarts.

Note: Storwize V7000 expansion canisters do not cache volume data or store state information in volatile memory. Therefore, they do not require battery power. If AC power to both power supplies in an expansion enclosure fails, the enclosure powers off. When AC power is restored to at least one of the power supplies, the controller restarts without operator intervention.

In a fully redundant system with two batteries and two canisters, there is enough charge in the batteries to support saving critical data from both canisters to a local drive twice. In a system with a failed battery, there is enough charge in the remaining battery to support saving critical data from both canisters to a local drive once.

If both node canisters shut down without writing the cache and state data to the local drive, the system is unable to restart without an extended service action. The system configuration must be restored. If any cache write data is lost, volumes must be restored from a backup. Therefore, it is important not to remove the canisters or the power supply units from the control enclosures unless directed to do so by the service procedures. Removing either of these components might prevent the node canister from writing its cache and state data to the local drive.

In case of an imminent power loss, strive to shut down the system cleanly, without triggering the data dump procedure. This process preserves battery charges for actual emergencies.

When the AC power is restored after a power outage that causes both canisters to save their critical data, the system restarts only when the batteries have sufficient charge to power both canisters for the duration of saving the critical data again.

If a second AC power outage occurs before the batteries have completed charging, the system starts in service state and does not permit I/O operations to be restarted until the batteries are half charged, which provides enough capacity to perform a clean system shut down in case of another power loss. The recharging process takes approximately 30 minutes.

The system automatically schedules maintenance battery discharge cycles to extend the lifetime of the batteries and ensure that the system can accurately measure the charge in the batteries. Discharge cycles ensure that the batteries have sufficient charge to protect the Storwize V7000 system. For more information about maintenance discharge cycles, see this IBM Knowledge Center web page.

To shut down your Storwize V7000 system, complete the following steps:

1. Ensure that no hosts are using the storage system, and that you stopped all FlashCopy mappings, remote copy relationships, data migration operations, and forced deletions.
2. In the GUI, click **Monitoring** → **System**. Then, click **System Actions** and choose **Power off System**, as shown in Figure 4-43.

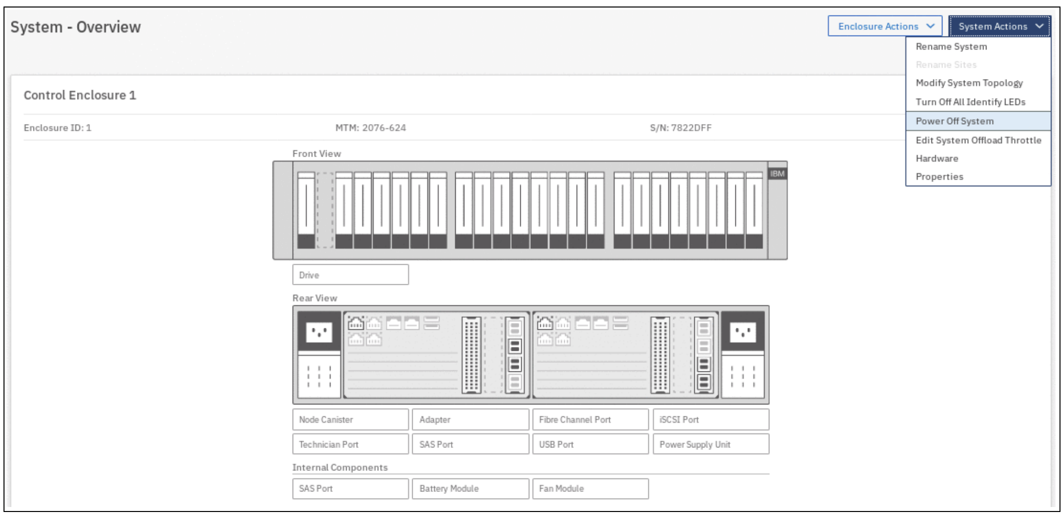


Figure 4-43 Initiating system power off

3. A confirmation window opens, as shown in Figure 4-44.

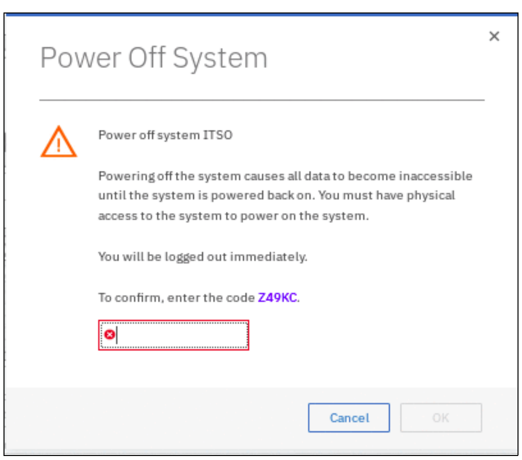


Figure 4-44 Power off warning window

The OK button is not active until you copy the displayed confirmation code into the provided text field. This step makes it more difficult to shut down the system by accident.

4. Enter the generated confirmation code and click **OK** to begin the shutdown process.

4.7.3 Changing the system topology to HyperSwap

The IBM HyperSwap function is a high availability feature that provides dual-site, active-active access to a volume. You can create an IBM HyperSwap topology system configuration where each I/O group in the system is physically on a different site. When used with HyperSwap volumes, these configurations can be used to maintain access to data on the system if site-wide outages occur.

To change the system topology to HyperSwap, complete the following steps:

1. In the GUI, click **Monitoring** → **System**. Then, click **System Actions** and choose **Modify System Topology**, as shown in Figure 4-45.

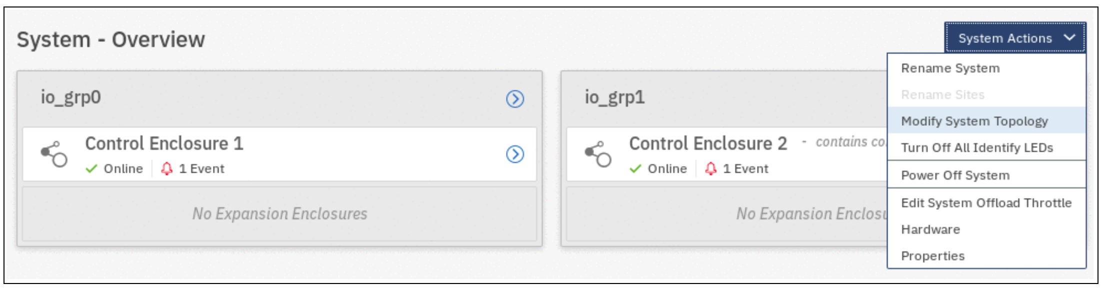


Figure 4-45 Starting the Modify System Topology wizard

2. The Modify Topology Wizard welcome window opens, as shown in Figure 4-46.

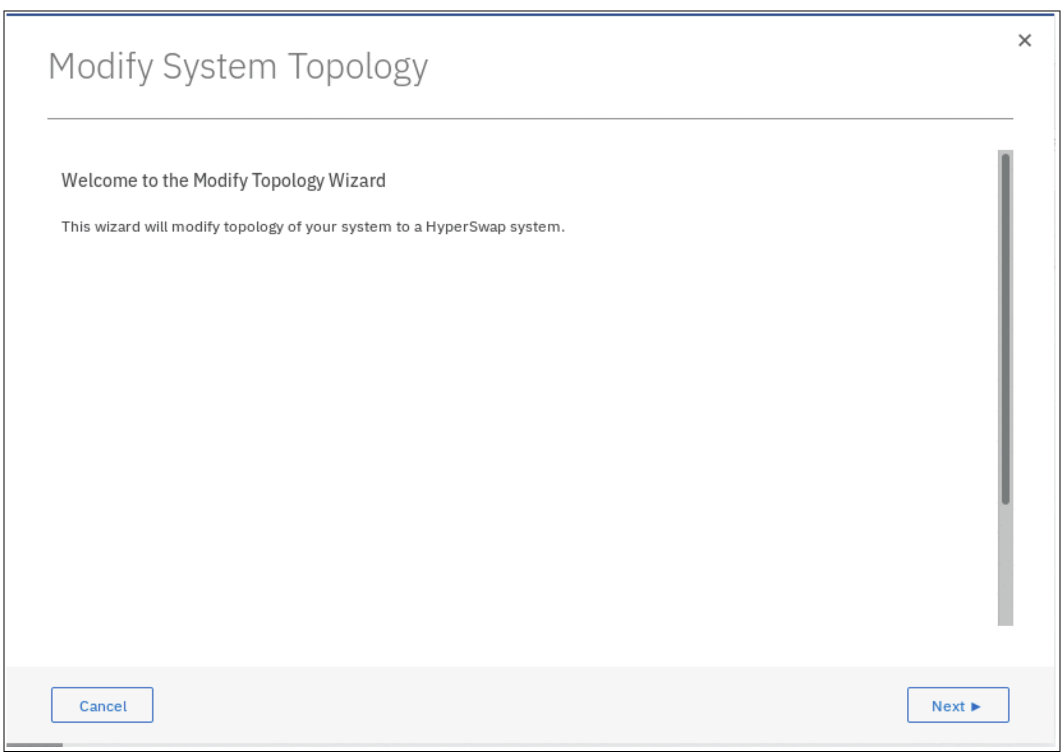


Figure 4-46 Modify Topology Wizard welcome window

3. Click **Next**.
4. Assign names to sites (see Figure 4-47). Click **Next**.

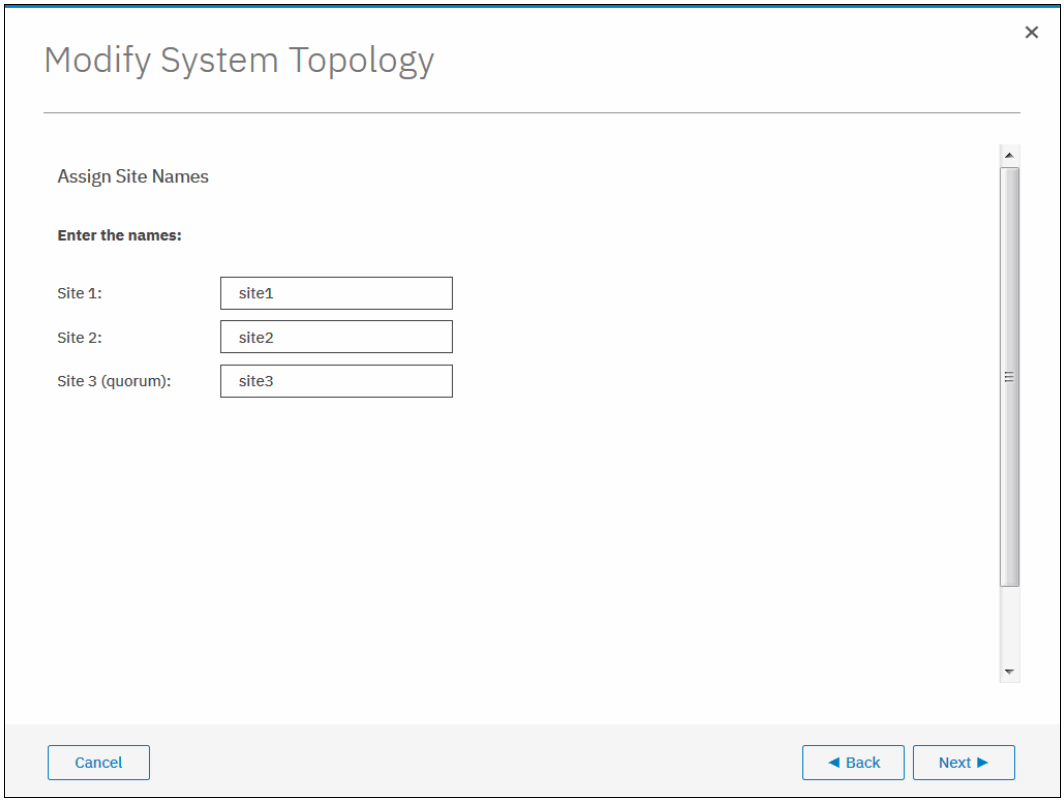


Figure 4-47 Modifying system topology – specifying site names

5. Assign I/O groups to sites. Click the marked icons in the center of the window to swap site assignments, as shown in Figure 4-48 on page 129. Click **Next** to proceed.

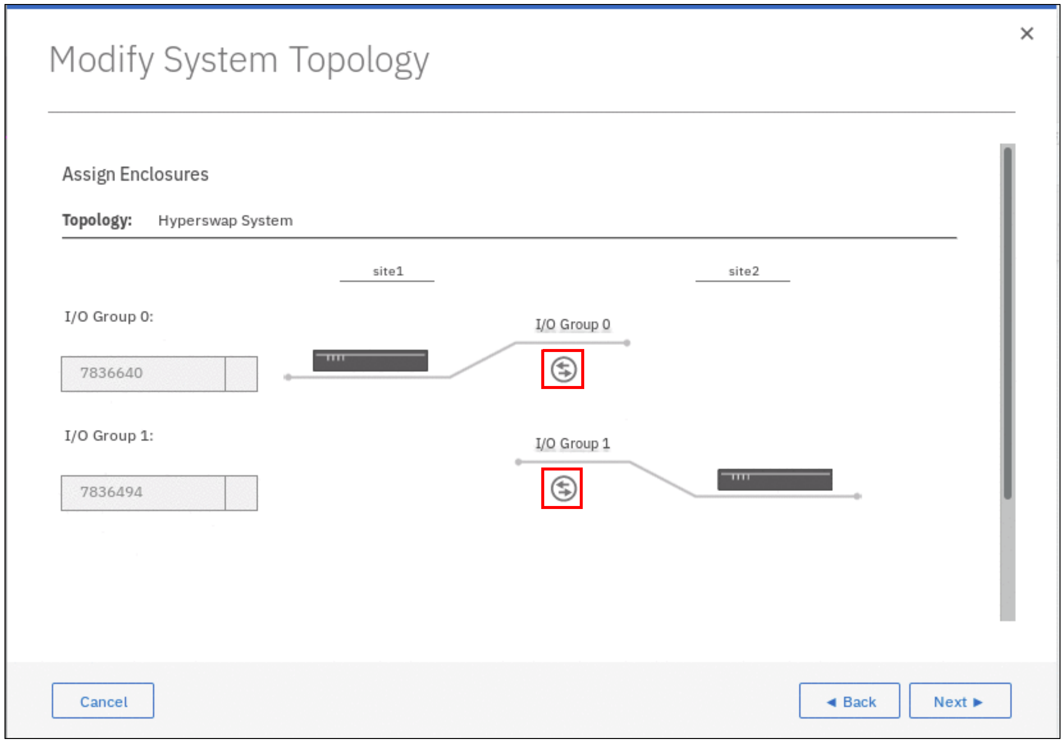


Figure 4-48 Specifying system topology

6. Set the bandwidth between the sites, as shown in Figure 4-49. You must specify the total bandwidth between the sites and what percentage of this bandwidth is reserved for background copies. Click **Next** to proceed.

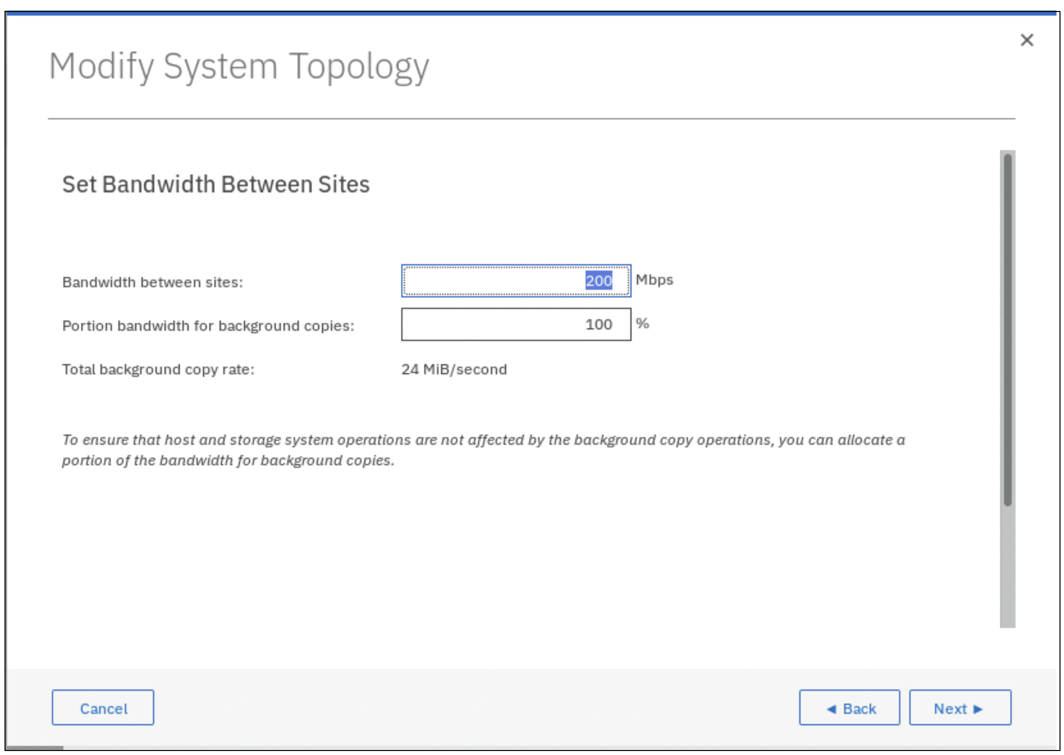


Figure 4-49 Setting the bandwidth between the sites

7. Review the summary and click **Finish** to complete system topology change, as shown in Figure 4-50.

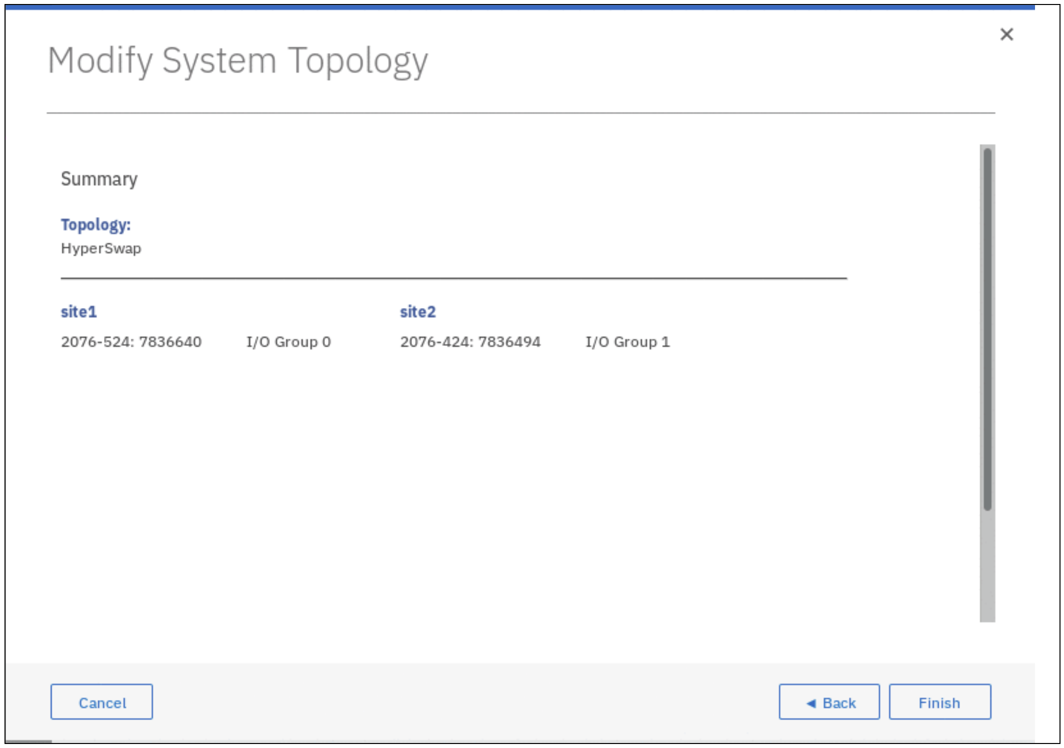


Figure 4-50 Modifying system topology Summary window

8. Before creating HyperSwap volumes, complete the following steps:

   a. Assign hosts to sites. You can use the **`chost -site site_id host_name`** CLI command to perform this step.

   b. Assign virtualized iSCSI-based back-end storage to sites. You can use the **`chcontroller –site site_id storage_name`** CLI command to perform this step.

      Alternatively, you can run Add iSCSI External Storage wizard. To access the wizard, select **Pools** → **External Storage** → **Add iSCSI External Storage**.

5

Graphical user interface

This chapter describes an overview of the IBM Spectrum Virtualize graphical user interface (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 IBM Storwize V7000 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, encryption). For more information, see Chapter 13, “RAS, monitoring, and troubleshooting” on page 673.

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 6, “Storage pools” on page 191), volumes (Chapter 7, “Volumes” on page 241), hosts (Chapter 8, “Hosts” on page 317), and Copy Services (Chapter 11, “Advanced Copy Services” on page 435) are described in separate chapters.

Demonstration: The IBM Client Demonstration Center includes a demonstration of the V8.2 GUI. For more information, see the IBM Client Demonstration Center (log in required).

This chapter includes the following topics:

5.1 Normal operations by using the GUI

This section describes useful tasks that use the GUI that help administrators to manage, monitor, and configure the IBM Storwize V7000 as quickly as possible. For example, we configure the IBM Storwize V7000 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 at any time. However, no locking mechanism exists, so be aware that if two users change the same object at the same time, the last action that is entered from the GUI is the action that takes effect.

IBM Spectrum Virtualize V8.2 brings some small improvements and feature changes to the Storwize V7000 GUI, following on from the major GUI redesign that occurred in V8.1. This chapter highlights these additions and limitations when compared with the previous version of V8.1.

> **Important:** Data entries that are made through the GUI are case-sensitive.
>
> You must enable Java Script in your browser. For Mozilla Firefox, Java Script is enabled by default and requires no other configuration steps. For more information, see these instructions at IBM Knowledge Center.

5.1.1 Access to GUI

To access the IBM Storwize V7000 GUI, enter into the address line of your web browser the IP address that was set during the initial setup process. You can connect from any workstation that can communicate with the system. The login window opens (see Figure 5-1).

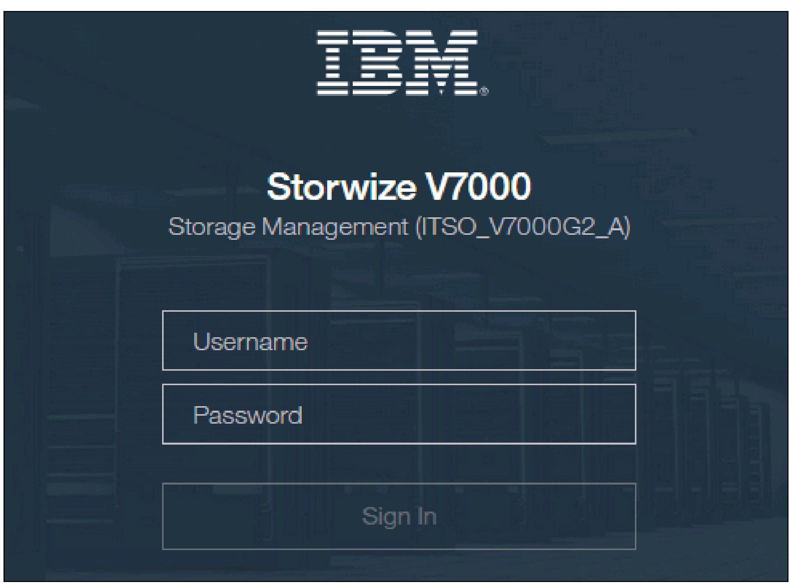


Figure 5-1 Log in window of Storwize V7000

Note: If you log in to the GUI by using a node that is the configuration node, you receive another option, “Service Assistant Tool”. Clicking this option takes you to the service assistant instead of the cluster GUI, as shown in Figure 5-2.

Figure 5-2 Log in Window of V7000 when connected to the configuration node

It is preferable 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 done by them. The *Superuser* account should be used for initial configuration only.

After a successful login, the V8.2 welcome window shows up with the new system dashboard (see Figure 5-3).

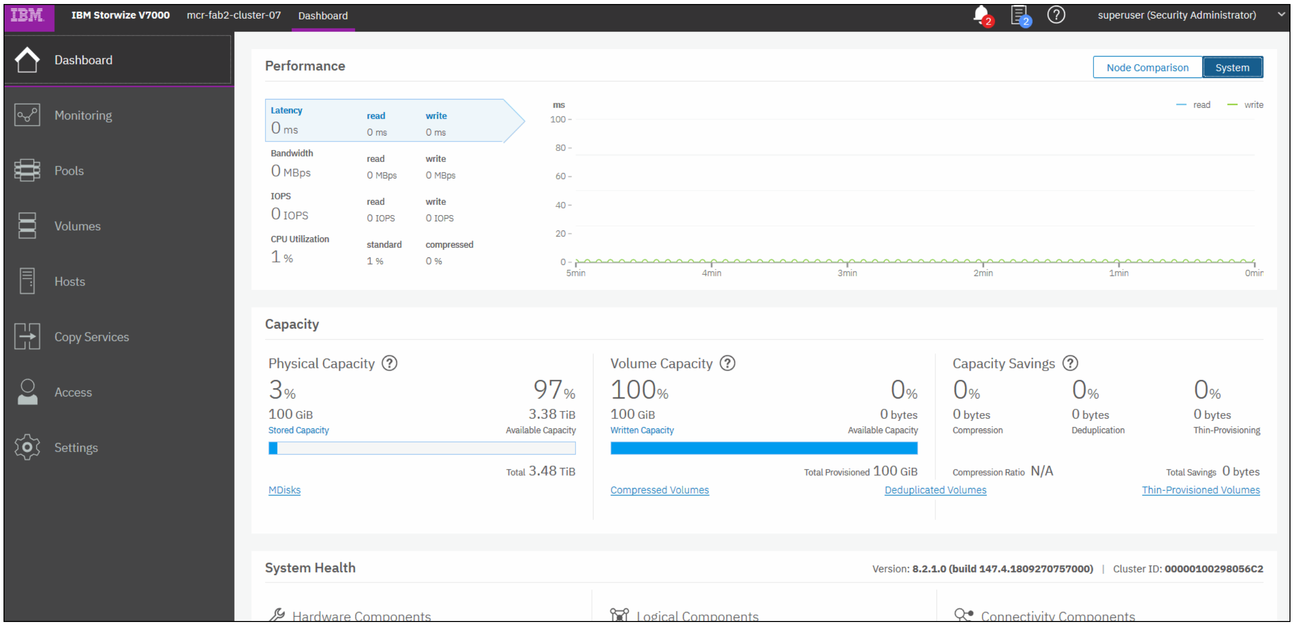


Figure 5-3 Welcome page with new dashboard

The Dashboard is divided into three sections:

- Performance

  This section provides important information about latency, bandwidth, IOPS, and CPU utilization. All of 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 5-4). The performance graph is updated with new data every 5 seconds.

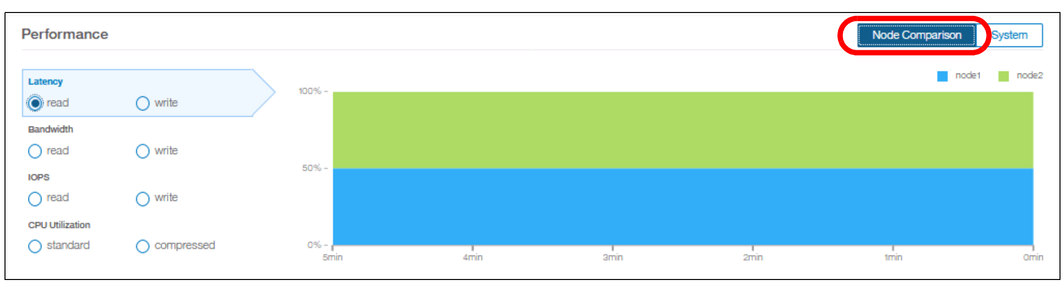


Figure 5-4 Performance statistics

- Capacity

  This section shows the current utilization of attached storage and its usage. Apart from the physical capacity, it also shows volume capacity and capacity savings. You can select the **Deduplicated Volumes** or **Thin Provisioned Volumes** options to display a complete list of either option in the volumes tab. New with V8.2 is the “Overprovisioned External Storage” section (highlighted in the red box in Figure 5-5 on page 135) which appear only when attached storage systems that are overprovisioned are present.

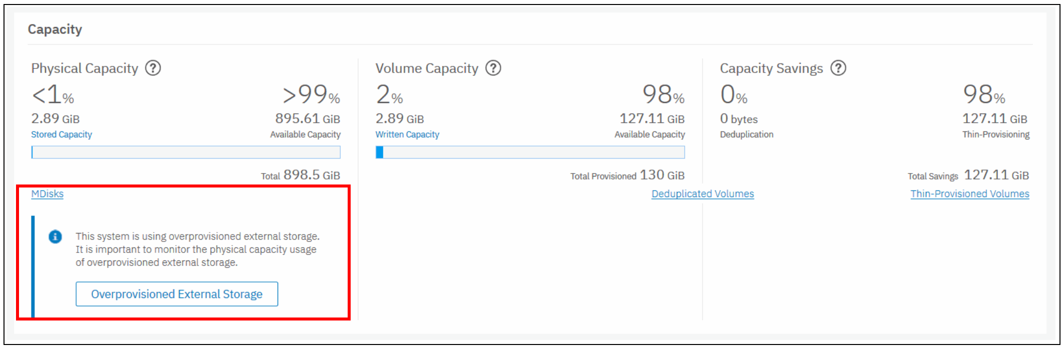


Figure 5-5 Capacity overview

Selecting this option provides a list of Overprovisioned External Systems that you can then click to see a list of related MDisks and Pools, as shown in Figure 5-7.

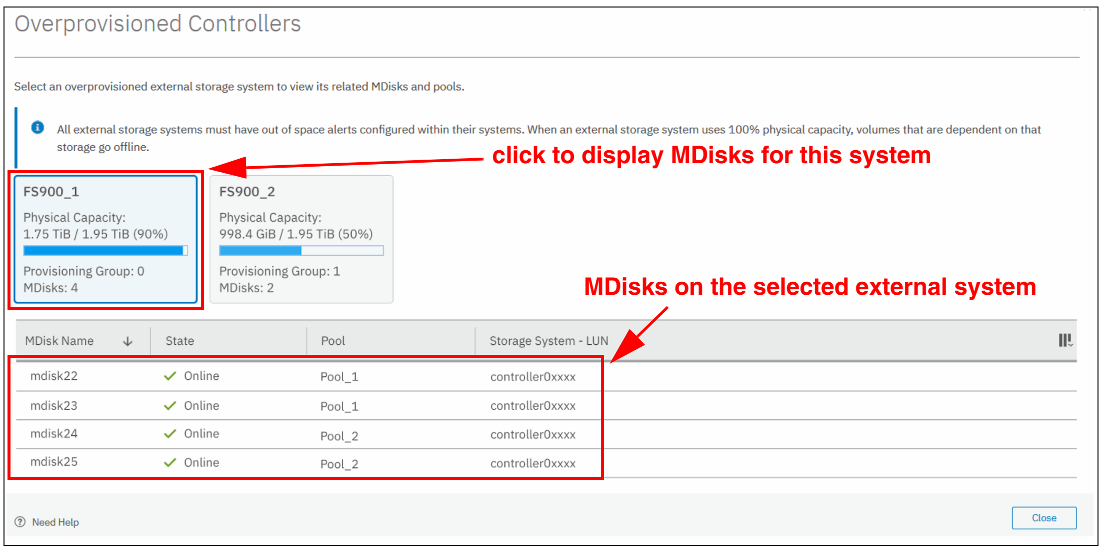


Figure 5-6 List showing overprovisioned external storage

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. When you click **Expand**, each component is listed as a subgroup. You can then navigate directly to the section of GUI where the component that you are interested in is managed from (see Figure 5-7).

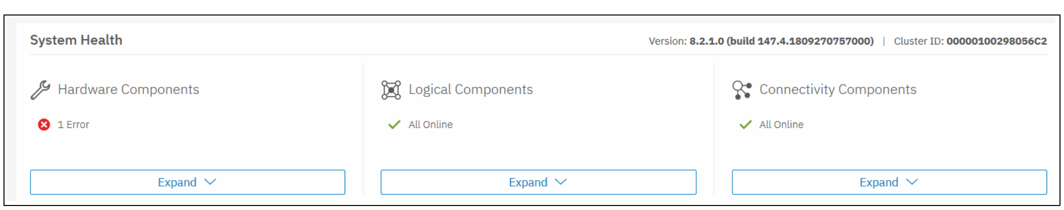


Figure 5-7 System Health overview window

The Dashboard in V8.2 appears as a welcome page instead of the system pane as in previous versions. This System overview was moved to the **Monitoring** → **System** menu.

Although the Dashboard pane provides key information about system behavior, the System menu is a preferred starting point to obtain the necessary details about your Storwize V7000 components. The System menu is described next.

5.2 GUI introduction

As shown in Figure 5-8, the former IBM Storwize V7000 GUI System pane was moved to **Monitoring** → **System**.

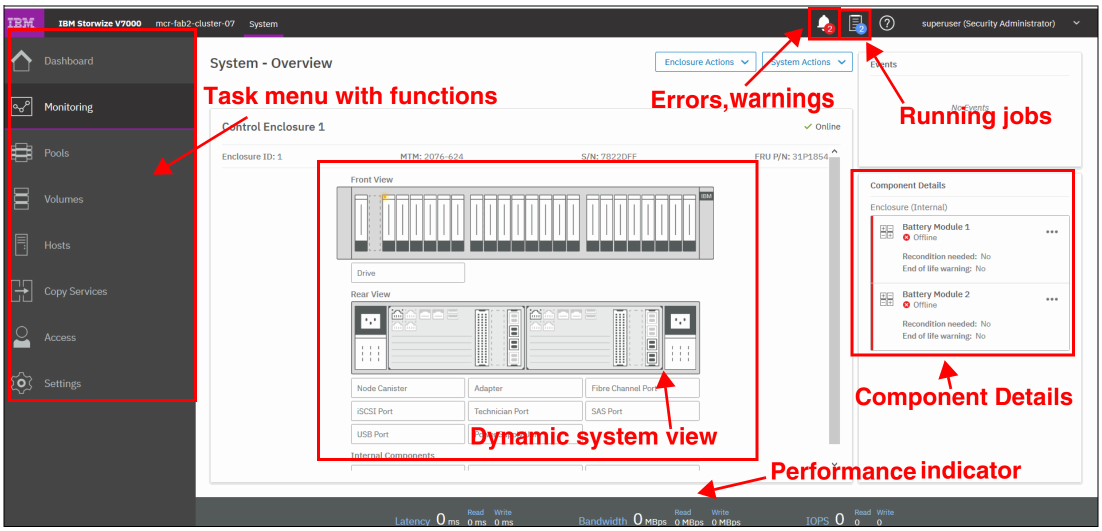


Figure 5-8 IBM Storwize V7000 System pane

5.2.1 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 5-9.

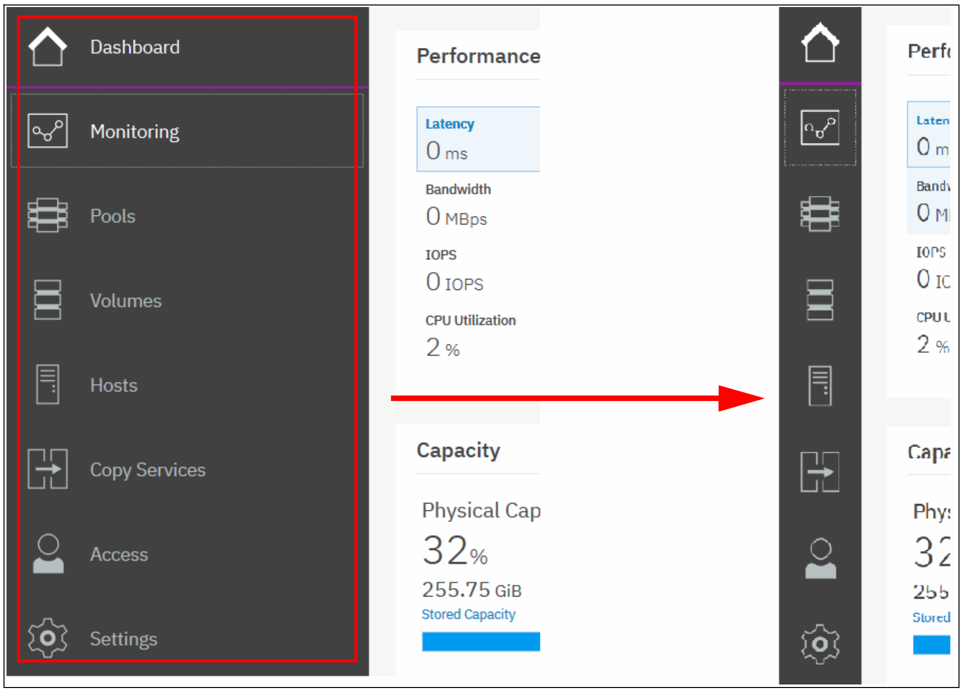


Figure 5-9 The task menu on the left side of the GUI

By reducing the horizontal size of your browser window, the wide task menu shrinks to the icons only.

5.2.2 Suggested tasks

After the initial configuration process is complete, IBM Spectrum Virtualize shows the information about suggested tasks that notify the administrator that several key IBM Storwize V7000 functions are not yet configured. If necessary, this indicator can be closed and these tasks can be performed at any time. Figure 5-10 shows the suggested tasks in the System pane.

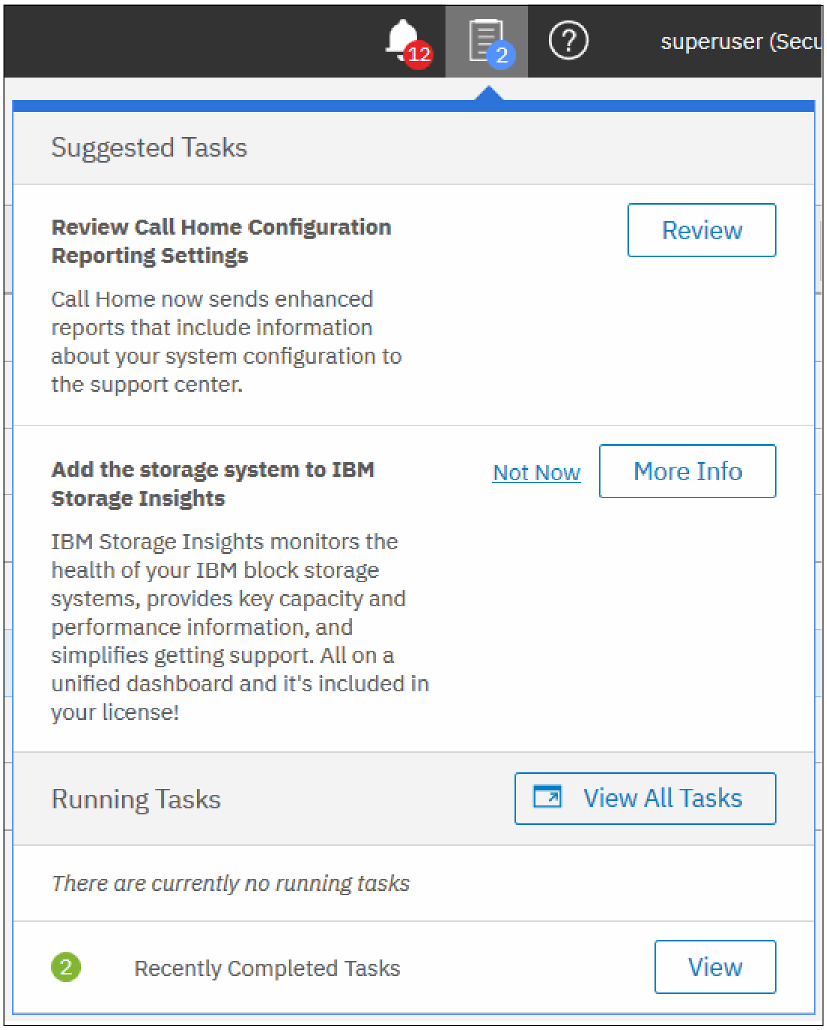


Figure 5-10 Suggested tasks

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 panes:

- Dashboard
- Monitoring
- Pools
- Volumes
- Hosts
- Copy Services

- Access
- Settings

5.2.3 Notification icons and help

Two notification icons are in the top navigation area of the GUI (see Figure 5-11). 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 most right icon offers a help menu with content associated with the current tasks and the currently opened GUI menu.

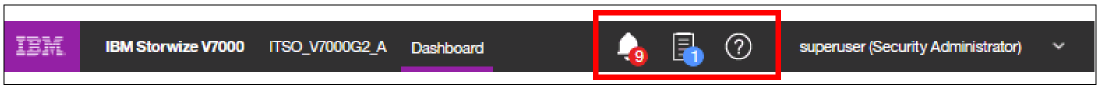


Figure 5-11 Notification area

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 5-12).

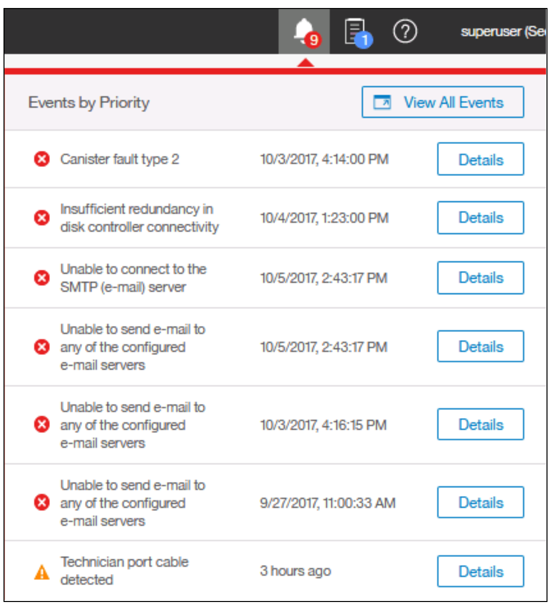


Figure 5-12 System alerts

You can navigate directly to the events menu by clicking **View All Events** option or see each event message separately by clicking the **Details** icon of the specific message, analyze the content, and eventually run the suggested fix procedure (see Figure 5-13).

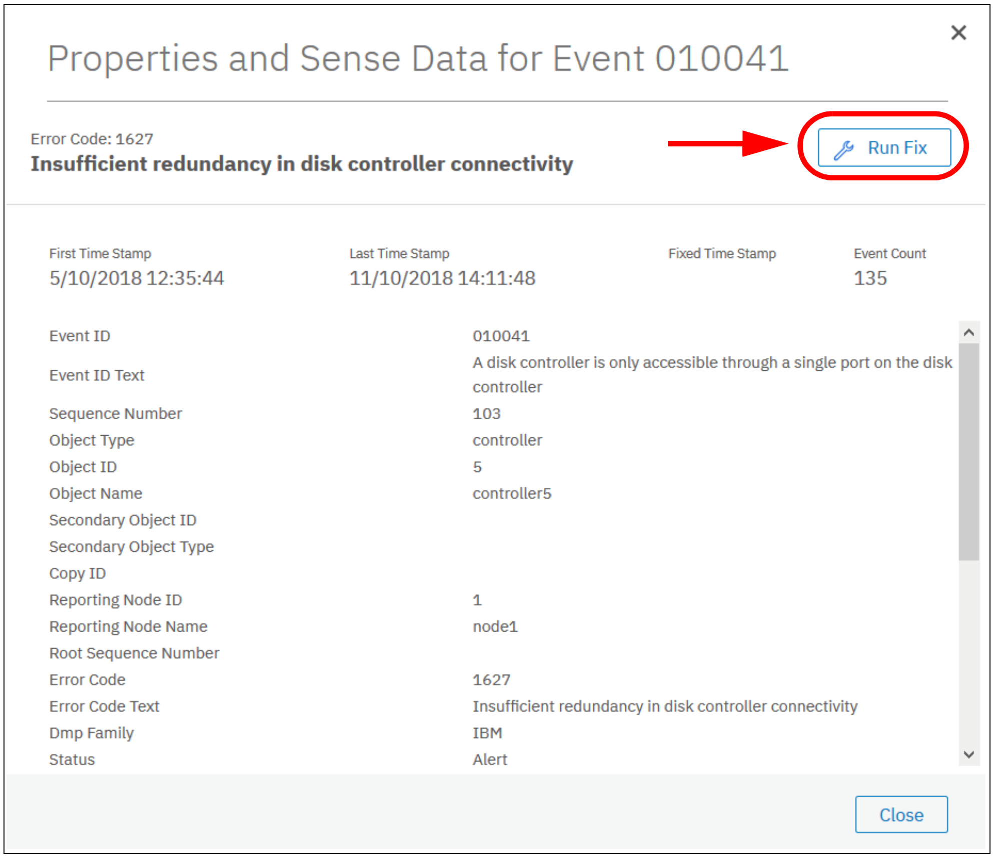


Figure 5-13 External storage connectivity loss

Running jobs 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 5-14 on page 141, we did not yet define any hosts attached to the systems. Therefore, the system suggests that we do so and offers us direct access to the associated host menu. Click **Run Task** to define the host according to the procedure that is explained in Chapter 8, “Hosts” on page 317. If you do not want to define any host now, click **Not Now** and the suggestion message disappears.

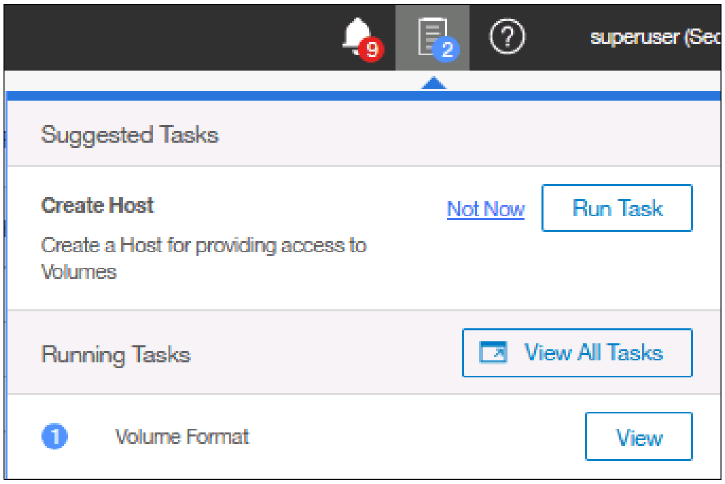


Figure 5-14 Storage allocation area

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 5-15.

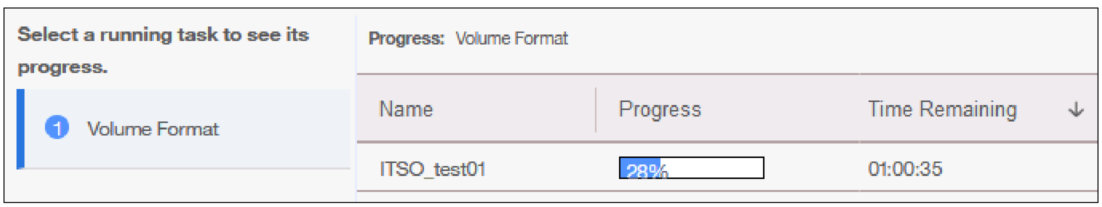


Figure 5-15 Details of running task

The following information can be displayed as part of running tasks:

- Volume migration
- Managed disk (MDisk) removal
- Image mode migration
- Extent migration
- IBM FlashCopy
- Metro Mirror and Global Mirror
- 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. Navigate to a wanted window, hold down the SHIFT or CTRL keys, and make your selection.

Holding down the SHIFT key, select the first item you want in your list, and then, select the last item. All items between the two you chose are also selected if you held the SHIFT key down, as shown in Figure 5-16 on page 142.

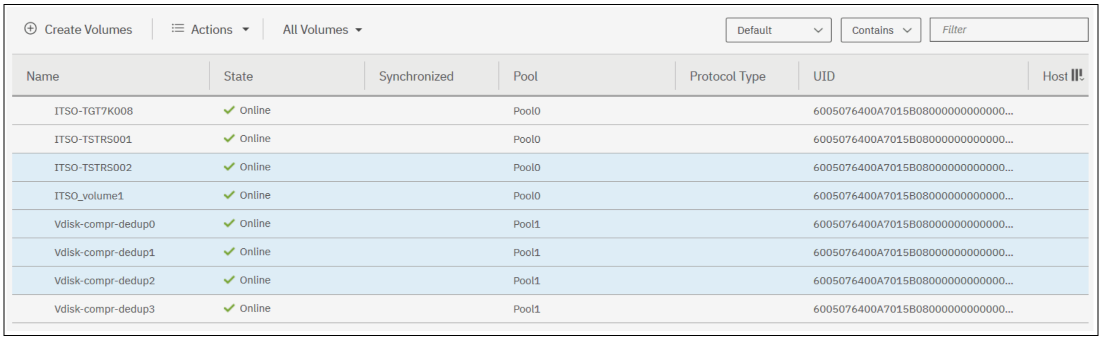


Figure 5-16 Selecting items using the SHIFT key

Holding down the CTRL key, select any items from the entire list. This process allows you to select items that do not appear in sequential order as shown in Figure 5-17.

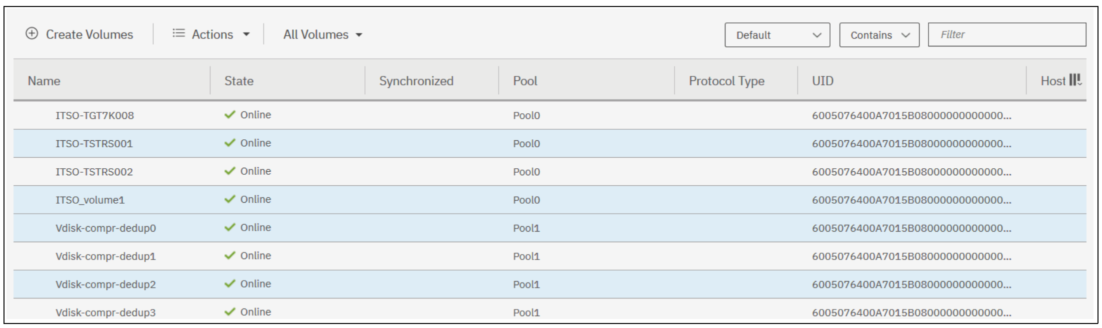


Figure 5-17 Selecting items using the CTRL key

You can also select items by using the in-built filtering function. For more information, see 5.3.1, “Content-based organization” on page 143.

Help

If you need help, you can select the (**?**) button, as shown in Figure 5-18. The first option brings up a new tab with plain text information about the pane you are on and its contents. The second option brings up the same information in IBM Knowledge Center; however, this option requires an internet connection, whereas the first option does not as the information is stored locally on the Storwize V7000.

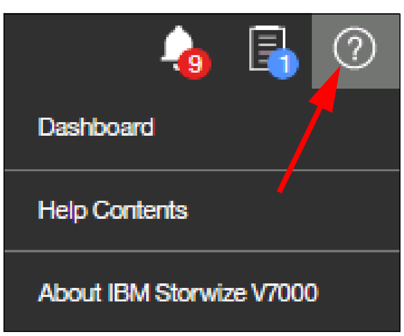


Figure 5-18 Access help menu

For example, in the Dashboard pane, you can open help information that is related to the dashboard-provided information, as shown in Figure 5-19.

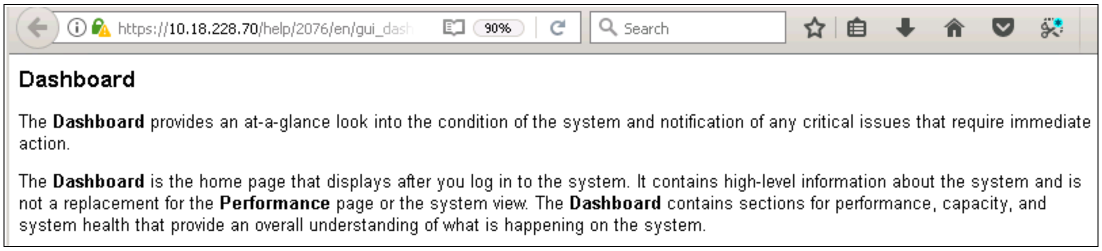


Figure 5-19 Example of Dashboard help content

Selecting the Help Contents option redirects you to the Storwize V7000 IBM Knowledge Center. However, it requires internet access from the workstation where the management GUI is started.

5.3 System View window

Starting with IBM Spectrum Virtualize release V7.4, the welcome window of the GUI changed from the well-known former Overview/system 3D pane to the new System pane. In V8.2, the system pane was changed again to the new System view pane, and the 3D view was removed, as shown in Figure 5-20.

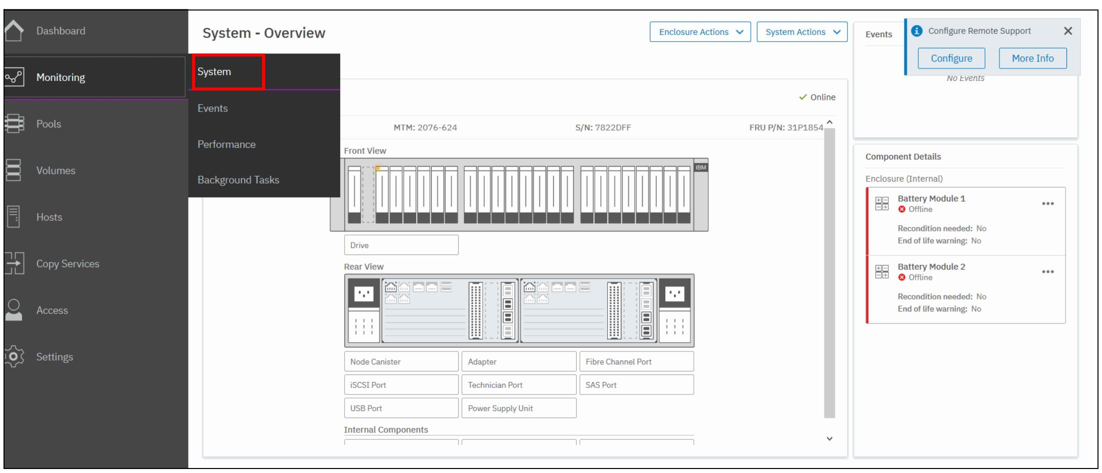


Figure 5-20 Opening the Overview pane

Next, we describe the structure of the pane and how to navigate to various system components to manage them more efficiently and quickly.

5.3.1 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 on the upper-right side of the window. Use this option if the list of object entries is too long and you want to search for something specific.

Complete the following steps to use search filtering:

1. In the Filter box, 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 wanted 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.

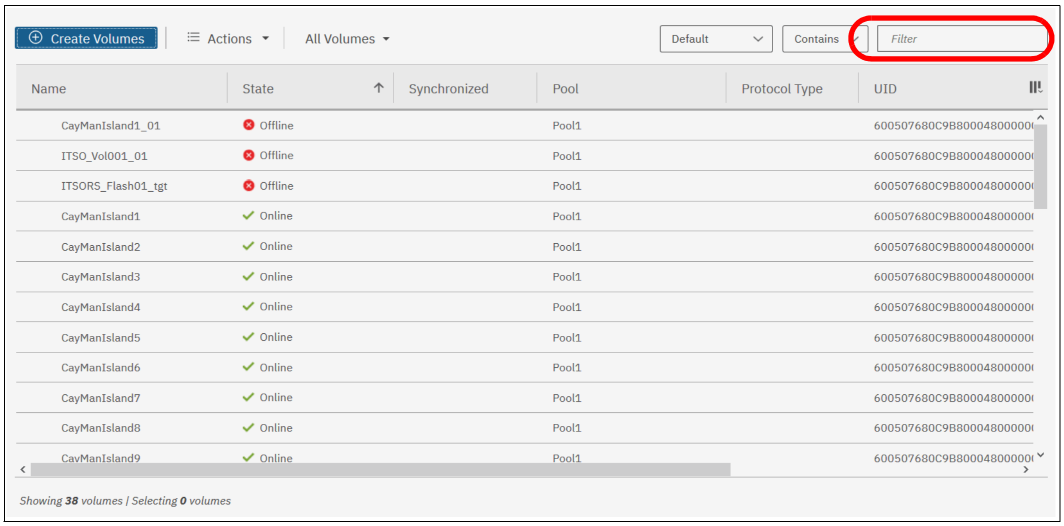


Figure 5-21 Filter search box

2. Enter the text string that you want to filter and press **Enter**.
3. By using this function, you can filter your table that is based on column names. In our example, a volume list is displayed that contains the names that include `cayman` somewhere in the name. `Cayman` is highlighted in amber, as are any columns that contain this information, as shown in Figure 5-22. The search option is not case-sensitive.

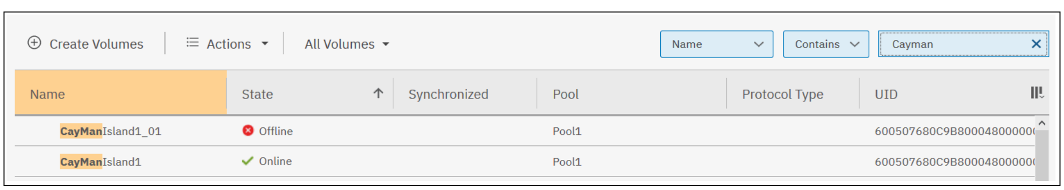


Figure 5-22 Show filtered rows

4. Remove this filtered view by clicking the **Reset Filter** icon, as shown in Figure 5-23.

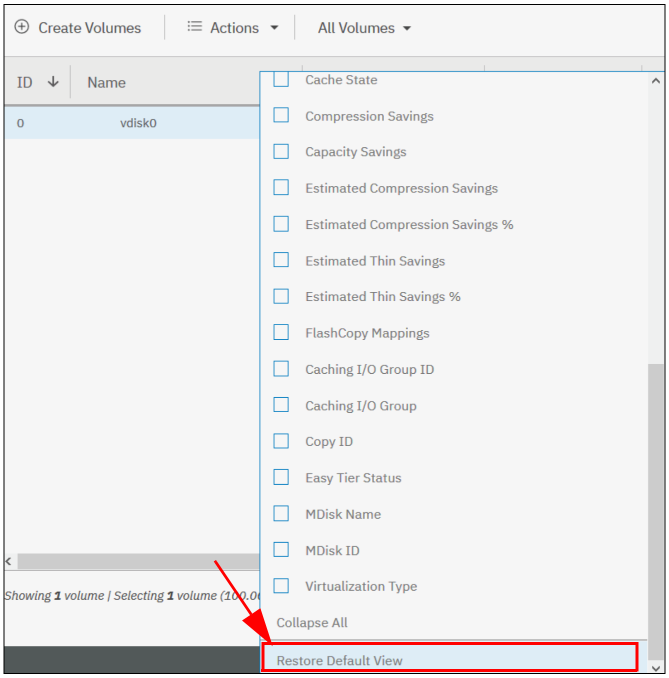


Figure 5-23 Reset filter

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 pane, 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 left corner of the table header. A list of all of the available columns displays, as shown in Figure 5-24.

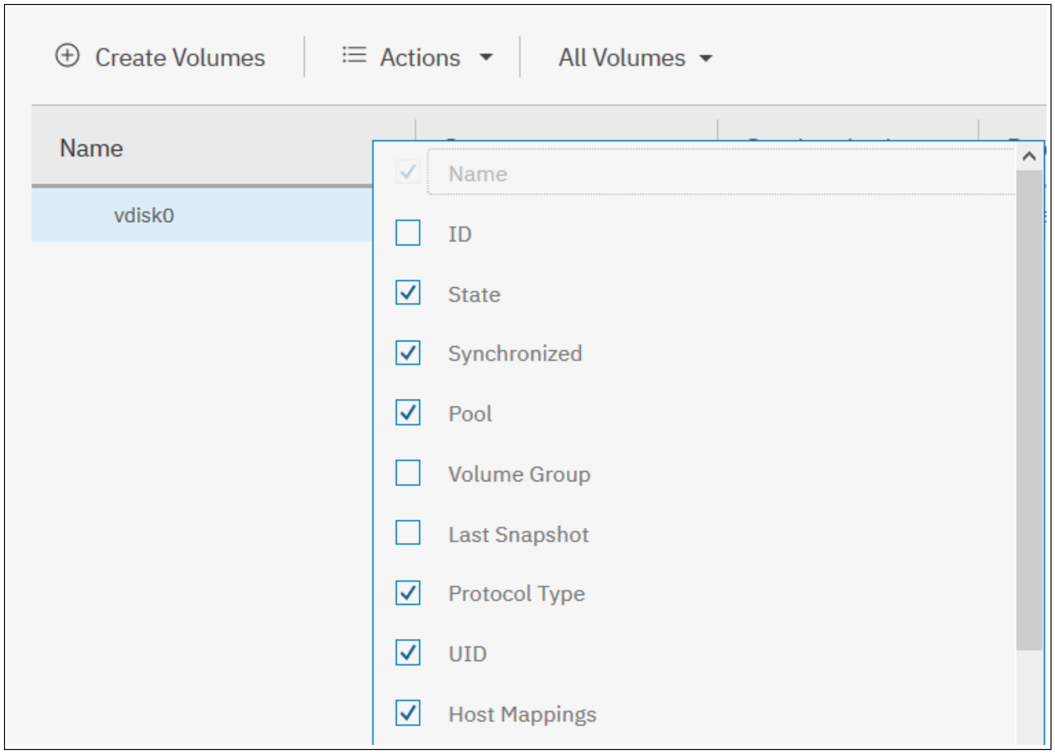


Figure 5-24 Add or remove details in a table

2. 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 5-25.

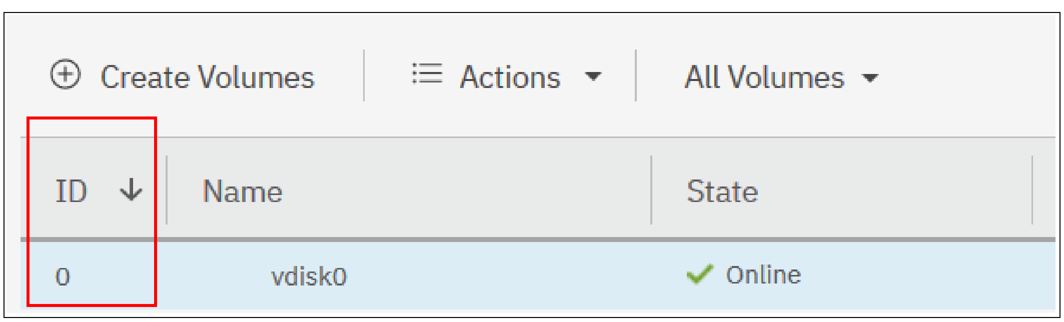


Figure 5-25 Table with an added ID column

3. You can repeat this process several times to create custom tables to meet your requirements.

4. Return to the default table view by selecting **Restore Default View** 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 left-clicking and moving the column right or left, as shown in Figure 5-26. We are attempting to move the Capacity column before the Pool column.

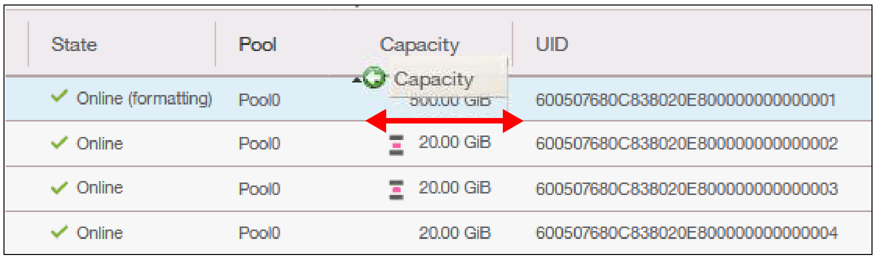


Figure 5-26 Reorganizing table columns

5.4 Monitoring menu

Click the **Monitoring** icon in left pane to open the Monitoring menu (see Figure 5-27). The Monitoring menu offers these navigation options:

- System: 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 5.4.1, “System overview” on page 148.
- 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 values (CSV) file. For more information, see 5.4.2, “Events” on page 152.
- Performance: This option reports the general system statistics that relate to the processor (CPU) utilization, host and internal interfaces, volumes, and MDisks. The GUI allows you to switch between megabytes per second (MBps) or IOPS. For more information, see 5.4.3, “Performance” on page 153.
- Background Tasks: The option shows the progress of all tasks running in the background as listed in “Running jobs and suggested tasks” on page 140

The following section describes the Monitoring menu options (see Figure 5-27).

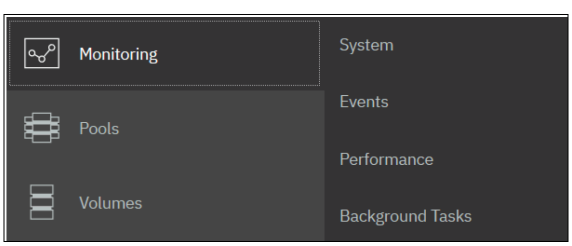


Figure 5-27 Monitoring menu

5.4.1 System overview

The System option on the Monitoring menu provides a general overview of your IBM Storwize V7000. If you have more than one control enclosure in a cluster, each enclosure has its own I/O group section (see Figure 5-28).

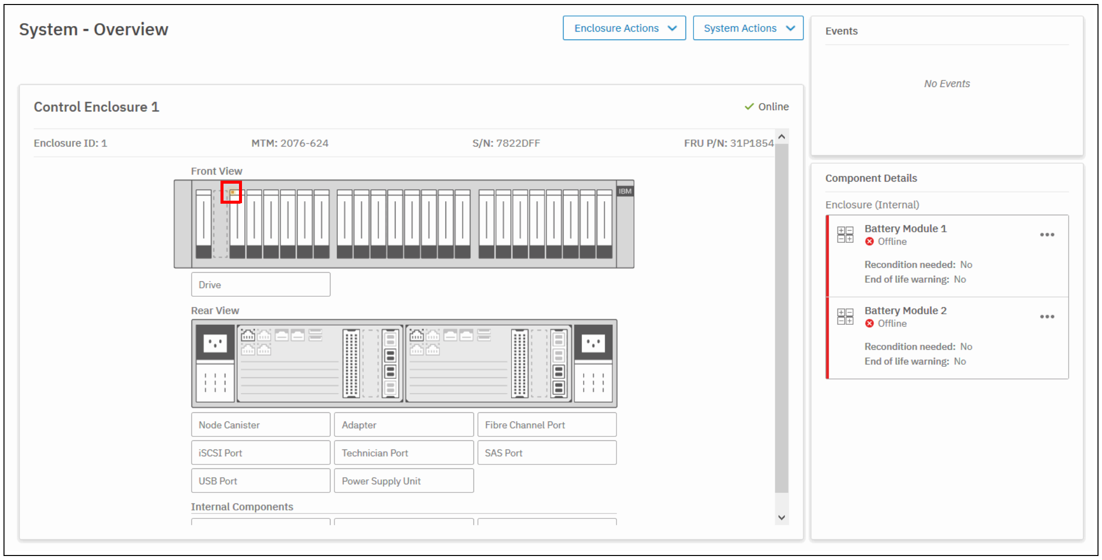


Figure 5-28 System overview

This view shows all external components in real time. In Figure 5-28, you can see that one of the drive's identify 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. For example, clicking a node brings up details, such as if the node is online, and which node is the configuration node, as shown in Figure 5-29. More information about the component can be seen on the right side under the Component Details section that shows when a component is selected.

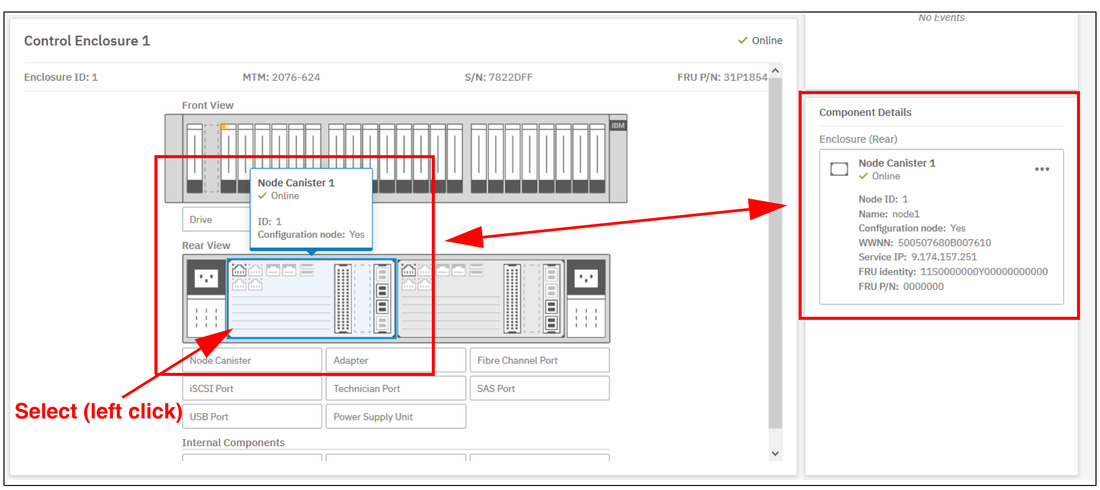


Figure 5-29 Showing node canister details

By right-clicking and selecting **Properties**, you see detailed technical parameters, such as ID, state (online or offline), drive capacity, and the drive use, as shown in Figure 5-30.

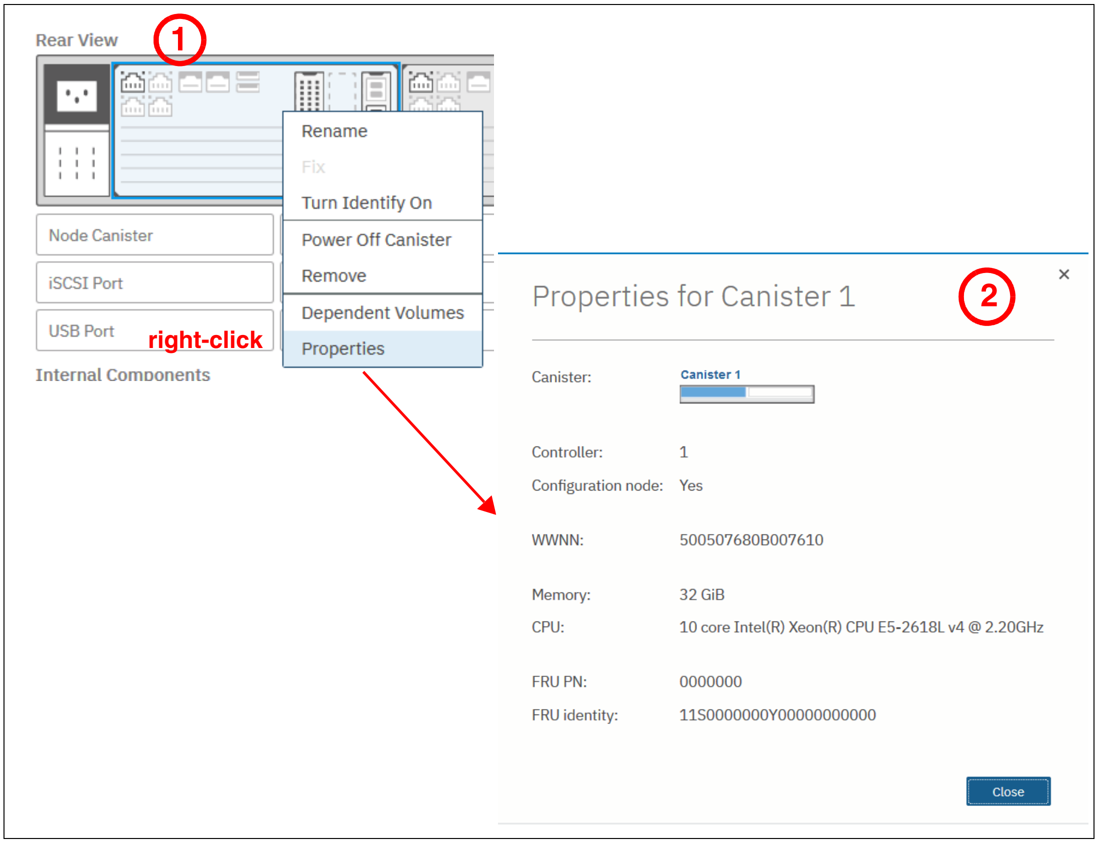


Figure 5-30 Canister information

In an environment with multiple IBM Storwize V7000 clusters, you can easily direct the onsite personnel or technician to the correct device by enabling the identification LED on the front pane. Click **Turn Identify On** in the menu that is shown in Figure 5-31.

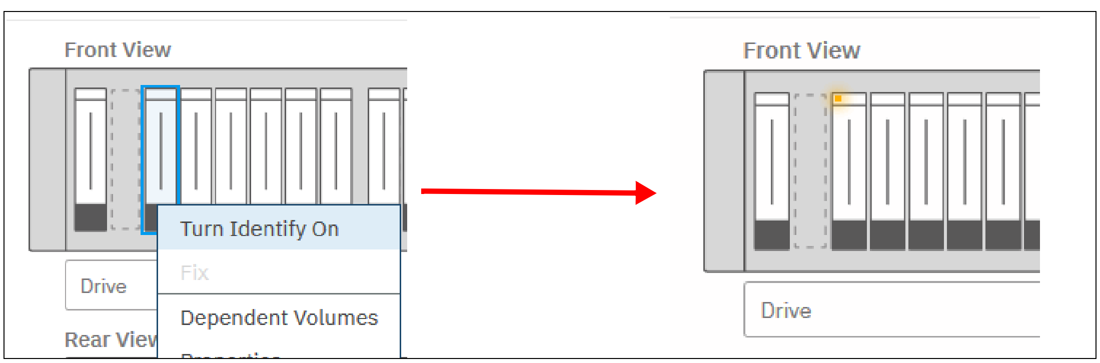


Figure 5-31 Identification LED

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.

After the confirmation, click **Turn LED Off** (see Figure 5-32).

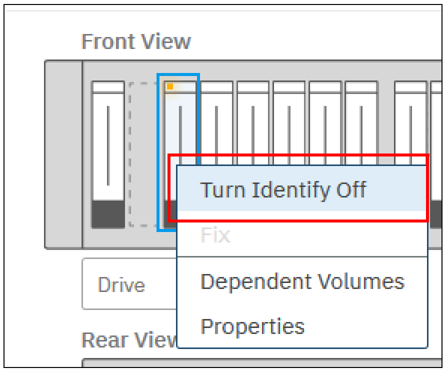


Figure 5-32 Turning the Identify LED off

Alternatively, you can use the command-line interface (CLI) to get the same results. Enter the following commands in this sequence:

1. Enter **`svctask chenclosure -identify yes 1`** (or enter **`chenclosure -identify yes 1`**).
2. Enter **`svctask chenclosure -identify no 1`** (or enter **`chenclosure -identify no 1`**).

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 show in the right pane, as with the external components.

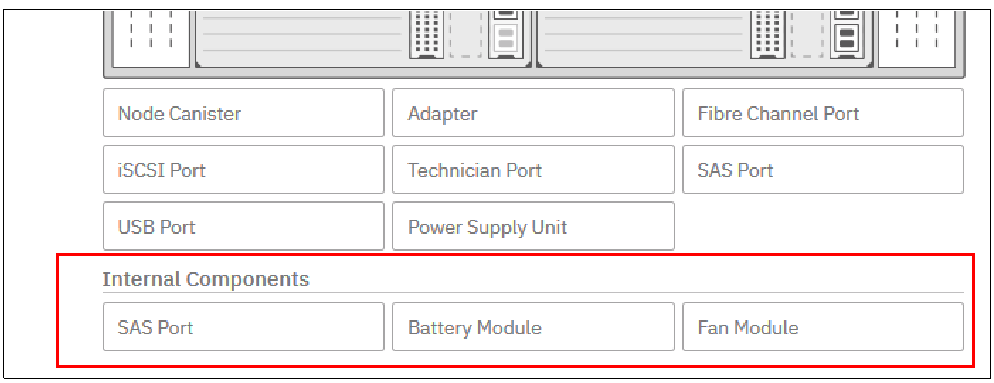


Figure 5-33 Viewing internal components

You can also choose **SAS Chain View** to view directly attached expansion enclosures, as shown in Figure 5-34. A useful view of the entire SAS chain is displayed, with selectable components that show port numbers and canister numbers, along with a cable diagram for easy cable tracking.

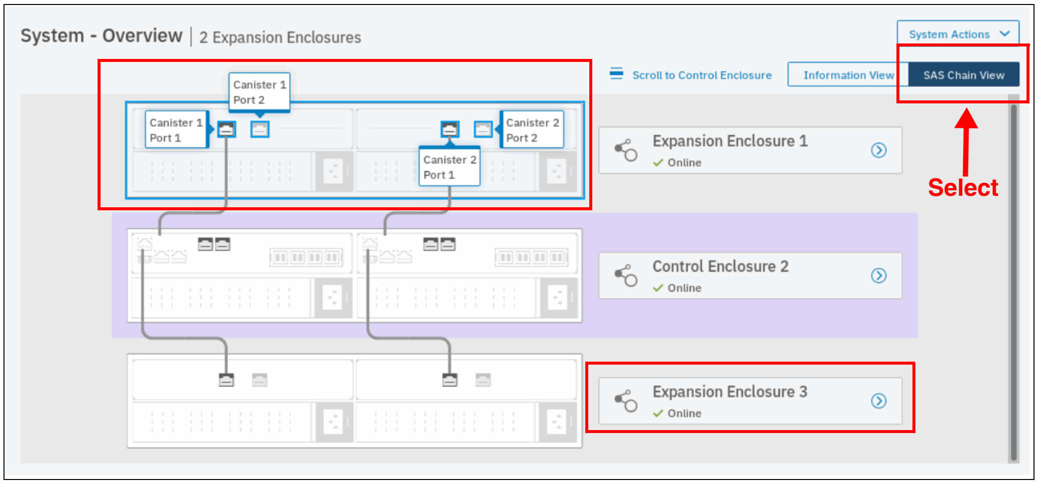


Figure 5-34 SAS Chain View

You can select any Enclosure to get more information, including serial number and model type, as shown in Figure 5-35 where Expansion Enclosure 3 is selected. You can also see the Events and Component Details areas on the right side of the pane, which shows information that relates to the enclosure or component you select.

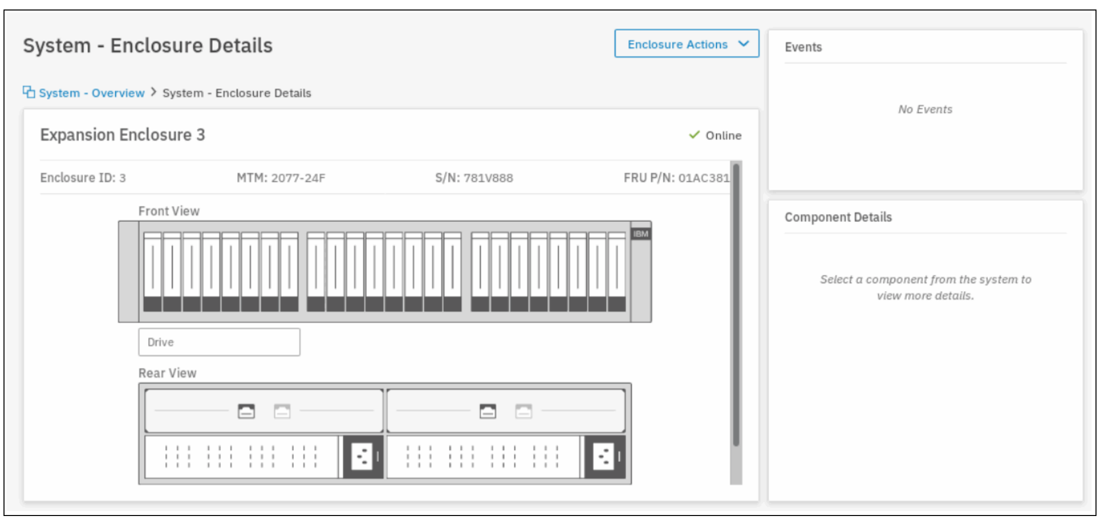


Figure 5-35 Enclosure Details

With directly attached expansion enclosures, the view is condensed to show all expansion enclosures on the right side, as shown in Figure 5-36. 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.

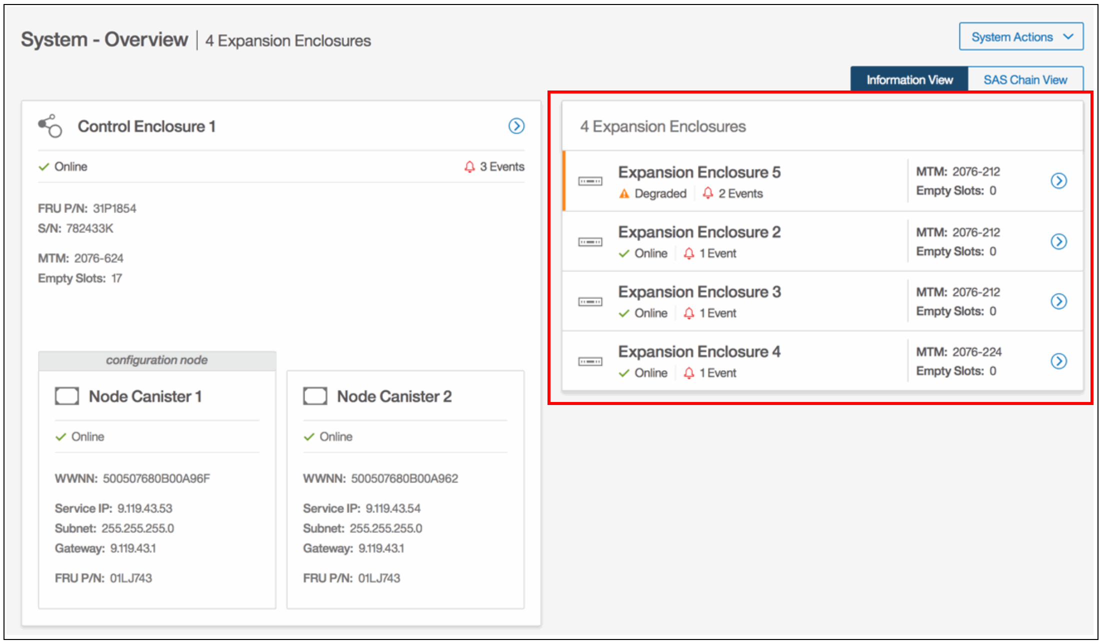


Figure 5-36   System Overview with attached enclosures

5.4.2 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 5-37 provides an example of records in the system Event log.

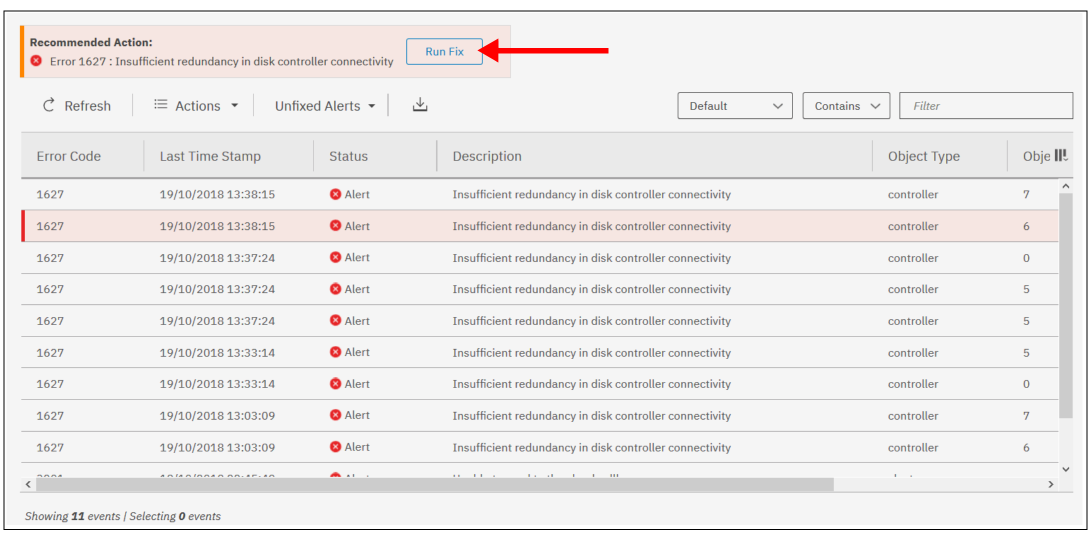


Figure 5-37   Event log list

For the error messages with the highest internal priority, perform corrective actions by running fix procedures. Click the **Run Fix** button (see Figure 5-37 on page 152). The fix procedure wizard opens, as shown in Figure 5-38.

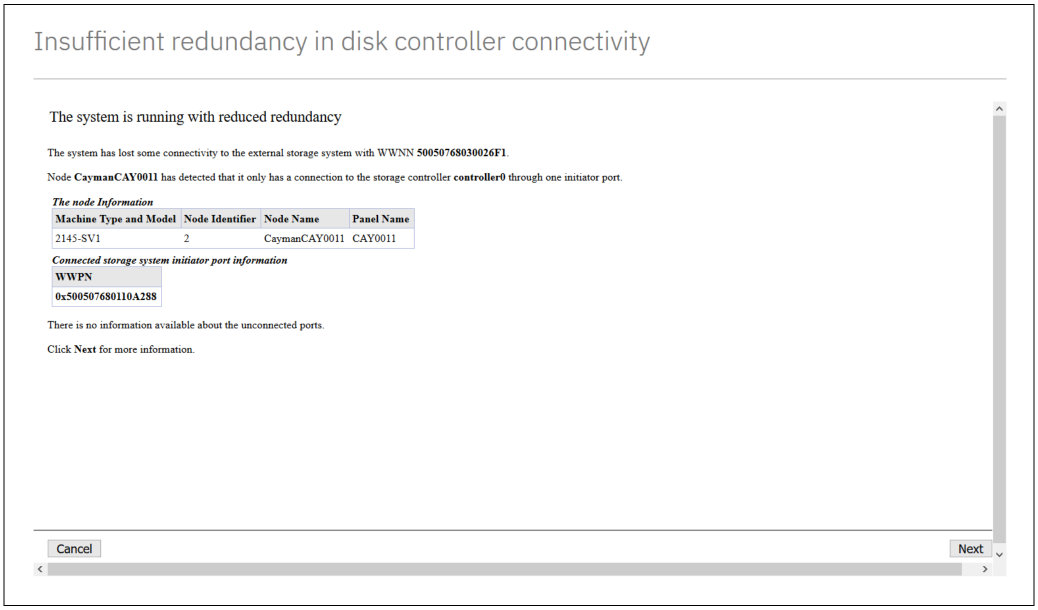


Figure 5-38 Performing fix procedure

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, "RAS, monitoring, and troubleshooting" on page 673.

5.4.3 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 5-39 on page 154).

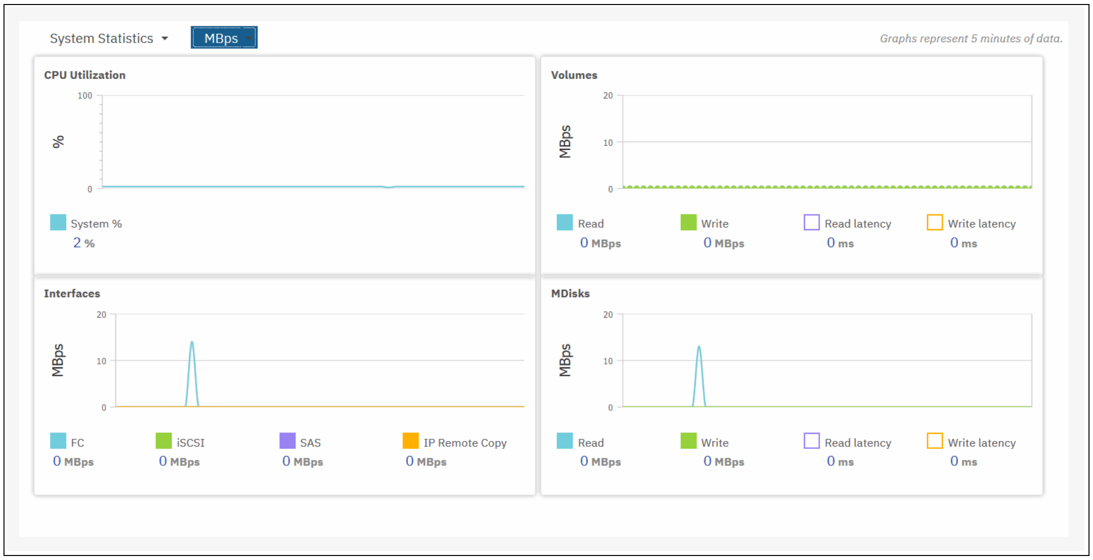


Figure 5-39 Performance statistics of the IBM Storwize V7000

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 time stamp, as shown in Figure 5-40.

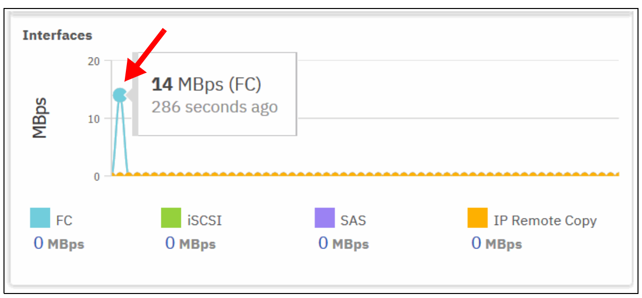


Figure 5-40 Sample details

The charts that are shown in Figure 5-40 represent 5 minutes of the data stream. For in-depth storage monitoring and performance statistics with historical data about your IBM Storwize system, use IBM Spectrum Control (formerly enabled by IBM Tivoli Storage Productivity Center for Disk and IBM Virtual Storage Center) or IBM Storage Insights.

5.4.4 Background tasks

This menu provides an overview of currently running tasks triggered by the administrator. In contrast to the Running jobs and Suggested tasks indication in the middle of top pane, it does not list the suggested tasks that administrators should consider performing. The overview provides more details than the indicator, as shown in Figure 5-41.

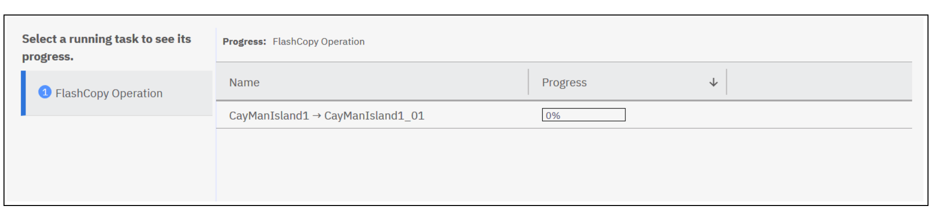


Figure 5-41 List of running tasks

You can switch between each type (group) of operation, but you cannot show them all in one list (see Figure 5-42).

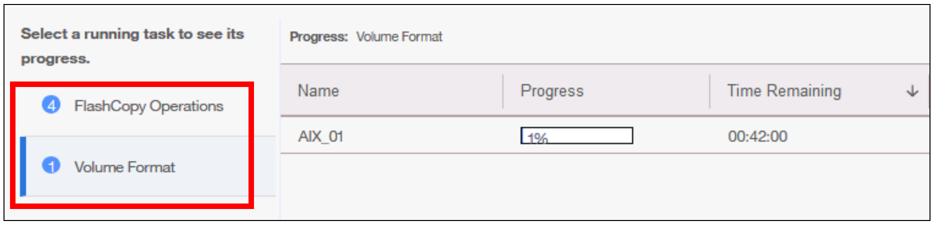


Figure 5-42 Switching between types of background tasks

5.5 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.

Pools menu contains the following items accessible from GUI (see Figure 5-43):

- Pools
- Volumes by Pool
- Internal Storage
- External Storage
- MDisks by Pool
- System Migration

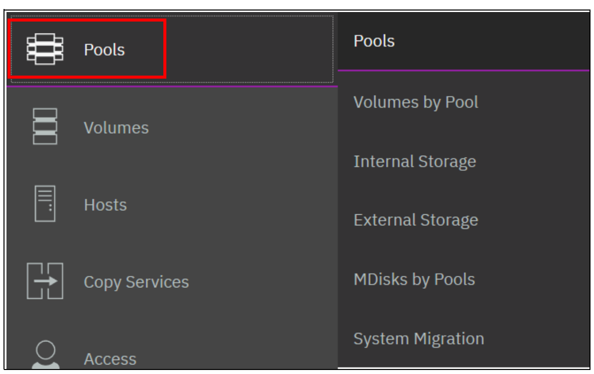


Figure 5-43 Pools menu

For more information about storage pool configuration and management, see Chapter 6, “Storage pools” on page 191.

5.6 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.

Volumes menu contains the following items, as shown in Figure 5-44 on page 156:

- Volumes
- Volumes by Pool
- Volumes by Host

- Cloud Volumes

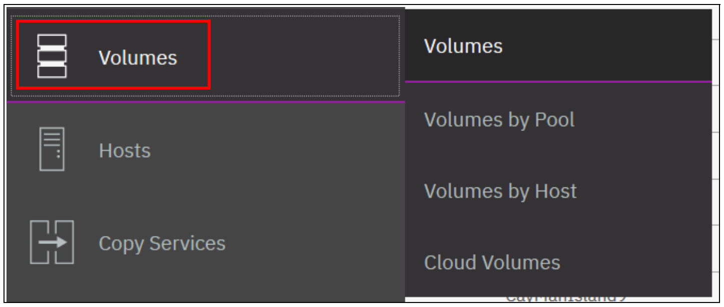


Figure 5-44 Volumes menu

For more information about these tasks and configuration and management process guidance, see Chapter 7, “Volumes” on page 241.

5.7 Hosts

A host system is a computer that is connected to the system through a Fibre Channel 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 Storwize V7000. Fibre Channel and SAS connections use WWPNs to identify the host interfaces to the systems.

Hosts menu consists of the following choices, as shown in Figure 5-45:

- Hosts
- Host Clusters
- Ports by Host
- Mappings
- Volumes by Host

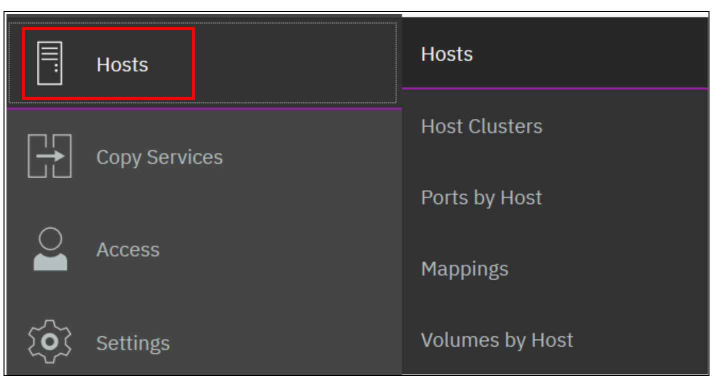


Figure 5-45 Hosts menu

For more information about configuration and management of hosts by using the GUI, see in Chapter 8, “Hosts” on page 317.

5.8 Copy Services

The IBM Spectrum Virtualize copy services and volumes copy operations are based on the IBM 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 for copying target volumes from their respective source volumes. This technique creates a consistent copy of data that spans multiple volumes.

The IBM SAN Volume Controller Copy Services menu offers the following operations in the GUI, as shown in Figure 5-46:

- FlashCopy
- Consistency Groups
- FlashCopy Mappings
- Remote Copy
- Partnership

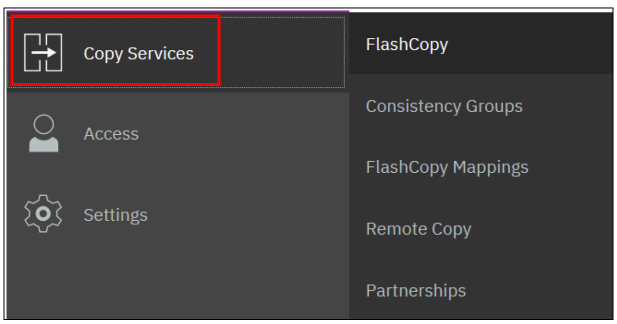


Figure 5-46 Copy Services in GUI

Because the Copy Services are one of the most important features for resiliency solutions, more technical information is available in Chapter 11, "Advanced Copy Services" on page 435.

5.9 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 two categories:

- Users
- Audit Log

In this section, we explain how to create, modify, or remove user, and how to see records in the audit log.

The Access menu is available from the left pane, as shown in Figure 5-47.

Figure 5-47 Access menu

5.9.1 Users

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 the IBM Storwize V7000, click **Access** → **Users**, as shown in Figure 5-48.

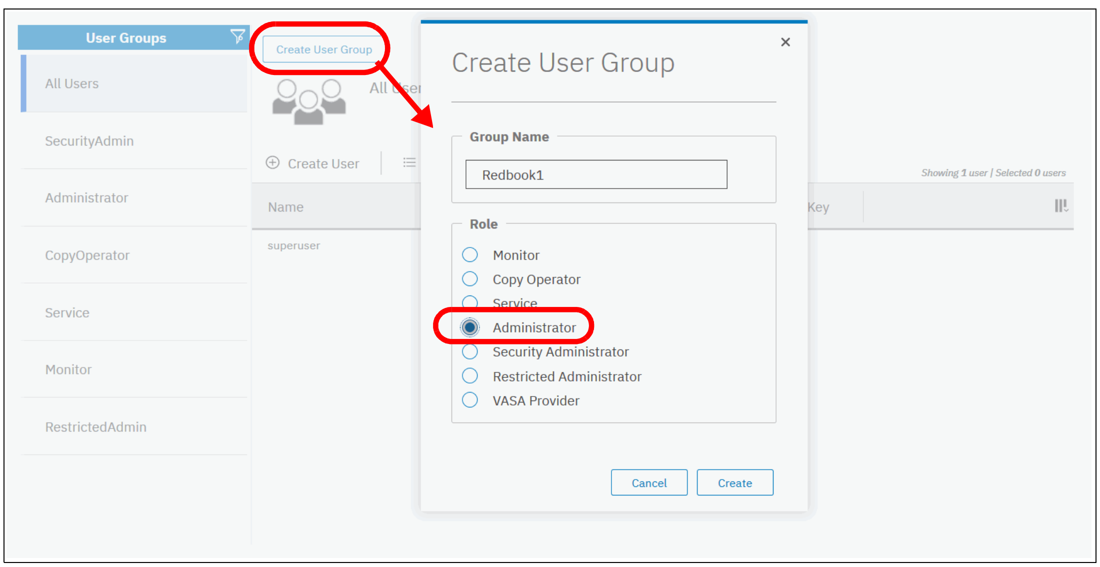


Figure 5-48 Defining User Group in Storwize V7000

The following privilege User group roles exist in the IBM Spectrum Virtualize:

- Security Administrator: Manages all functions of the systems, except tasks that are associated with commands `satask` and `sainfo`.
- Administrator: Has full rights in the system, except those commands that are related to user management and authentication.
- Restricted Administrator: Has the same rights as Administrators, except removing volumes, host mappings, hosts, or pools. This option is ideal for support personnel.
- Copy Operators: Can start, stop, or pause any FlashCopy-based operations.

- Monitor users: Have access to all viewing operations. They cannot change any value or parameters of the system.
- Service users: Can set the time and date on the system, delete dump files, add and delete nodes, apply service, and shut down the system. They have access to all views.
- VASA Provider users: Can manage VMware vSphere Virtual Volumes (VVOLs).

Registering a new user

After you defined your group (in our example Redbook1 with Administrators privileges), you can register a new user within this group. Click **Create User** and select **MyGroup** (see Figure 5-49).

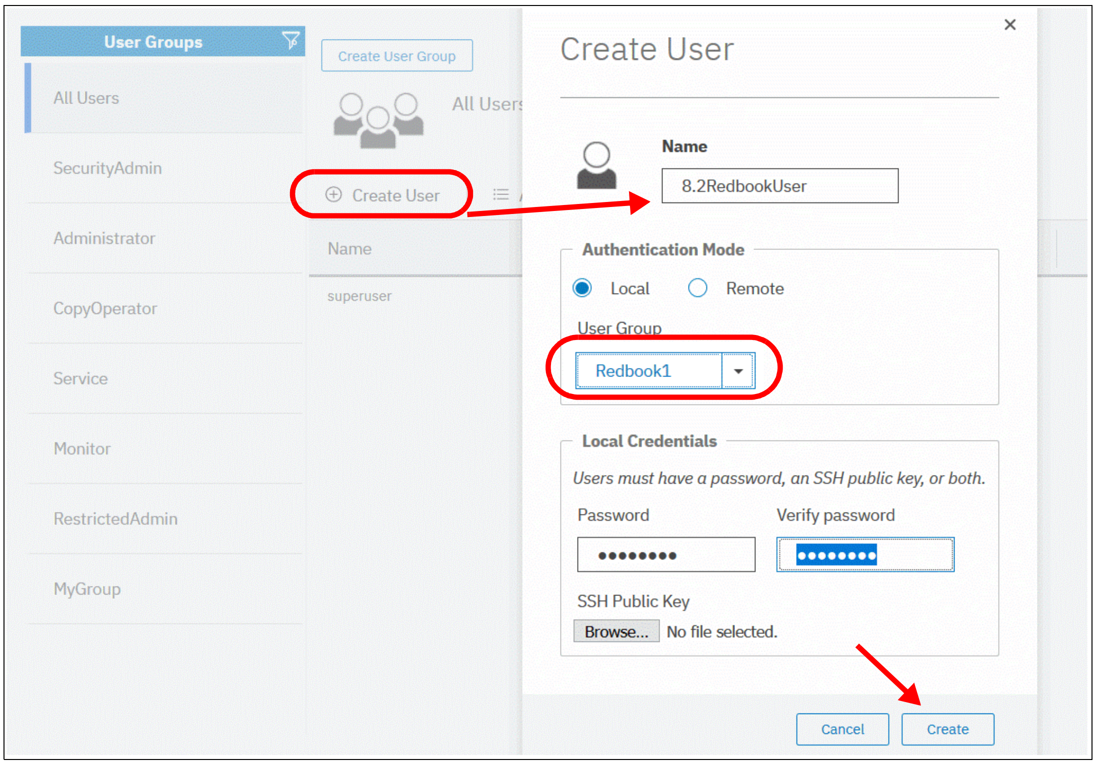


Figure 5-49 Registering new user account

Deleting a user

To remove a user account, right-click the user in the All Users list and select **Delete**, as shown in Figure 5-50.

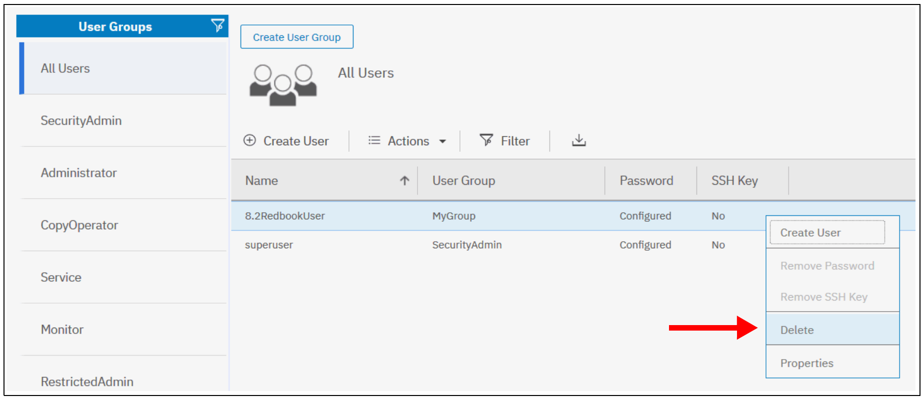


Figure 5-50   Deleting user account

> **Attention:** When you click **Delete**, the user account is directly deleted from Storwize V7000. No other confirmation request is presented.

Resetting user password

To set a new password for the user, right-click the user (or click the **Actions** button) and select **Properties**. In this window, you can assign the user to a different group or reset their password (see Figure 5-51).

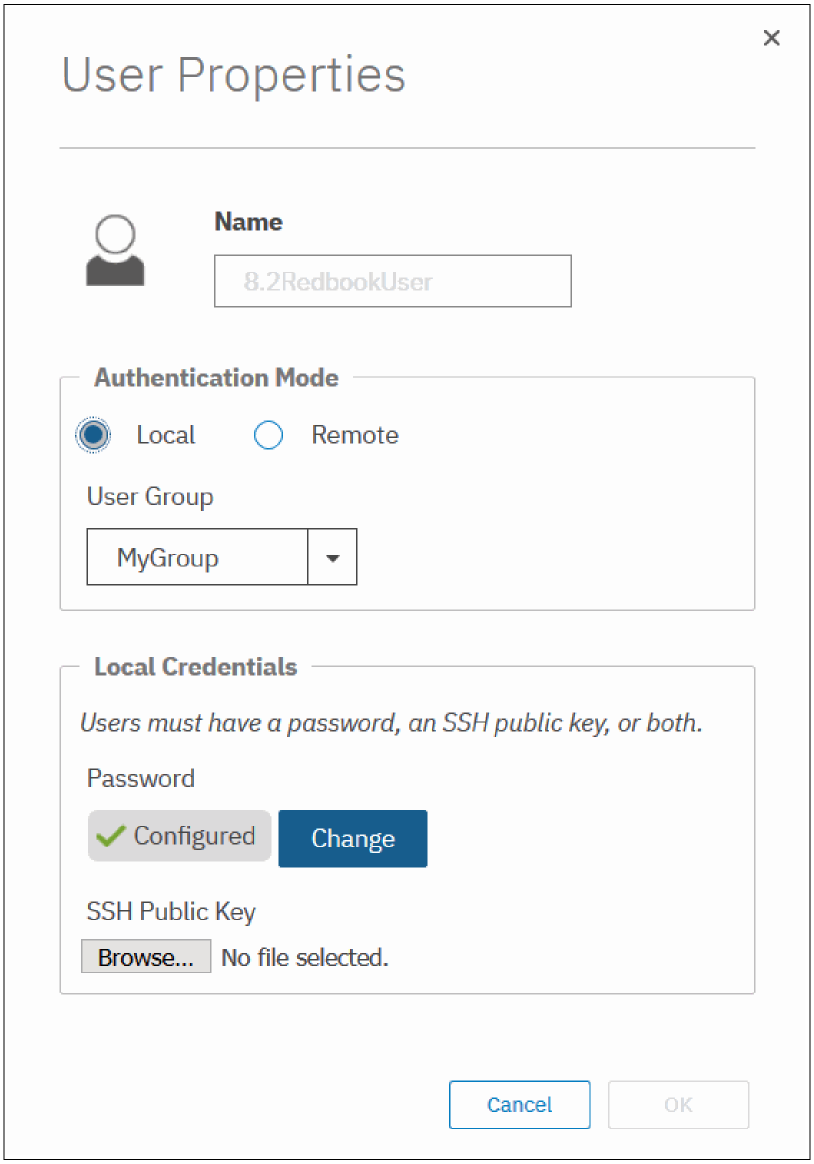


Figure 5-51 Changing user's password

5.9.2 Audit log

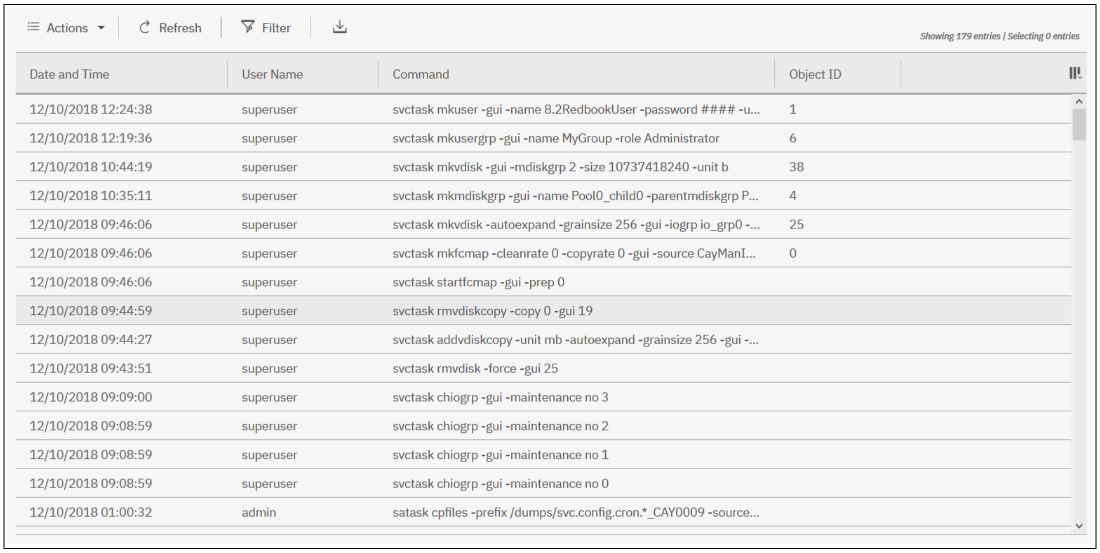
An audit log documents actions that are submitted through the management GUI or the CLI. You can use the audit log to monitor user activity on your system.

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 5-52.



Actions Refresh Filter

Showing 179 entries | Selecting 0 entries

Date and Time	User Name	Command	Object ID
12/10/2018 12:24:38	superuser	svctask mkuser -gui -name 8.2RedbookUser -password #### -u...	1
12/10/2018 12:19:36	superuser	svctask mkusergrp -gui -name MyGroup -role Administrator	6
12/10/2018 10:44:19	superuser	svctask mkvdisk -gui -mdiskgrp 2 -size 10737418240 -unit b	38
12/10/2018 10:35:11	superuser	svctask mkmdiskgrp -gui -name Pool0_child0 -parentmdiskgrp P...	4
12/10/2018 09:46:06	superuser	svctask mkvdisk -autoexpand -grainsize 256 -gui -iogrp io_grp0 -...	25
12/10/2018 09:46:06	superuser	svctask mkfcmap -cleanrate 0 -copyrate 0 -gui -source CayManI...	0
12/10/2018 09:46:06	superuser	svctask startfcmap -gui -prep 0	
12/10/2018 09:44:59	superuser	svctask rmvdiskcopy -copy 0 -gui 19	
12/10/2018 09:44:27	superuser	svctask addvdiskcopy -unit mb -autoexpand -grainsize 256 -gui -...	
12/10/2018 09:43:51	superuser	svctask rmvdisk -force -gui 25	
12/10/2018 09:09:00	superuser	svctask chiogrp -gui -maintenance no 3	
12/10/2018 09:08:59	superuser	svctask chiogrp -gui -maintenance no 2	
12/10/2018 09:08:59	superuser	svctask chiogrp -gui -maintenance no 1	
12/10/2018 09:08:59	superuser	svctask chiogrp -gui -maintenance no 0	
12/10/2018 01:00:32	admin	satask cpfiles -prefix /dumps/svc.config.cron.*_CAY0009 -source...	

Figure 5-52 Audit log

> **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.

5.10 Settings

Use the Settings pane to configure system options for notifications, security, IP addresses, and preferences that are related to display options in the management GUI (see Figure 5-53).

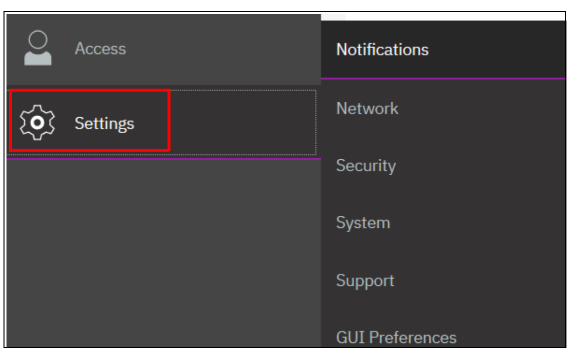


Figure 5-53 Settings menu

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 the support center when significant events are detected. Any combination of these notification methods can be used simultaneously.

- Network: Use the Network pane to manage the management IP addresses for the system, service IP addresses for the nodes, and iSCSI and Fibre Channel configurations. The system must support Fibre Channel or Fibre Channel over Ethernet connections to your storage area network (SAN).
- Security: Use the Security pane to configure and manage remote authentication services.
- System: Navigate to the System menu item to manage overall system configuration options, such as licenses, updates, and date and time settings.
- Support: Helps to configure and manage connections, and upload support packages to the support center.
- GUI Preferences: Configure welcome message after login, refresh internals and GUI logout timeouts.

These options are described next.

5.10.1 Notifications menu

IBM Storwize V7000 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 IBM Storwize V7000.

To view the SNMP configuration, use the System window. Point to the Settings icon and click **Notification** → **SNMP** (see Figure 5-54).

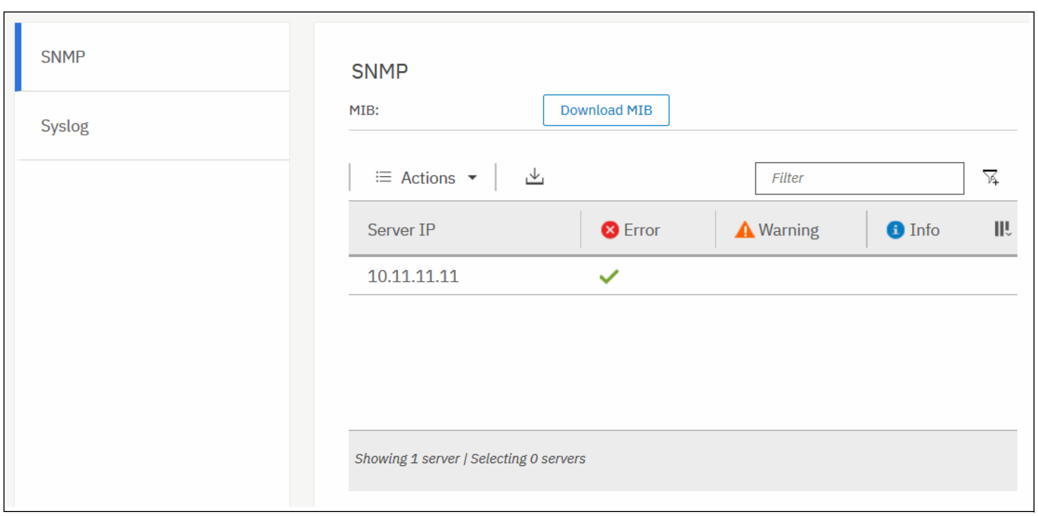


Figure 5-54 Setting SNMP server and traps

From this window (see Figure 5-54 on page 163), 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.

- Server Port

  The remote port number for the SNMP server. The remote port number must be a value of 1 - 65535.

- Community

  The SNMP community is the name of the group to which devices and management stations that run SNMP belong.

- 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.

  To remove an SNMP server, click the Minus sign (**-**). To add another SNMP server, click the Plus sign (**+**).

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 pane to view the Syslog messages that are sent by the IBM Storwize V7000. To view the Syslog configuration, use the System pane and point to **Settings** and click **Notification** → **Syslog** (see Figure 5-55).

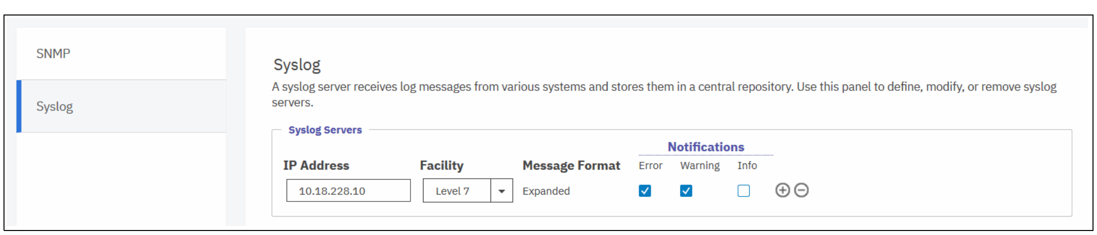


Figure 5-55 Setting Syslog messages

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.

- 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 detail about the event.
  - The expanded format provides more details about the event.

- 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 whether any corrective action is necessary.

    > **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.

To remove a syslog server, click the Minus sign (**-**). To add another syslog server, click the Plus sign (**+**).

The syslog messages can be sent in concise message format or expanded message format.

Example 5-1 shows a compact format syslog message.

Example 5-1 Compact syslog message example

```
IBM2145 #NotificationType=Error #ErrorID=077001 #ErrorCode=1070 #Description=Node
CPU fan failed #ClusterName=V7kCluster1 #Timestamp=Wed Oct 02 08:00:00 2018 BST
#ObjectType=Node #ObjectName=Node1 #CopyID=0 #ErrorSequenceNumber=100
```

Example 5-2 shows an expanded format syslog message.

Example 5-2 Full format syslog message example

```
IBM2145 #NotificationType=Error #ErrorID=077001 #ErrorCode=1070 #Description=Node
CPU fan failed #ClusterName=V7kCluster1 #Timestamp=Wed Oct 02 08:00:00 2018 BST
#ObjectType=Node #ObjectName=Node1 #CopyID=0 #ErrorSequenceNumber=100 #ObjectID=2
#NodeID=2 #MachineType=2076524#SerialNumber=1234567 #SoftwareVersion=8.2.1.0
```

```
(build 147.6.1810180824000)#FRU=fan 24P1118, system board 24P1234
#AdditionalData(0->63)=000000002100000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000#Additional
Data(64-127)=00000000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000
```

5.10.2 Network

This section describes how to view the network properties of the IBM Storwize V7000 system. The network information can be obtained by choosing **Network**, as shown in Figure 5-56.

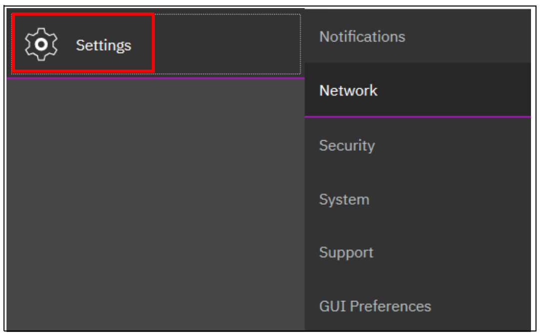


Figure 5-56 Accessing network information

Configuring the network

The procedure to set up and configure IBM Storwize V7000 network interfaces is described in Chapter 4, “Initial configuration” on page 87.

Management IP addresses

To view the management IP addresses of the 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 Figure 5-57.

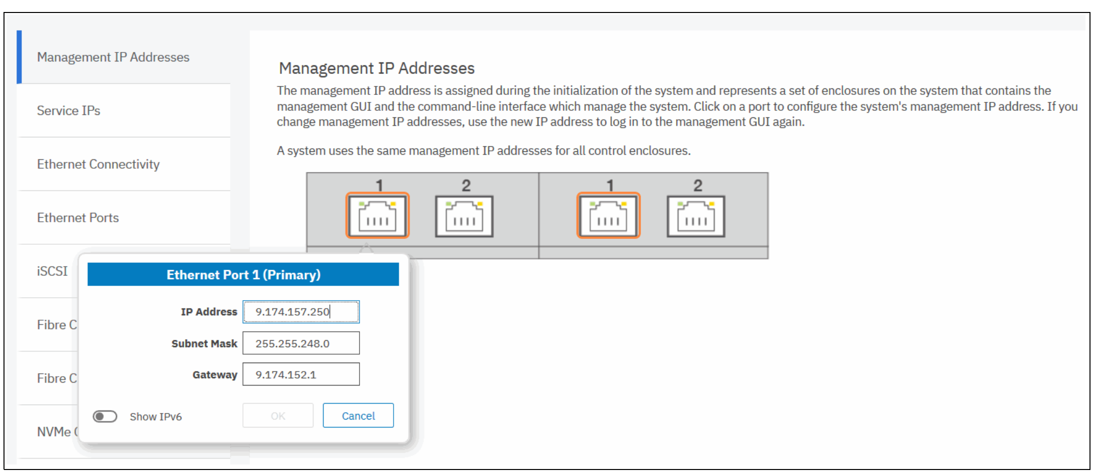


Figure 5-57 Viewing the management IP addresses

Service IP information

To view the Service IP information of your IBM Spectrum Virtualize, select **Settings** → **Network**, as shown in Figure 5-56 on page 166. Click the **Service IP Address** option to view the properties, as shown in Figure 5-58.

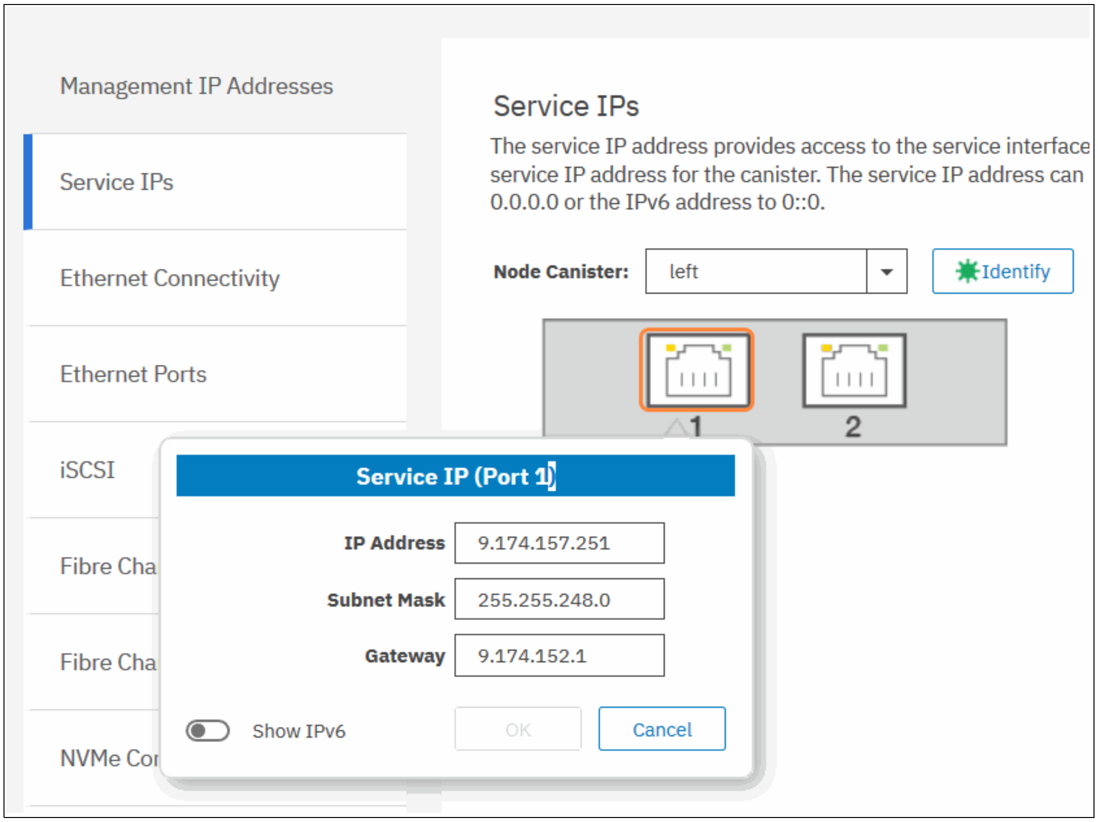


Figure 5-58 Viewing service IP address

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 configure to support IPv4 or IPv6.

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 5-59 on page 168.

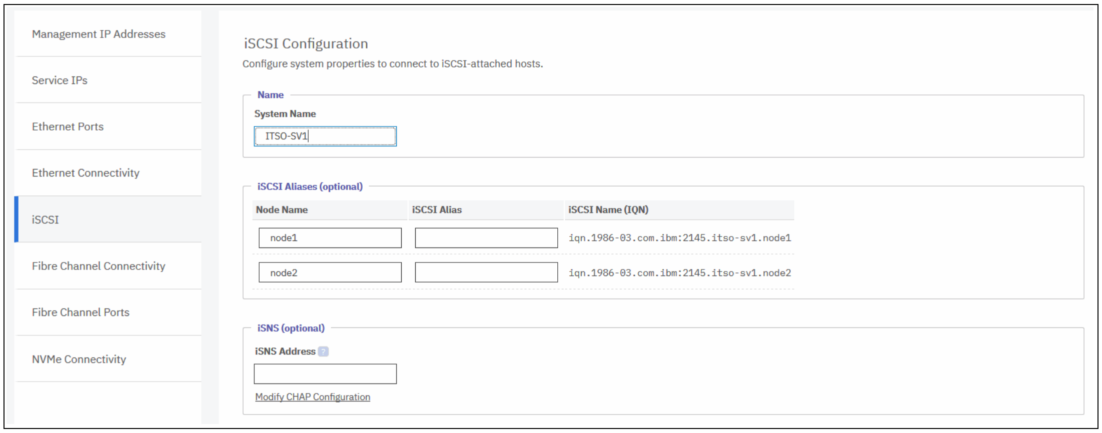


Figure 5-59 iSCSI Configuration

The following parameters can be updated:

- System Name

  It is important to set the system name correctly because it is part of the 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:

  a. Click an iSCSI alias.
  b. 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.

- iSNS and CHAP

  You can specify the IP address for the iSCSI Storage Name Service (iSNS). Host systems use the iSNS server to manage iSCSI targets and for iSCSI discovery.

  You can also enable Challenge Handshake Authentication Protocol (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/host definition. You can create an iSCSI host definition without the use of a CHAP.

Fibre Channel information

As shown in Figure 5-60, you can use the Fibre Channel Connectivity pane to display the FC connectivity 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

View the Fibre Channel Connectivity, as shown in Figure 5-60.

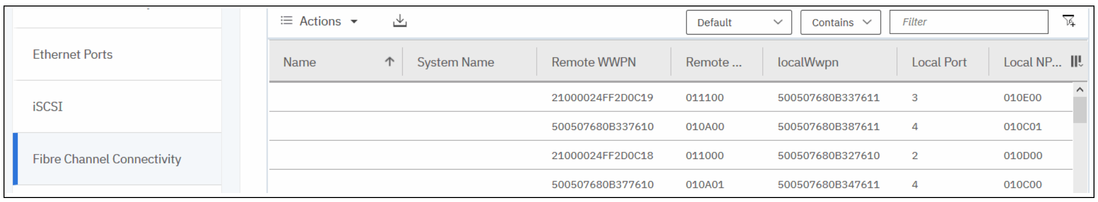


Figure 5-60 Fibre Channel connections

In the Fibre Channel Ports pane, you can use this view to display how the Fibre Channel 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 is allowed to 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 5-61).

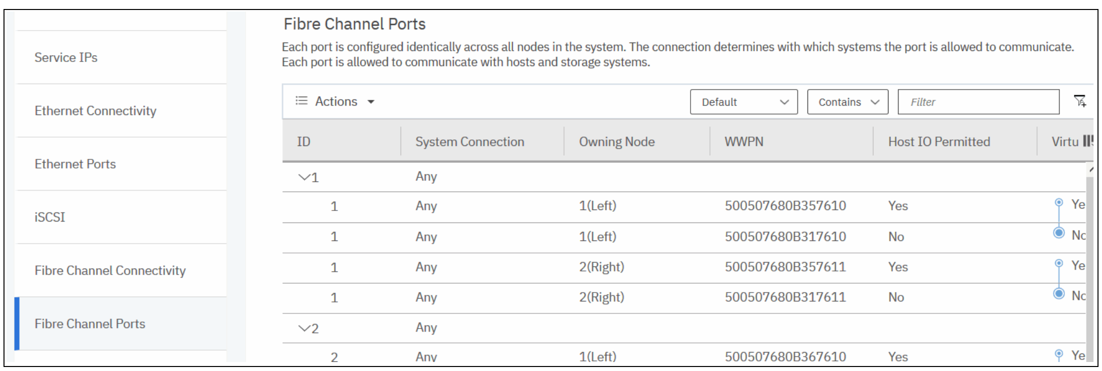


Figure 5-61 Viewing Fibre Channel Port properties

5.10.3 Security menu

Use the Security option from the Settings menu (as shown in Figure 5-62 on page 170) to view and change security settings, authenticate users, and manage secure connections.

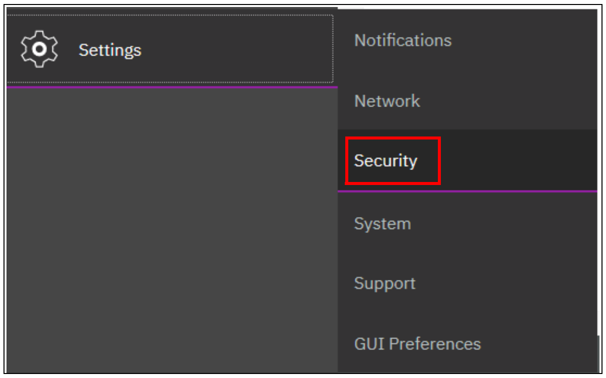


Figure 5-62 Security menu

Remote Authentication

In the Remote Authentication pane, you can configure remote authentication with LDAP as shown in Figure 5-63. By default, the Storwize V7000 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.

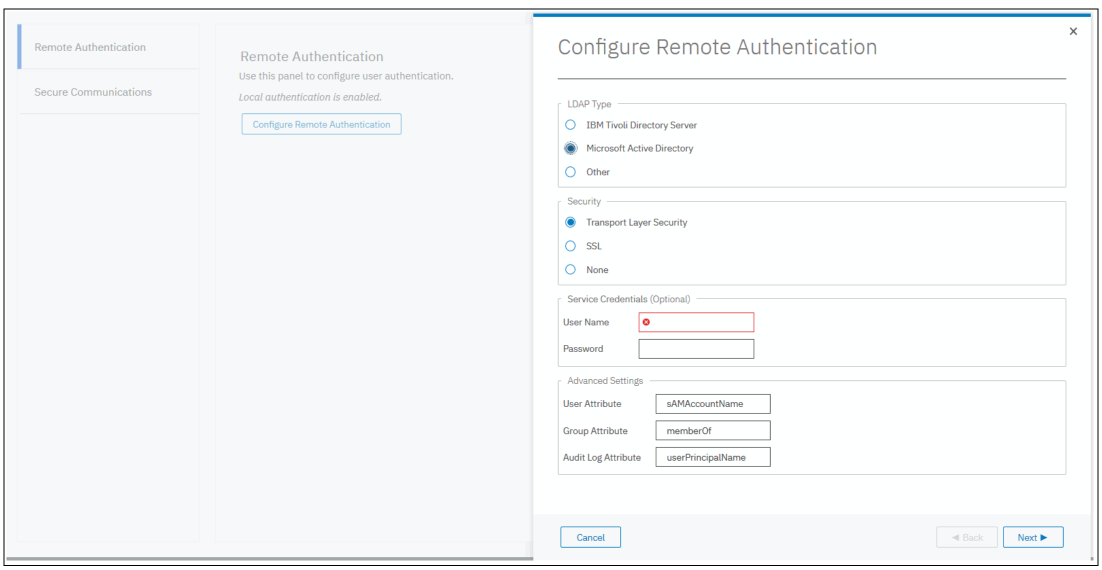


Figure 5-63 Configuring Remote Authentication

Secure Connections

To enable or manage secure connections, select the secure connections pane, as shown in Figure 5-64 on page 171. Before you create a request for either type of certificate, ensure that your current browser does not have restrictions on 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.

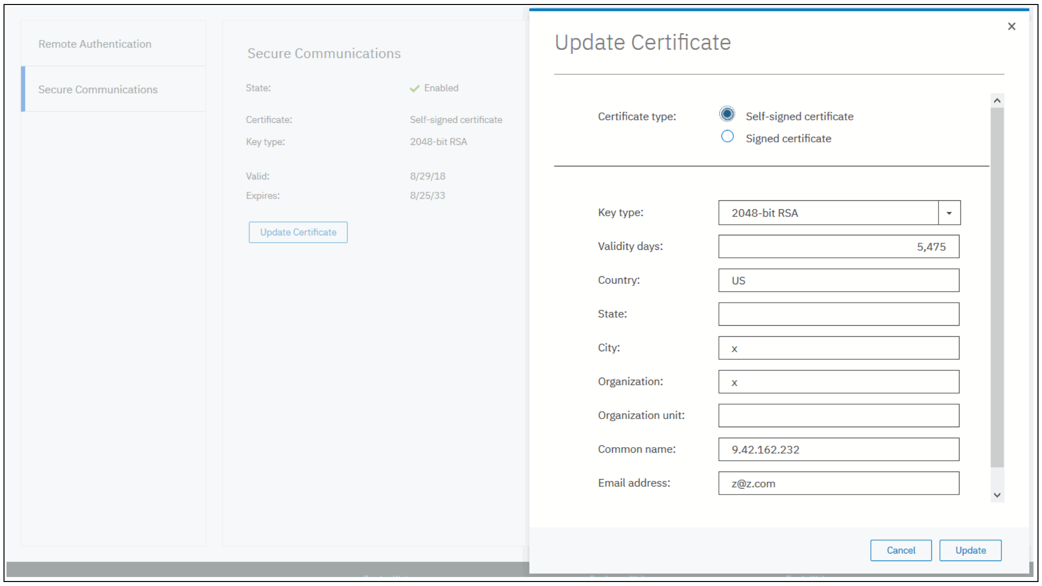


Figure 5-64 Configuring Secure Communications

5.10.4 System menus

Click the **System** option from the Settings menu (see Figure 5-65) to view and change the time and date settings, work with licensing options, download configuration settings, work with VMware VVOLs and IP Quorum, or download software upgrade packages.

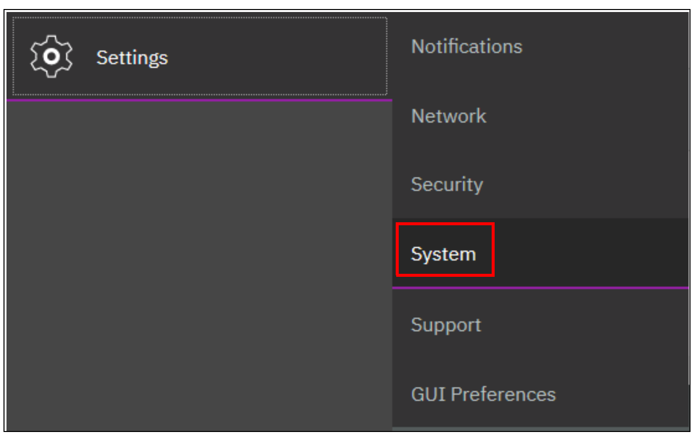


Figure 5-65 System option

Date and time

Complete the following steps to view or configure the date and time settings:

1. From the main System pane, point to **Settings** and click **System**.
2. In the left column, select **Date and Time**, as shown in Figure 5-66 on page 172.

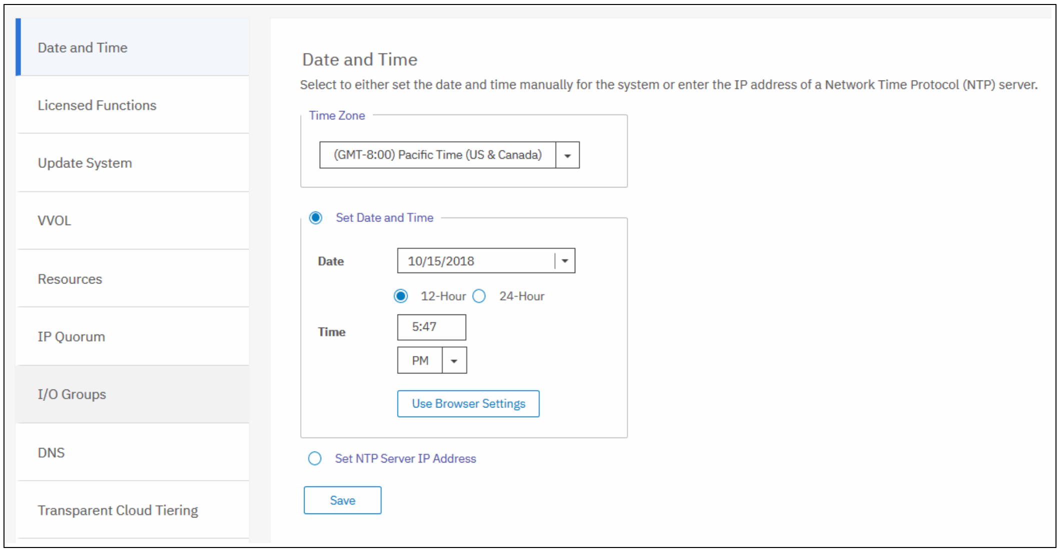


Figure 5-66 Date and Time window

3. From this pane, 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 5-67. You can click **Use Browser Settings** to automatically adjust the date and time of your Storwize V7000 system with your local workstation date and time.

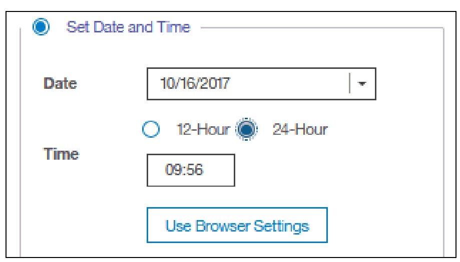


Figure 5-67 Set Date and Time window

     - 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 5-68.

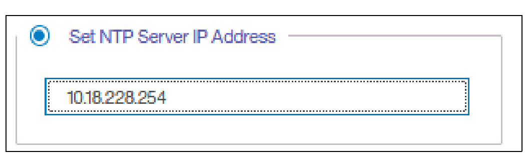


Figure 5-68 Set NTP Server IP Address window

4. Click **Save**.

Licensing

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.

The IBM Storwize V7000 supports 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.

Complete the following steps to view or configure the licensing settings:

1. From the main Settings pane, point to **Settings** and click **System**.
2. In the left column, select **Licensed Functions**, as shown in Figure 5-69.

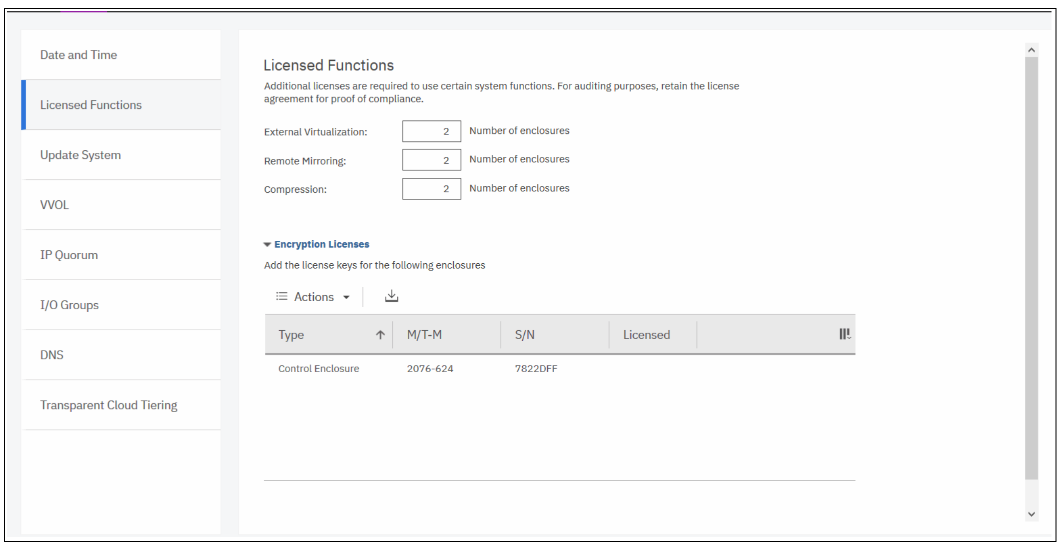


Figure 5-69 Licensing window

3. In the Licensed Functions pane, you can view or set the licensing options for the IBM Storwize V7000 for the following elements (limits are in TiB):

   – External Virtualization

     Specifies the number of enclosures of external storage systems that are attached to your IBM Storwize V7000. Data can be migrated from storage systems to your system that uses the external virtualization function within 45 days of purchase of the system without purchase of 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 managed disks for a quorum disk and are not providing any capacity for volumes.

   – Remote Mirroring Limit

     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.

     The remote-mirroring function is licensed per enclosure. It allows the use of 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 license settings apply only to the system on which you are configuring license settings. For remote-copy partnerships, a license is also required on any remote systems that are in the partnership.

- Real-time Compression Limit

  One license is required for each control or expansion enclosure and each enclosure in any external storage systems that use compression.

- Transparent Cloud Tiering

  Enter the total number of enclosures that are licensed to create snapshots of volume data to cloud storage.

- 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.

Update System

For more information about the update procedure that uses the GUI, see Chapter 13, “RAS, monitoring, and troubleshooting” on page 673.

VMware Virtual Volumes

IBM Spectrum Virtualize V7.6 and later can manage VMware vSphere VVOLs directly in cooperation with VMware. It enables VMware virtual machines to get the assigned disk capacity directly from IBM Storwize V7000 rather than from the ESXi data store. That technique enables storage administrators to control the appropriate usage of storage capacity, and to enable enhanced features of storage virtualization directly to the virtual machine (such as replication, thin-provisioning, compression, and encryption).

VVOL management is enabled in IBM Storwize V7000 in the System section, as shown in Figure 5-70. The NTP server must be configured before enabling VVOL management. It is strongly advised to use the same NTP server for ESXi and for IBM Storwize V7000.

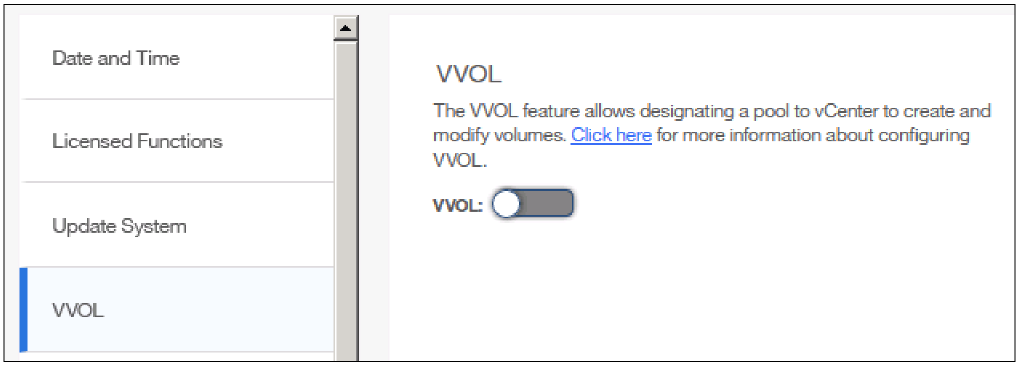


Figure 5-70 Enabling VVOLs management

Restriction: You cannot enable VVOLs support until the NTP server is configured in Storwize V7000.

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

IP Quorum

Starting with IBM Spectrum Virtualize V7.6, a new feature was introduced for enhanced stretched systems, the IP Quorum application. By using an IP-based quorum application as the quorum device for the third site, no Fibre Channel connectivity is required. Java applications run on hosts at the third site.

To start with IP Quorum, complete the following steps:

1. If your IIBM Storwize V7000 is configured with IP addresses version 4, click **Download IPv4 Application,** or select **Download IPv6 Application** for systems running with IP version 6. In our example, IPv4 is the option, as shown in Figure 5-71.

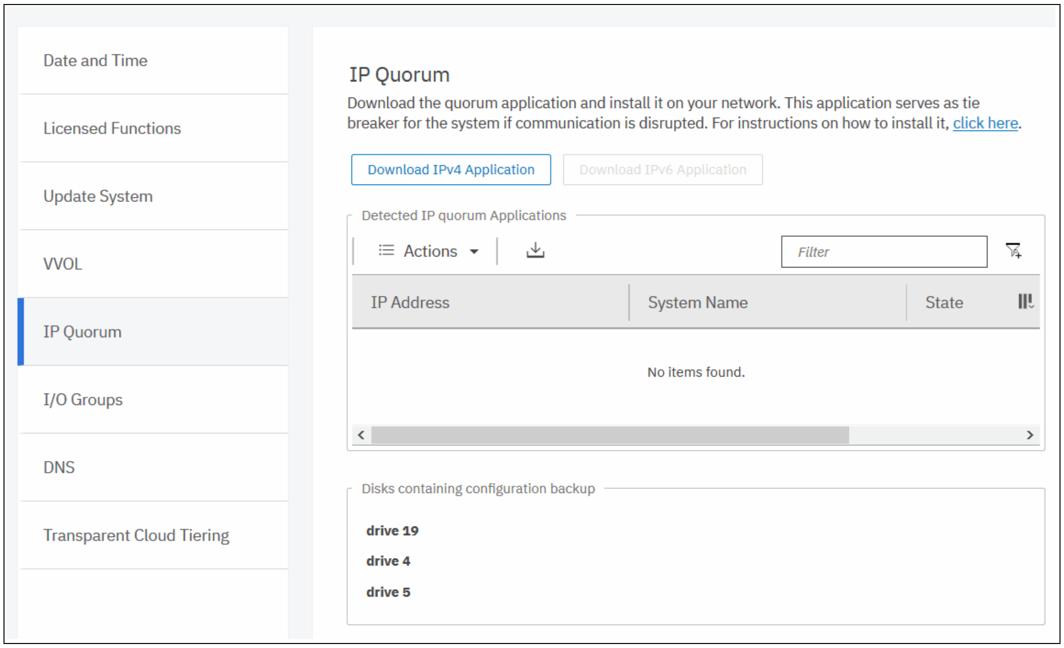


Figure 5-71 IP Quorum Settings

2. Click **Download IPv4 Application** and the IBM Spectrum Virtualize generates an IP Quorum Java application, as shown in Figure 5-72. The application can be saved and installed in a host that is to run the IP quorum application.

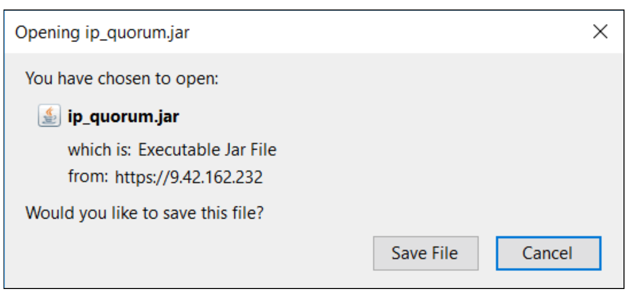


Figure 5-72 IP Quorum Java Application

3. On the host, you must use the Java command line to start the IP quorum application. Change to the folder where the application is located and run **java -jar ip_quorum.jar**.

I/O Groups

For ports within an I/O group, you can enable virtualization of Fibre Channel ports that are used for host I/O operations. With N_Port ID virtualization (NPIV), the Fibre Channel 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 around node unpends. 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 Fibre Channel and virtual ports that are currently 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 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 clicking **Settings** → **System** → **I/O Groups**, as shown in Figure 5-73.

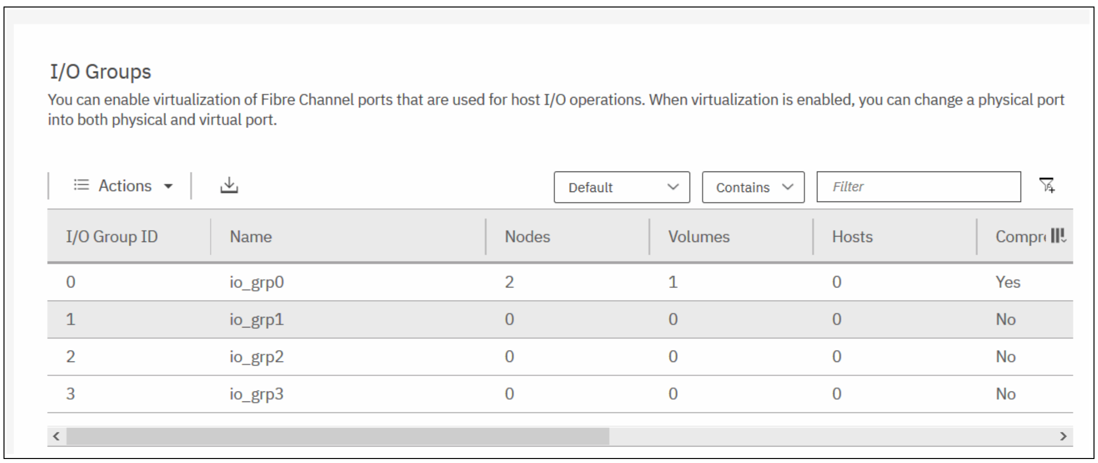


Figure 5-73 I/O Group port virtualization

You can change the status of the port by right-clicking the wanted I/O group and selecting **Change Target Port**, as shown in Figure 5-74.

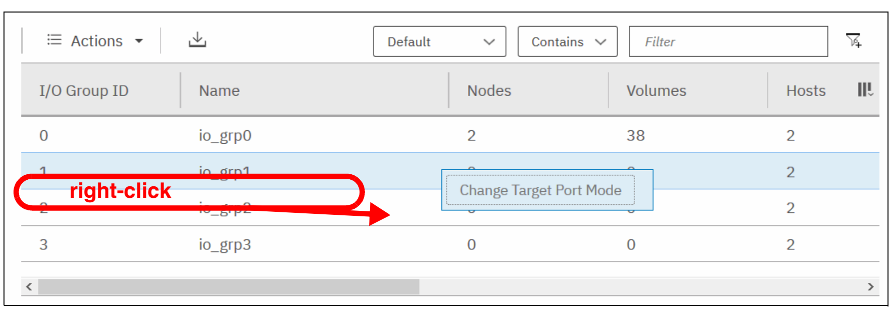


Figure 5-74 Changing port mode

Domain Name Server

Introduced within V7.8, IBM Spectrum Virtualize allows domain name server (DNS) entries to be manually set up in the IBM Storwize V7000. The information about the DNS servers in the IBM Storwize V7000 helps the system to access the DNS servers to resolve names of the computer resources that are in the external network.

To view and configure DNS server information in IBM Spectrum Virtualize, complete the following steps:

1. In the left pane, click the **DNS** icon and enter the **IP address** and the **Name** of each DNS server. The IBM Spectrum Virtualize supports up two DNS Servers, IPv4 or IPv6 (see Figure 5-75).

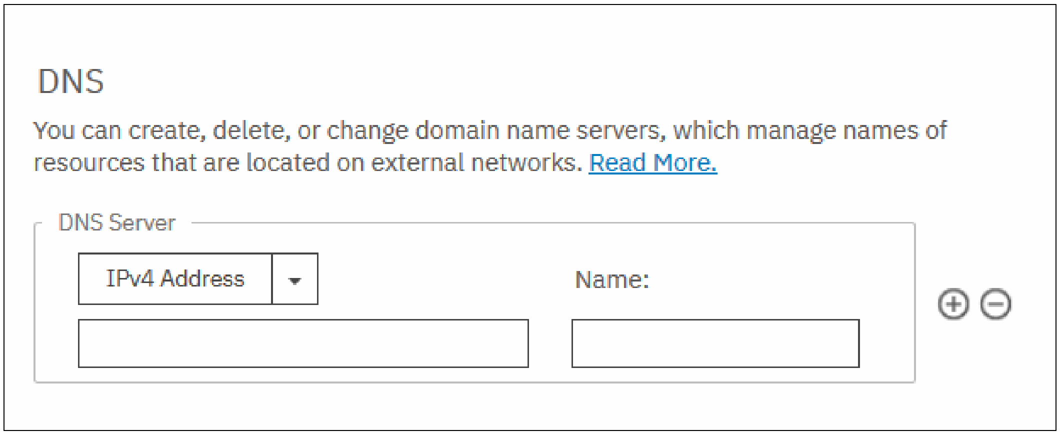


Figure 5-75 DNS information

2. Click **Save** after you complete entering the DNS server information.

Transparent Cloud Tiering

Transparent cloud tiering 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 to store copies of volume data in private or public cloud storage.

With transparent cloud tiering, administrators can move older data to cloud storage to free up capacity on the system. Point-in-time snapshots of data can be created on the system and then copied and stored on the cloud storage. An external cloud service provider manages the cloud storage, which reduces storage costs for the system. Before data can be copied to cloud storage, a connection to the cloud service provider must be created from the system.

A cloud account is an object on the system that represents a connection to a cloud service provider by using a particular set of credentials. These credentials differ depending on the type of cloud service provider that is being specified. Most cloud service providers require the host name of the cloud service provider and an associated password, and some cloud service providers also require certificates to authenticate users of the cloud storage.

Public clouds use certificates that are signed by well-known certificate authorities. Private cloud service providers can use self-signed certificate or a certificate that is signed by a trusted certificate authority. These credentials are defined on the cloud service provider and passed to the system through the administrators of the cloud service provider. A cloud account defines whether the system can successfully communicate and authenticate with the cloud service provider 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 cloud service provider. Migration between providers is not supported.

Note: Before enabling Transparent Cloud Tiering, consider the following requirements:

- Ensure that DNS server 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 encryption enabled.

Each cloud service provider requires different configuration options. The system supports the following cloud service providers:

- SoftLayer®
- OpenStack Swift
- Amazon S3

To view your IBM Spectrum Virtualize cloud provider settings, from the IBM Storwize V7000 Settings pane, point to **Settings** and click **System**. Then, select **Transparent Cloud Tiering**, as shown in Figure 5-76.

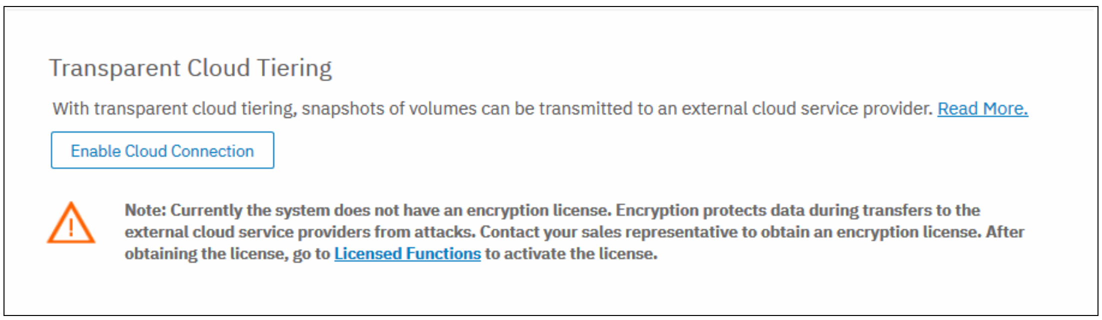


Figure 5-76 Transparent cloud tiering settings

B using this view, you can enable and disable features of your Transparent Cloud Tiering and update the system information concerning your cloud service provider. This pane allows you to set the following options:

- Cloud service provider
- Object Storage URL
- The Tenant or the container information that is associated to your cloud Object Storage
- User name of the cloud object account
- API Key
- The container prefix or location of your object
- Encryption
- Bandwidth

5.10.5 Support menu

Use the Support pane to configure and manage connections and upload support packages to the 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 service 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 5-77:

– 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 responsible for the system

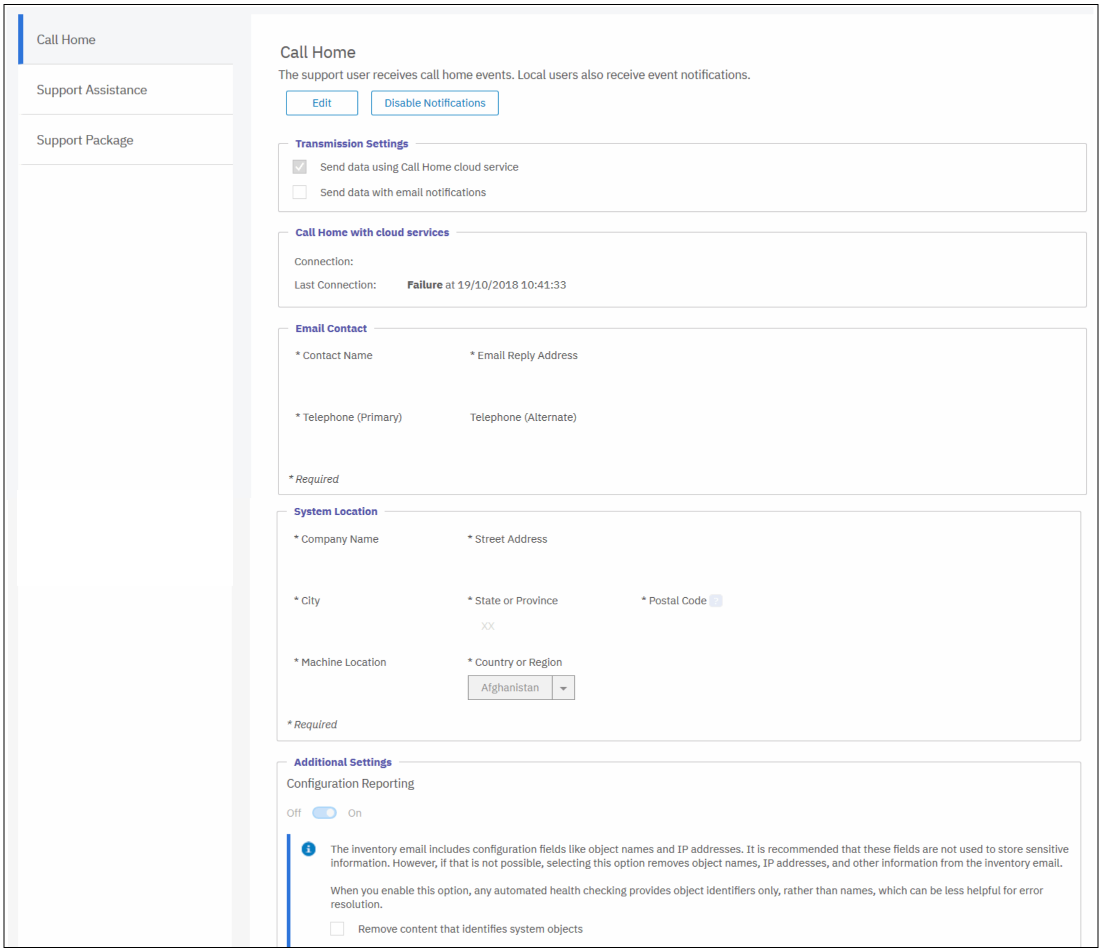


Figure 5-77 Call Home Settings

► Support assistance

This option enables support personnel to access the system to complete troubleshooting and maintenance tasks. You can configure e local support assistance, where 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 support center and system. More access controls can be added by the system administrator.

► Support Package

If support assistance is configured on your systems, you can automatically or manually upload new support packages to the support center to help analyze and resolve errors on the system.

The menus are available under **Settings** → **Support**, as shown in Figure 5-78.

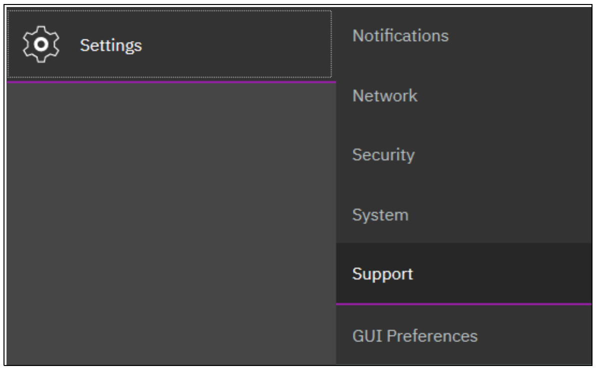


Figure 5-78 Support menu

More information about how the Support menu helps with troubleshooting your system or how to back up your systems, see Chapter 13, “RAS, monitoring, and troubleshooting” on page 673.

5.10.6 Setting GUI preferences

The GUI Preferences menu consists of the following options:

- Login
- General

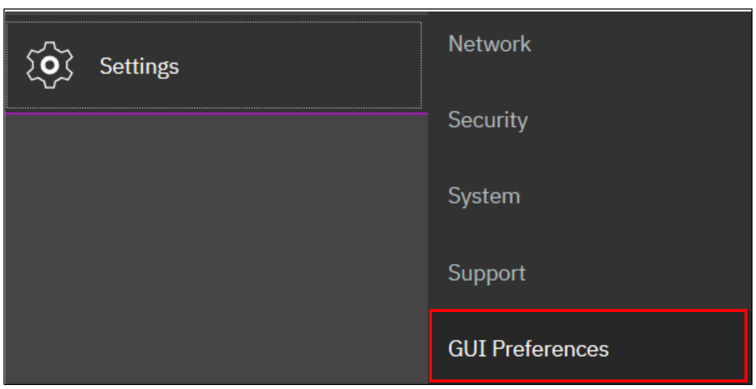


Figure 5-79 GUI Preferences

Login message

IBM Spectrum Virtualize V7.6 and later 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 5-80).

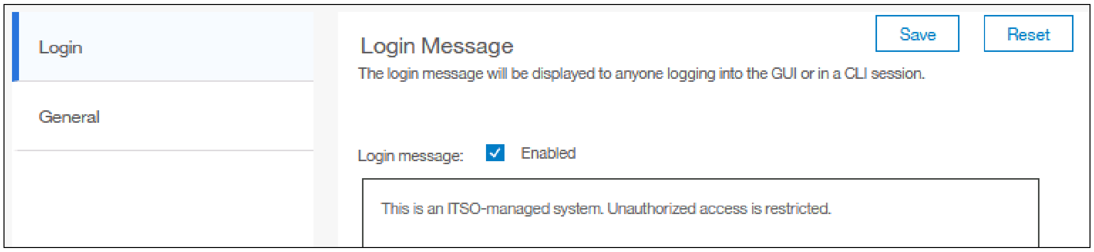


Figure 5-80 Enabling login message

The resulting log-in screen is shown in Figure 5-81.

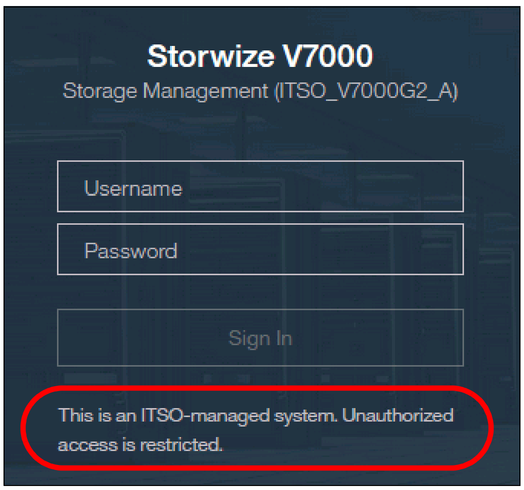


Figure 5-81 Welcome message in GUI

The banner message also appears in the CLI login prompt window, as shown in Figure 5-82.

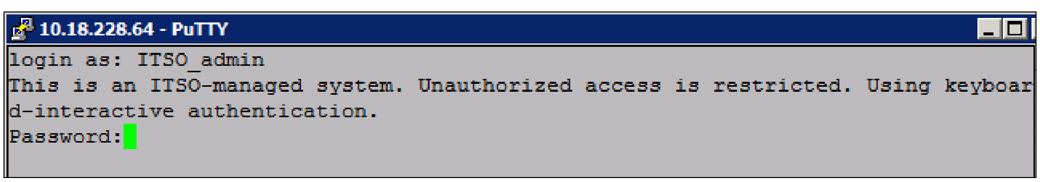


Figure 5-82 Banner message in CLI

General settings

The General Settings menu allows the user to refresh the GUI cache, set the low graphics mode option, and tenable advanced pools settings.

Complete the following steps to configure general GUI preferences:

1. From the Settings pane, point to **Settings** and click **GUI Preferences** (see Figure 5-83).

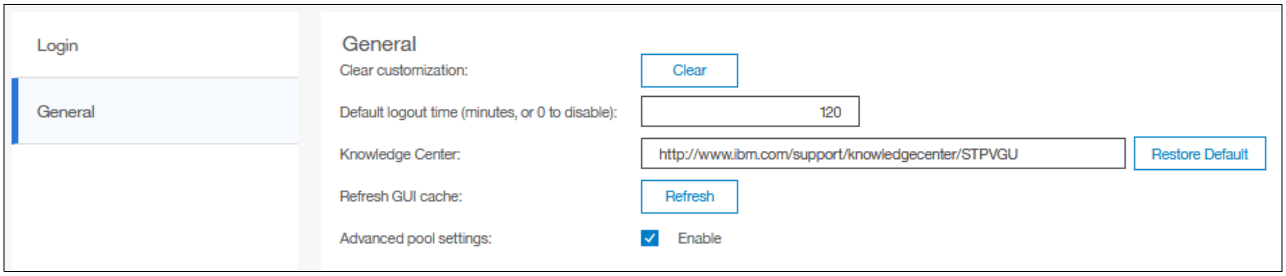


Figure 5-83 General GUI Preferences window

2. You can configure the following elements:
   - Refresh GUI cache

     This option causes the GUI to refresh all of 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.
   - Clear Customization

     This option deletes all GUI preferences that are stored in the browser and restores the default preferences.
   - IBM Knowledge Center

     You can change the URL of IBM Knowledge Center for IBM Spectrum Virtualize.
   - The accessibility option enables Low graphic mode when the system is connected through a slower network.
   - Advanced pool settings allow you to select the extent size during storage pool creation.
   - Default log out time in minutes after inactivity in the established session.

5.11 Additional frequent tasks in GUI

This section describes additional options and tasks available in the GUI of your IBM Storwize V7000 that are frequently used by administrators.

5.11.1 Renaming components

These sections provide guidance about how to rename your system and canisters.

Renaming IBM Storwize V7000 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. The use of the 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 exceptions exist. The name can be a maximum of 15 characters for Remote Copy relationships and groups. The `lsfabric` command displays long object names that are truncated to 15 characters for nodes and systems. V5.1.0 systems display truncated volume names when they are partnered with a version V6.1.0 or later system that has volumes with long object names (`lsrcrelationshipcandidate` or `lsrcrelationship` 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 called ABC and an MDisk called ABC, but you cannot have two volumes called ABC.
- The default object name is valid (object prefix with an integer).
- Objects can be renamed to their current names.

To rename the system from the System pane, complete the following steps:

1. Click **Actions** in the upper-left corner of the System pane, as shown in Figure 5-84.

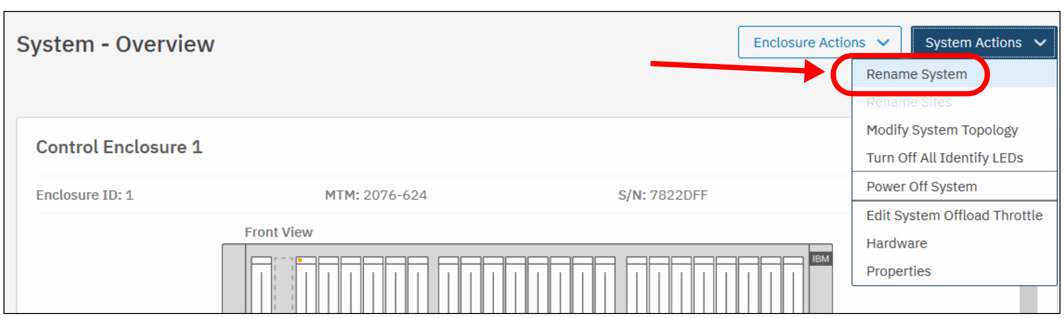


Figure 5-84 Actions on the System pane

2. The pane opens (see Figure 5-85 on page 185). Specify a new name for the system and click **Rename**.

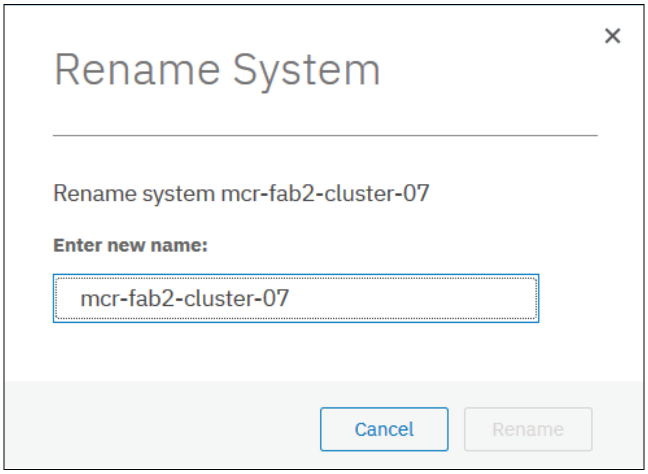


Figure 5-85   Rename the system

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.

3. If you are certain that you want to change the system name, click **Yes**.

Warning: When you rename your system, the iSCSI name (IQN) automatically changes because it includes system name by default. Therefore, this change needs more actions on iSCSI-attached hosts.

Rename a node canister

To rename a node canister, complete the following steps:

1. Navigate to the system overview window and right-click a node that you want to rename, as shown in Figure 5-86.

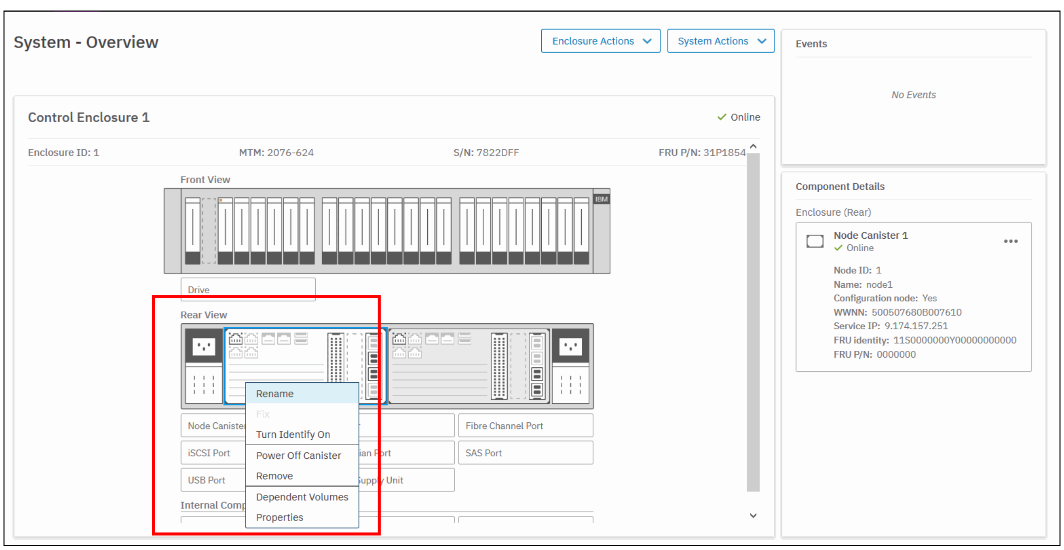


Figure 5-86   Renaming a node on the system overview window

2. Enter the new name of the node and click **Rename** (see Figure 5-87).

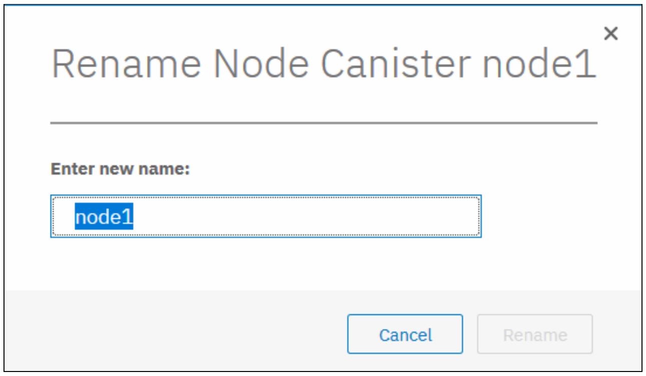


Figure 5-87 Enter the new name of the node

Warning: Changing the node canister name causes an automatic IQN update and requires the reconfiguration of all iSCSI-attached hosts.

5.11.2 Working with enclosures

The following section describes how to add or remove expansion enclosure to or from your IBM Storwize V7000 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 pane 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 5-88. The new enclosure is showing as unmanaged, which means it is not currently part of the system.

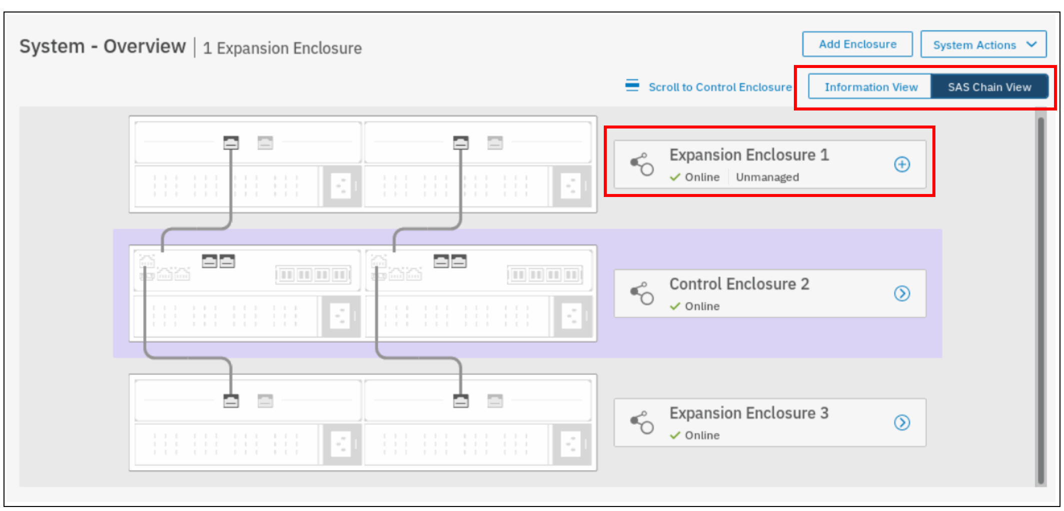


Figure 5-88 Newly detected expansion enclosure

2. At this point, you can select the **+** next to the enclosure you want to add or click the **Add Enclosure** button at the top. These buttons only show 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 5-89.

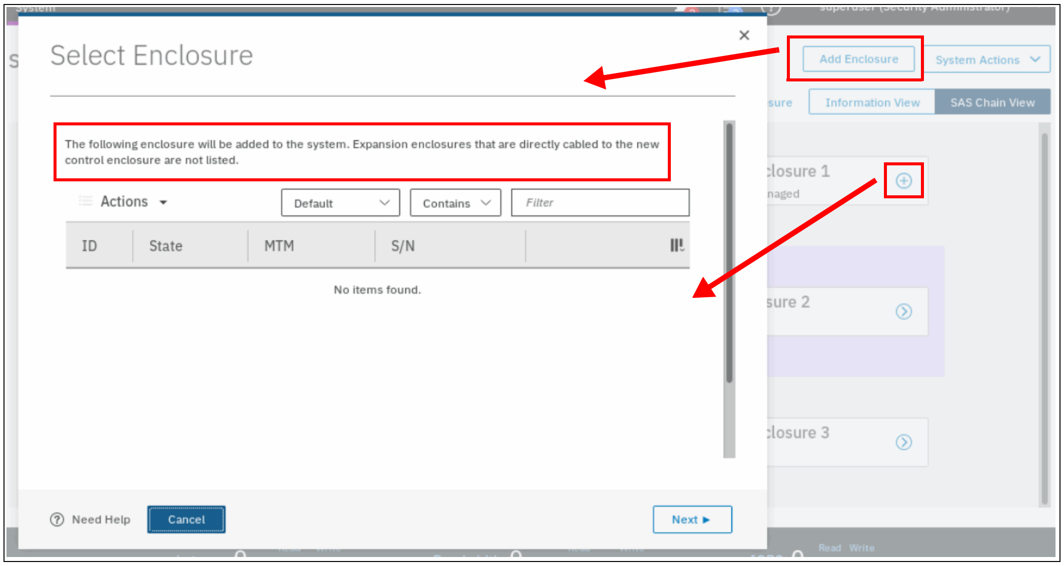


Figure 5-89 Adding an enclosure

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 5-90.

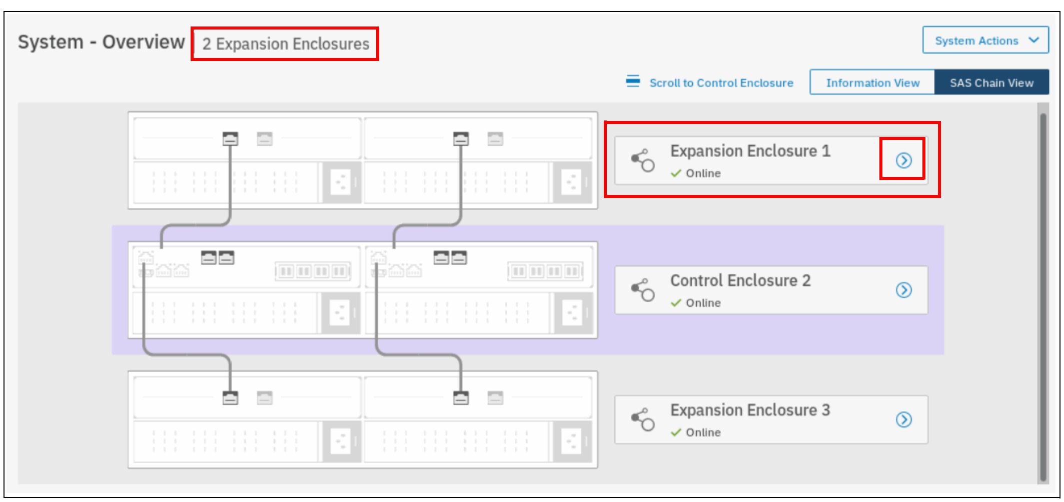


Figure 5-90 Enclosure successfully added

Removing enclosure

The enclosure removal procedure includes its logical detachment from the system by using GUI and physical unmount from the rack. The IBM Storwize V7000 guides you through this process. Complete the following steps:

1. In the System pane that is available from the Monitoring menu, select **>** of the enclosure you want to remove. The Enclosure Details pane opens. You can then click **Enclosure Actions** and select **Remove**, as shown in Figure 5-91.

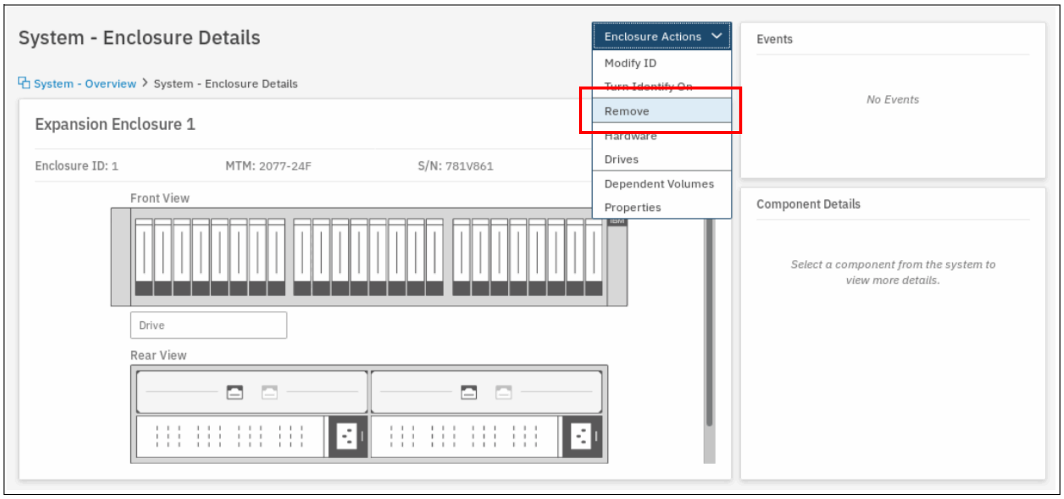


Figure 5-91 Selecting enclosure for removal

2. The system asks if you want to remove the enclosure. All disk drives in the removed enclosure must be in the *Unused* state. Otherwise, the removal process fails (see Figure 5-92).

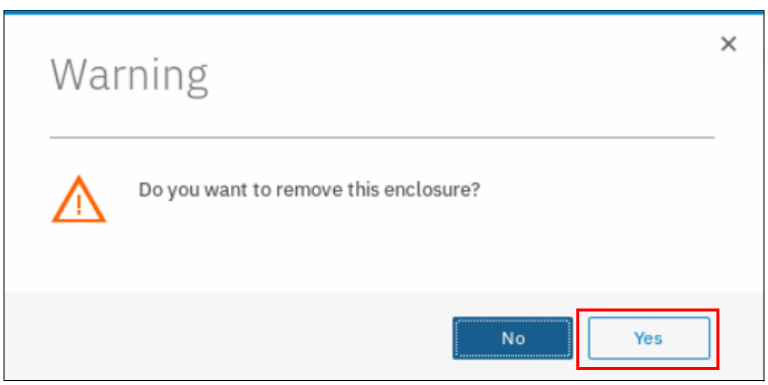


Figure 5-92 Confirm the removal

3. After the enclosure is logically removed from the system (set to *unmanaged* state), the IBM Storwize V7000 reminds you about steps necessary for physical removal, such as power off, uncabling, dismantling from the rack, and secure handling (see Figure 5-93).

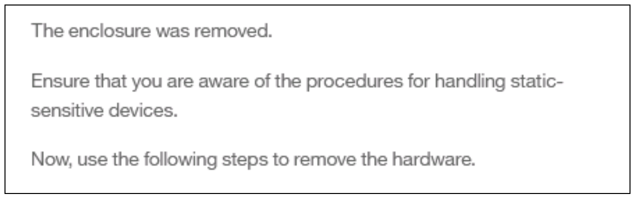


Figure 5-93 Enclosure removed

As part of the enclosure removal process, consult 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.

5.11.3 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, you log on to the service assistant and identify the configuration node, as shown in Figure 5-94.

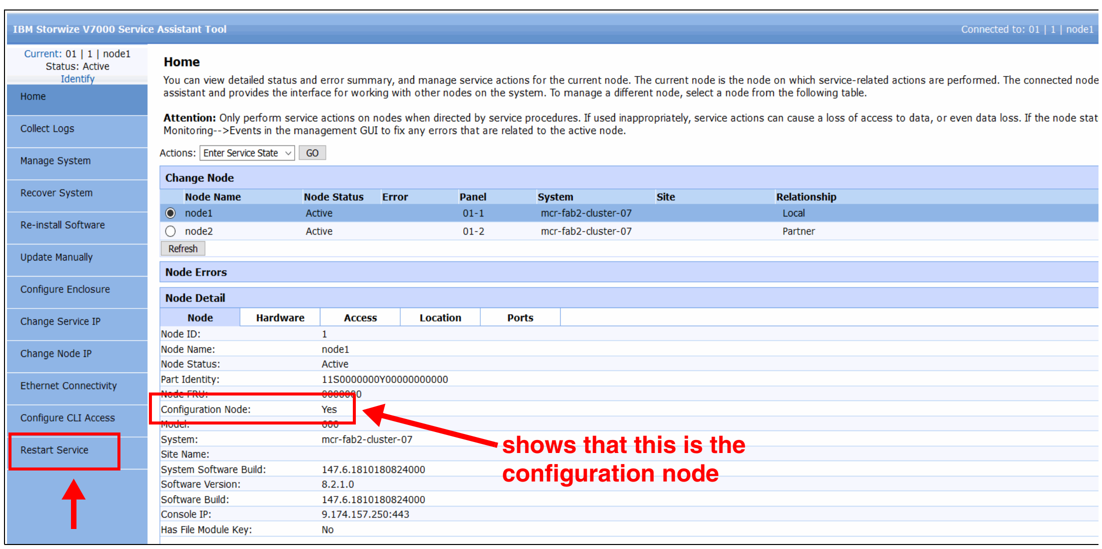


Figure 5-94 Identifying the configuration node on the service assistant

After the process is complete, you can navigate to **Restart Service**, as shown in Figure 5-95.

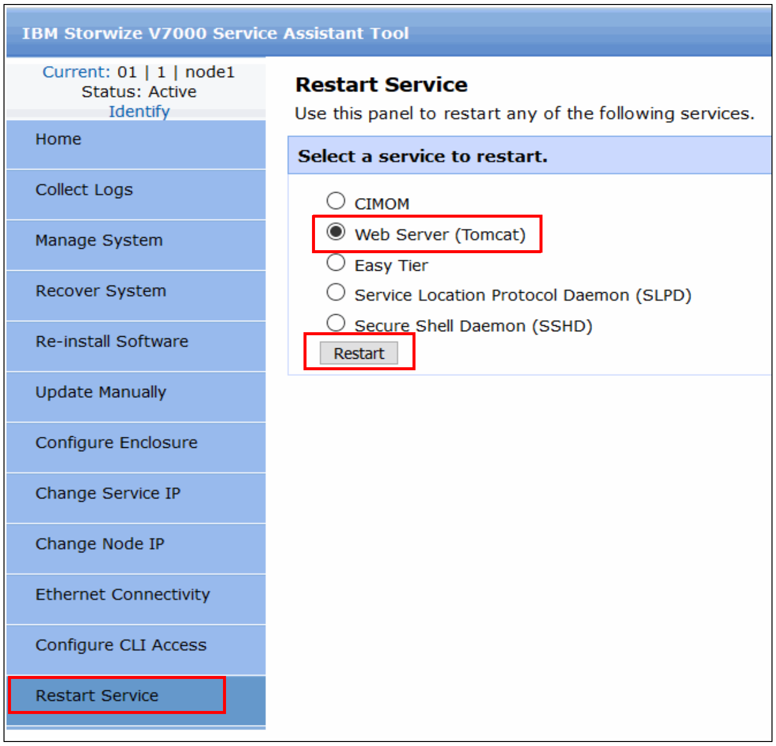


Figure 5-95 Restarting the Tomcat Web Server

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 is not affected). After 5 minutes, check to see whether GUI access was restored.

6

Storage pools

This chapter describes how IBM Storwize V7000 manages physical storage resources. All storage resources that are under the system control are managed by using *storage pools* or *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. Alternatively, you can configure the storage to your own requirements by using the command-line interface (CLI).

This chapter includes the following topics:

6.1 Working with storage pools

A managed disk (MDisk) is a logical unit (LU) of physical storage. MDisks are either RAID arrays that are constructed of internal storage disks or flash disk modules, or LUs that are exported from external storage systems. Storage pools act as a container for MDisks by dividing the MDisks into extents. Storage pools provision the available capacity from the extents to volumes.

Figure 6-1 provides an overview of how storage pools, MDisks, and volumes are related.

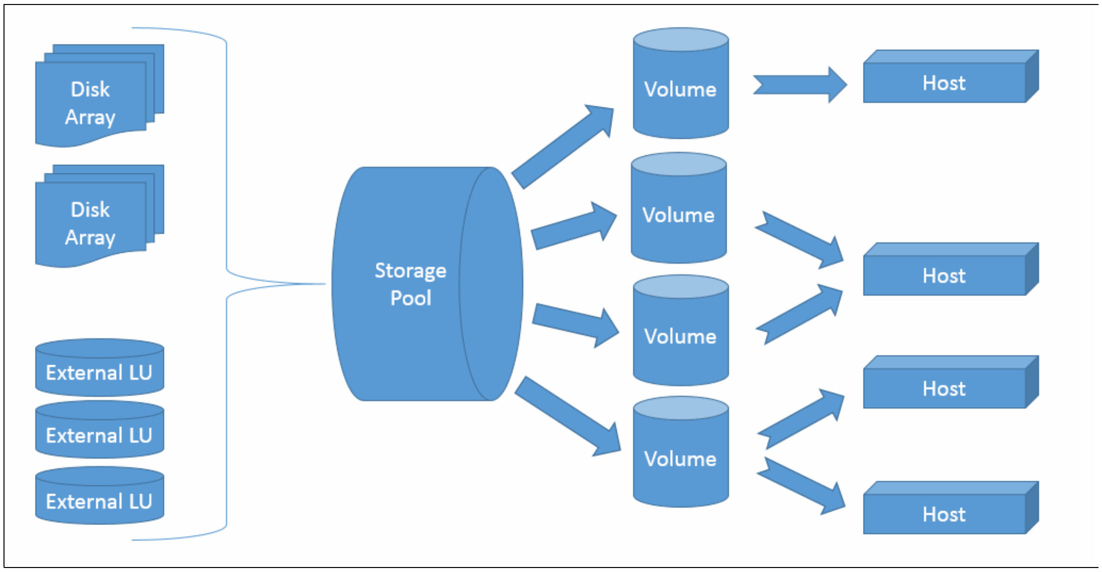


Figure 6-1 Relationship between MDisks, storage pools, and volumes

Storwize V7000 organizes storage into pools to ease storage management and make it more efficient. All MDisks in a pool are split into extents of the same size, and volumes are created out of the available extents. The extent size is a property of the storage pool and cannot be changed after the pool is created. It is possible to add MDisks to an existing pool to provide additional extents or to remove extents from the pool by deleting the MDisk.

Storage pools can be further divided into subcontainers that are called *child pools*. Child pools are created from existing capacity that is allocated to a parent pool. They have a similar extent size setting as their parent pool and can be used for volume operation.

When you create a storage pool, you can define it to have the “Data reduction” feature enabled to trigger creation of a *Data Reduction Pool (DRP)*. Data Reduction Pools increase the existing infrastructure capacity utilization by leveraging new efficiency functions. Data Reduction Pools automatically deallocate and reclaim capacity of thin-provisioned volumes that contain deleted data. This reclaimed capacity can be reused by other volumes. Data Reduction Pools also support data deduplication and compression features.

For more information on DRP planning and implementation, see *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430.

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. That way resources become free for more critical I/O operations.

You can create either an IOPS or bandwidth limit or both. You can use storage pool throttles to avoid overwhelming the back-end storage. Parent and child pool throttles are independent of each other. A child pool can have higher throttle limits than its parent pool.

In general, you manage storage as follows:

1. Create storage pools (standard or DRP), depending on your solution needs.
2. Add managed storage to pools by using one of the following options:
   – Create array MDisks out of internal drives or flash modules.
   – Assign MDisks provisioned from external storage systems to pools.
3. Create volumes on pools and map them to hosts or host clusters.

You manage storage pools either in the Pools pane of the GUI or with the CLI. To access the Pools pane, click **Pools** → **Pools**, as shown in Figure 6-2.

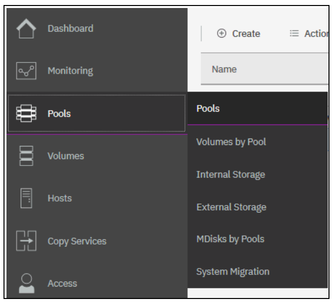


Figure 6-2 Accessing the Storage Pool pane

The pane 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 with CLI, run the **lsmdiskgrp** command without any parameters, as shown on Example 6-1.

Example 6-1 lsmdiskgrp output (some columns are not shown)

```
IBM_2145:ITSO-SV1:superuser>
id name  status mdisk_count vdisk_count capacity extent_size free_capacity
0  Pool0 online 1           5           821.00GB 1024        771.00GB
1  Pool1 online 1           5           7.21TB   4096        6.14TB
```

6.1.1 Creating storage pools

To create a storage pool, you can navigate to **Pools** → **MDisks by Pools** and click **Create Pool**, or navigate to **Pools** → **Pools** and click **Create**, as shown in Figure 6-3.

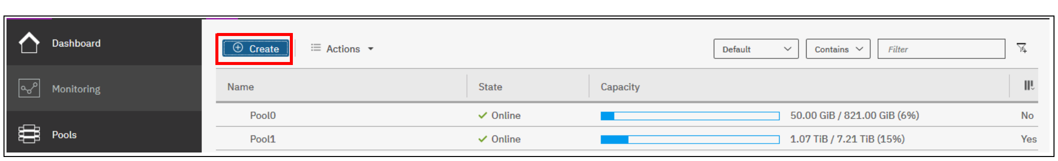


Figure 6-3 Option to create a storage pool in the Pools pane.

Both alternatives open the dialog box that is shown in Figure 6-4 on page 194.

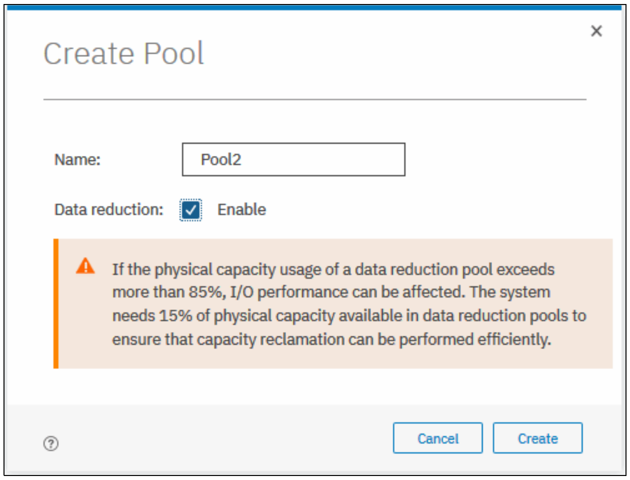


Figure 6-4 Create Pool dialog box

Mark the **Data Reduction** check box to create the Data Reduction Pool. Leaving it unmarked creates a standard storage pool.

A standard storage pool that is created by using the GUI has a default extent size of 1 GB. Data Reduction Pools have a default extent size of 4 GB. The size of the extent is selected at creation time and cannot be changed later.

If you want to specify a different extent size, you can enable this option by clicking **Settings** → **GUI Preferences** → **General** and checking **Advanced pool settings**, as shown in Figure 6-5.

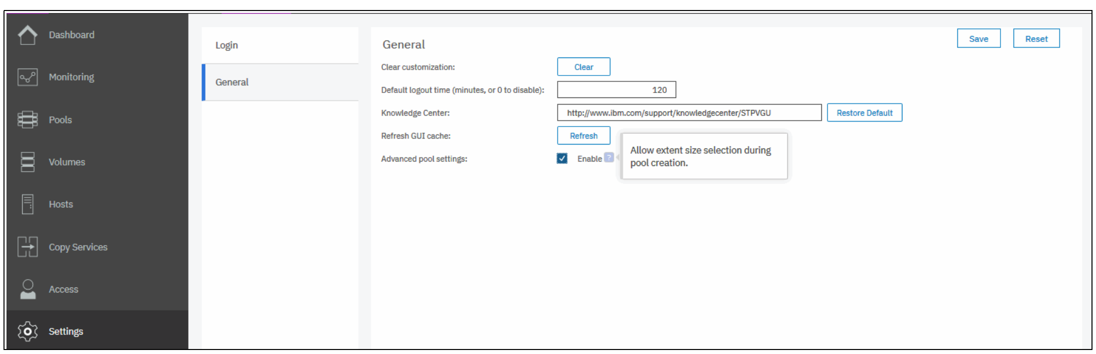


Figure 6-5 Advanced pool settings

When advanced pool settings are enabled, you can additionally select an extent size at creation time, as shown in Figure 6-6 on page 195.

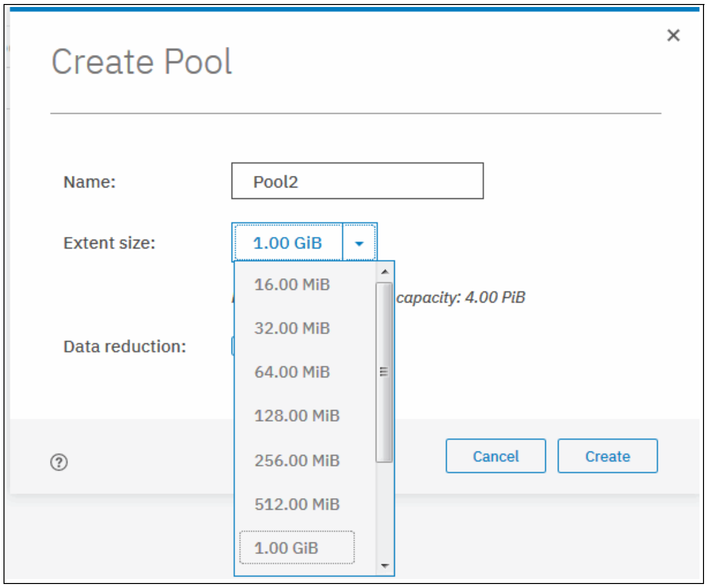


Figure 6-6   Creating a pool with advanced settings selected

Note: Do not create Data Reduction Pools with small extent sizes. See the following alert:
http://www.ibm.com/support/docview.wss?uid=ibm10718413

For more information about differences between standard pools and Data Reduction Pools, and for extent size planning, see Chapter 3, “Planning” on page 43.

If an encryption license is installed and enabled, you can additionally select whether the storage pool is encrypted, as shown in Figure 6-7. 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 603.

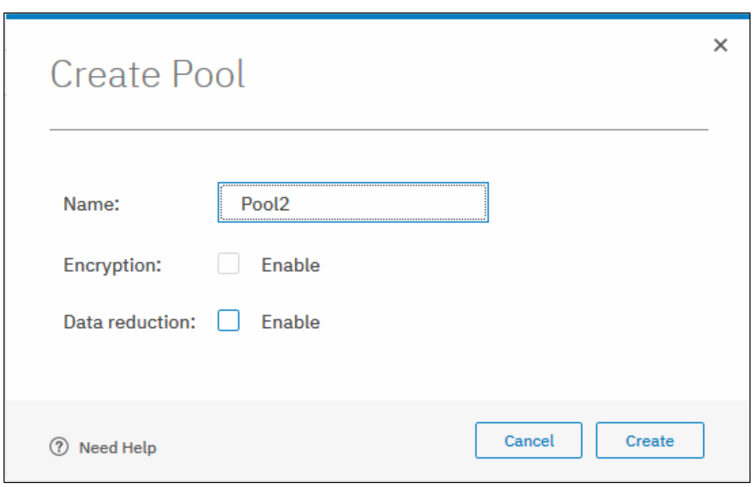


Figure 6-7   Creating a pool with Encryption enabled

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 calledvolumes that are calledvolumes called ABC.
- The default object name is valid (object prefix with an integer).
- Objects can be renamed to their current names.

The new pool is created and is included in the list of storage pools with zero bytes, as shown in Figure 6-8.

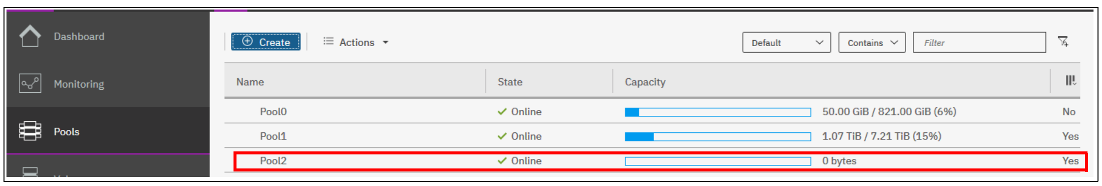


Figure 6-8 Newly created empty pool

To perform this task with the CLI, the **mkmdiskgrp** command is used. The only required parameter is extent size. It is specified with the **-ext** parameter, which must have one of the following values: 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, or 8192 (MB). To create Data Reduction Pool, specify **-datareduction yes**. In Example 6-2, the command creates DRP named “Pool2” with no MDisks in it.

Example 6-2 mkmdiskgrp command

```
IBM_Storwize:ITSOV7K:superuser>mkmdiskgrp -name Pool2 -datareduction yes -ext 8192
MDisk Group, id [2], successfully created
```

6.1.2 Managed disks in a storage pool

A storage pool is created as an empty container, with no storage assigned to it. Storage is then added in the form of **MDisks**. An MDisk can be either an array from internal storage (as an array of drives) or an 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. You cannot have an array that does not belong to any storage pool. They cannot be moved between storage pools. It is only possible to destroy an array by removing it from a pool and to re-create it with a new pool.

External MDisks can exist outside the pool. You can assign them to storage pools and remove them from storage pools. The MDisk object remains on a system, but its state (mode of operations) might change.

MDisks are managed by using the MDisks by Pools pane. To access the MDisks by Pools page, browse to **Pools** → **MDisks by Pools**, as shown in Figure 6-9.

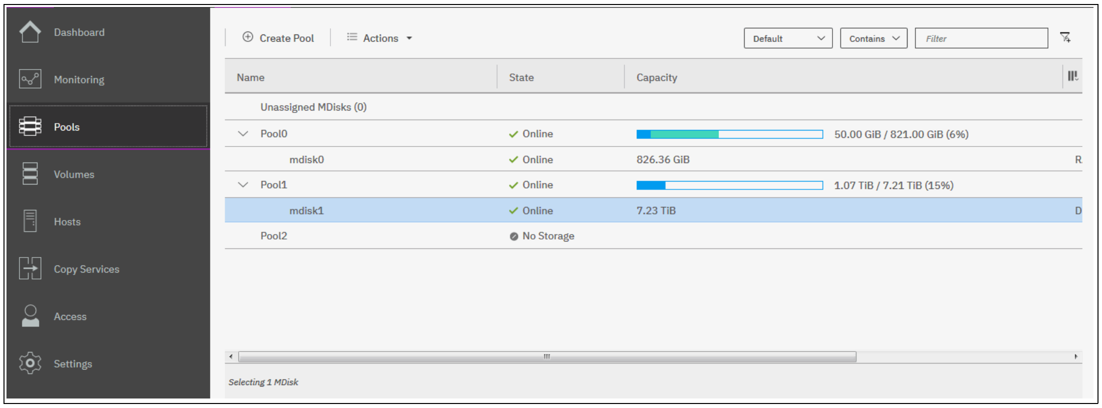


Figure 6-9 MDisks by pool

The pane lists all the MDisks available in the system under the storage pool to which they belong. Both arrays and external MDisks are listed. For instructions on operations with array MDisks, see 6.2, “Working with internal drives and arrays” on page 207. To implement a solution with external MDisks, see 6.3, “Working with external controllers and MDisks” on page 227.

To list all MDisks visible by the system with the CLI, use the `lsmdisk` command without any parameters. If required, you can filter output to include only external or only array type MDisks.

6.1.3 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 6-10. Alternatively, right-click the storage pool.

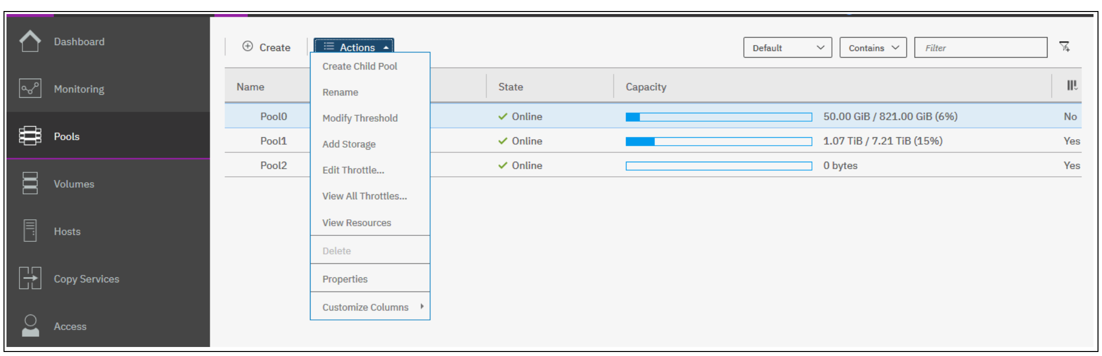


Figure 6-10 Pools actions menu

Create child pool

Select **Create Child Pool** to open a dialog window to create a child storage pool. For more information about child storage pools and a detailed description of this wizard, see 6.1.4, “Child storage pools” on page 203. It is not possible to create a child pool from an empty pool or from a Data Reduction Pool.

Rename

Select **Rename** to modify the name of a storage pool. Enter the new name and click **Rename** in the dialog window.

To do it with the CLI, use the **chmdiskgrp** command. Example 6-3 shows how to rename Pool2 to StandardStoragePool. If successful, the command returns no output.

Example 6-3 Using chmdiskgrp to rename storage pool

```
IBM_Storwize:ITSOV7K:superuser>chmdiskgrp -name StandardStoragePool Pool2
IBM_Storwize:ITSOV7K:superuser>
```

Modify Threshold

The storage pool threshold refers to the percentage of storage capacity that must be in use for a warning event to be generated. 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.

Note: The warning is generated only the first time that the threshold is exceeded by the used-disk 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 is indicated with a red bar as shown in Figure 6-11.

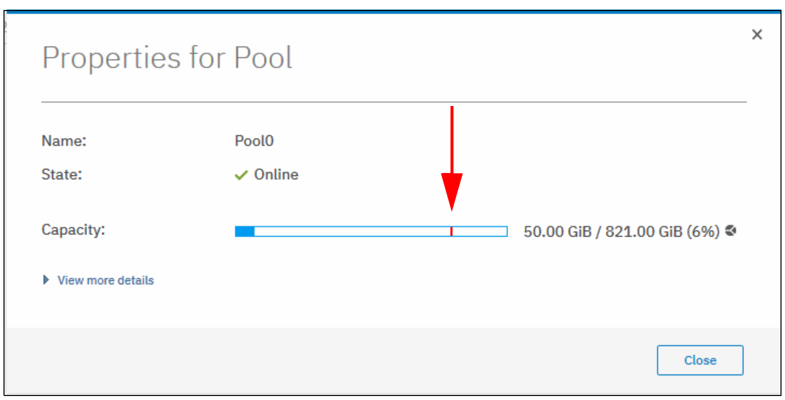


Figure 6-11 Pool properties

To do it with CLI, use the **chmdiskgrp** command. You can specify the threshold not only with a percentage of the storage pool size, as you can through the GUI. Also, you can set an exact size. Example 6-4 shows warning threshold set to 750 GB for Pool0.

Example 6-4 Changing warning threshold level with CLI

```
IBM_Storwize:ITSOV7K:superuser>chmdiskgrp -warning 750 -unit gb Pool0
IBM_Storwize:ITSOV7K:superuser>
```

Add storage

This action starts the configuration wizard, which assigns storage to the pool, as shown in Figure 6-12.

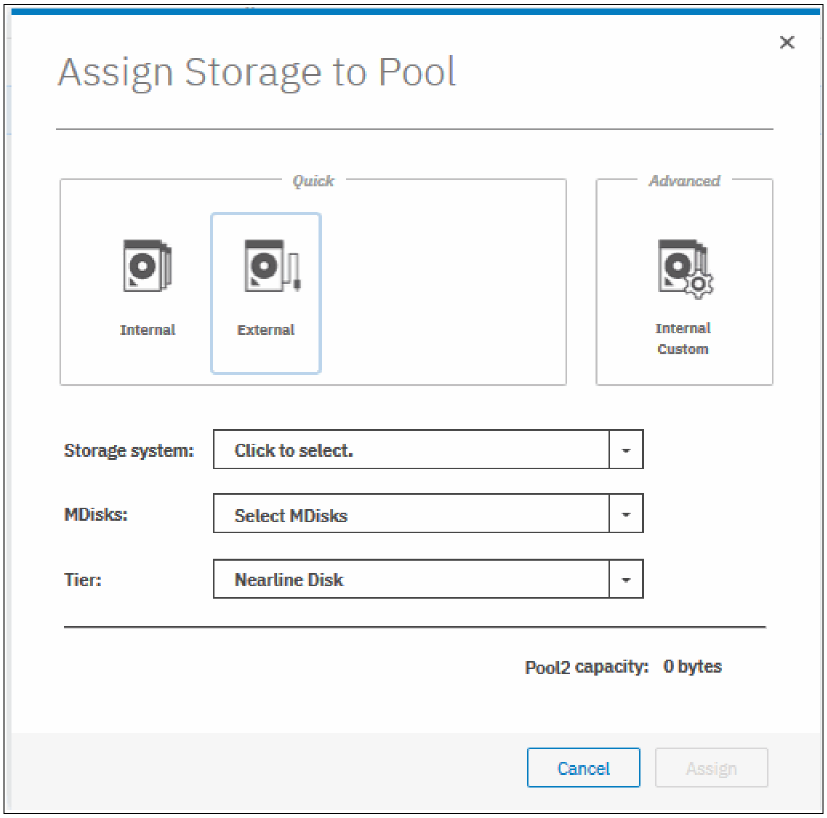


Figure 6-12 Add storage wizard

If **Internal** or **Internal Custom** is chosen, the system guides you through array MDisk creation. If **External** is selected, the system guides you through the selection of external storage MDisk. If no external storage is attached, the **External** option is not shown.

More details for each type of storage are provided in next sections.

Edit Throttle

Click this option to access the window where you set the Pool's throttle.

You can define a throttle for IOPS, bandwidth, or both as shown in Figure 6-13 on page 200:

- **IOPS limit** indicates the limit of configured IOPS (for both reads and writes combined).
- **Bandwidth limit** indicates the bandwidth, in megabytes per second (MBps) limit. You can also specify limit in GBps or TBps.

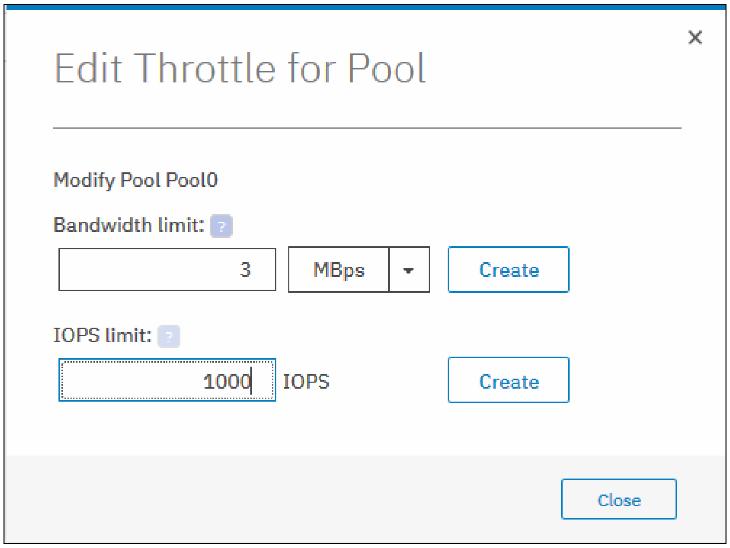


Figure 6-13 Editing throttle for a pool

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 200 MBps throttle is defined on a volume of that pool, the I/O operations are limited to 100 MBps.

A child pool throttle is independent of its parent pool throttle. Limits for a child can be higher than limits for a parent storage pool.

If a throttle already exists for the storage pool, the dialog box that is shown in Figure 6-13 also shows the **Remove** button that is used to delete the throttle.

To set storage pool throttle with CLI, use the **mkthrottle** command. Example 6-5 shows a storage pool throttle, named iops_bw_limit, set to 3 Mbps and 1000 IOPS, on Pool0.

Example 6-5 Setting storage pool throttle with CLI

```
IBM_Storwize:ITSOV7K:superuser>mkthrottle -type mdiskgrp -iops 1000 -bandwidth 3
-name iops_bw_limit -mdiskgrp Pool0
Throttle, id [0], successfully created.
```

To remove a throttle with CLI, use the **rmthrottle** command. It needs throttle ID or throttle name to be supplied as an argument, as shown on Example 6-6. Command returns no feedback if executed successfully.

Example 6-6 Removing pool throttle with CLI

```
IBM_Storwize:ITSOV7K:superuser>rmthrottle iops_bw_limit
IBM_Storwize:ITSOV7K:superuser>
```

View all Throttles

You can display defined throttles from Pools pane. 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 6-14 on page 201.

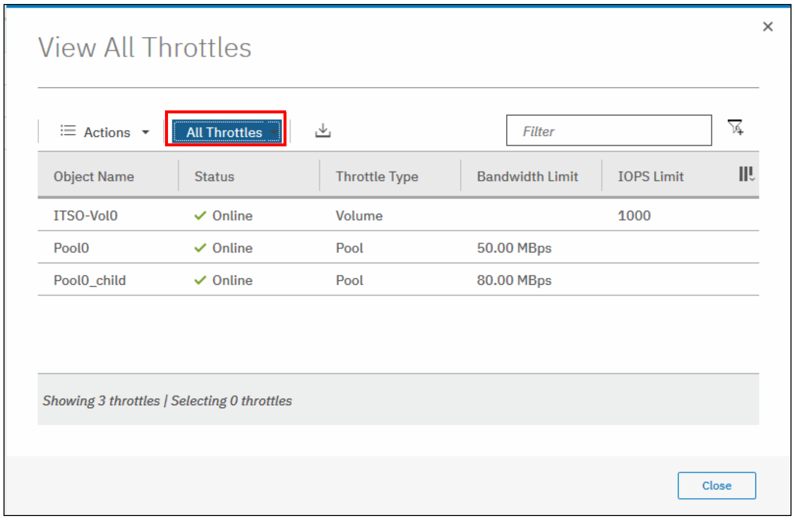


Figure 6-14 Viewing all throttles

To see a list of created throttles with CLI, use **`lsthrottle`**. When you run the command without arguments, it displays a list of all throttles on the system. To list only storage pool throttles, specify **`-filtervalue throttle_type=mdiskgrp`** parameter.

View Resources

Click this option to browse a list of Mdisks that are included into the storage pool, as shown in Figure 6-15.

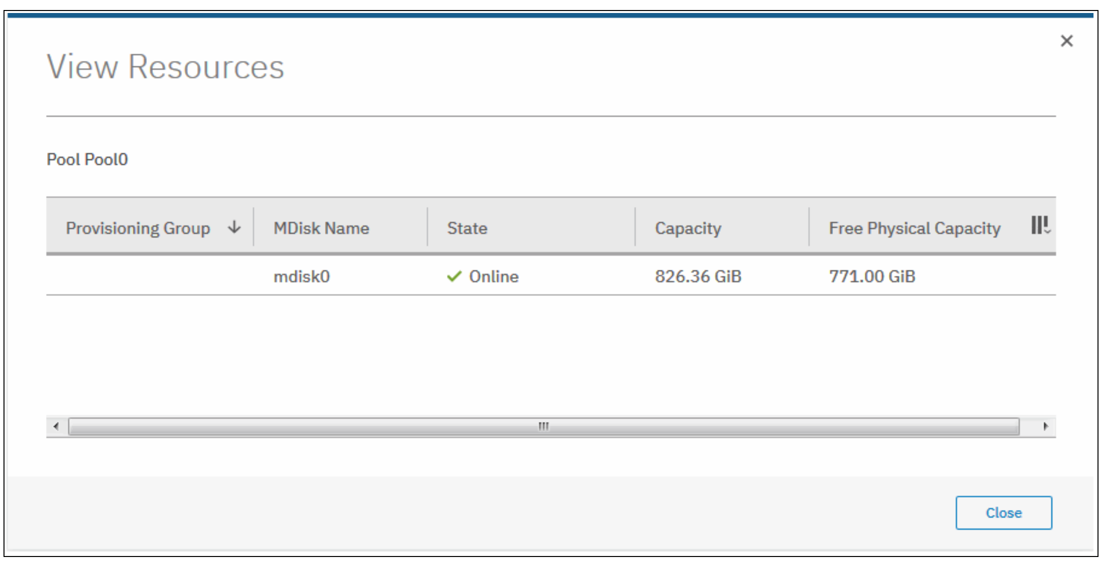


Figure 6-15 List of resources in the storage pool

To list storage pool resources with CLI, use **`lsmdisk`**. You can filter output to display MDisk objects that belong only to a single MDisk group (storage pool), as shown in Example 6-7 on page 202.

Example 6-7   Using lsmdisk (some columns are not shown)

```
IBM_Storwize:ITSOV7K:superuser>lsmdisk -filtervalue mdisk_grp_name=Pool0
id name   status mode  mdisk_grp_id mdisk_grp_name capacity ctrl_LUN_#
0  mdisk0 online array 0            Pool0          826.4GB
```

Delete

A storage pool can be deleted with 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, **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 7, “Volumes” on page 241.

After you delete a pool, the following actions occur:

- All the external MDisks in the pool return to a status of *Unmanaged*.
- All the array mode MDisks in the pool are deleted and all member drives return to a status of *Candidate*.

To delete storage pool with CLI, use `rmmdiskgrp`.

> **Note:** Be extremely careful when you use `rmmdiskgrp` command with `-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 these cannot be recovered.

Properties

Select **Properties** to display information about the storage pool. Additional information is available if you expand the **View more details** section. By hovering over the elements of the window and clicking **[?]**, you see a short description of each property, as shown in Figure 6-16.

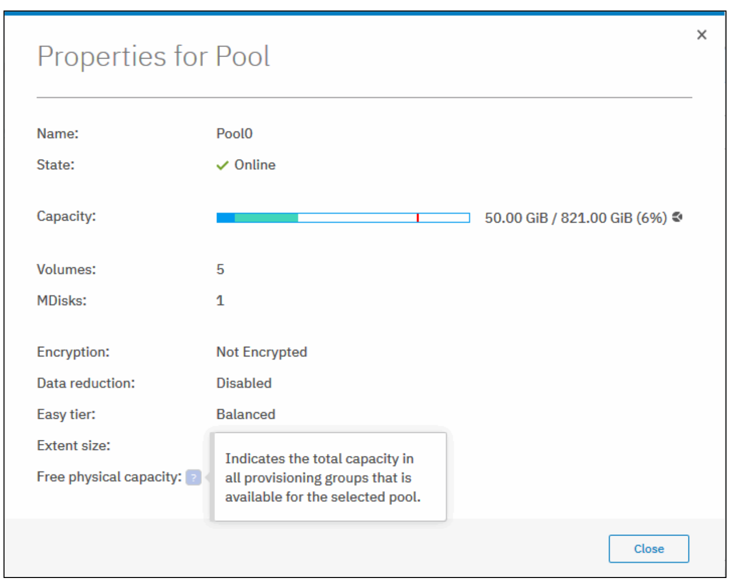


Figure 6-16   Pool properties and details

Use **lsmdiskgrp** with storage pool name or ID as a parameter to display detailed information on it with CLI, as shown in Example 6-8.

Example 6-8 lsmdiskgrp output (partially shown)

```
IBM_Storwize:ITSOV7K:superuser>lsmdiskgrp Pool0
id 0
name Pool0
status online
mdisk_count 1
vdisk_count 5
capacity 821.00GB
extent_size 1024
free_capacity 571.00GB
<...>
```

6.1.4 Child storage pools

A *child storage pool* is a storage pool that is created within a storage pool. The storage pool in which the child storage pool is created is called *parent storage pool*.

Unlike a parent pool, a child pool does not contain MDisks. Its capacity is provided exclusively by the parent pool in the form of extents. The capacity of a child 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 must be smaller than the free capacity of the parent pool.

A child pool cannot be created from a data reduction pool.

Child pools are useful when the capacity allocated to a specific set of volumes must be controlled. For example, child pools can be used with VMware vSphere Virtual Volumes (VVols). Storage administrators can restrict access of VMware administrators to only a part of the storage pool and prevent volumes creation from affecting the rest of the parent storage pool.

Child pools can also be useful when strict control over thin-provisioned volumes expansion is needed. For example, you might create a child pool with no volumes in it to act as an emergency set of extents. That way, if the parent pool ever runs out of free extent, you can use the ones from the child pool.

Child pools can also be used when a different encryption key is needed for different sets of volumes.

Child pools inherit most properties from their parent pools, and these cannot be changed. The inherited properties include these:

- Extent size
- Easy Tier setting
- Encryption setting, but only if the parent pool is encrypted

Creating a child storage pool

To create a child pool, browse to **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 6-17 on page 204.

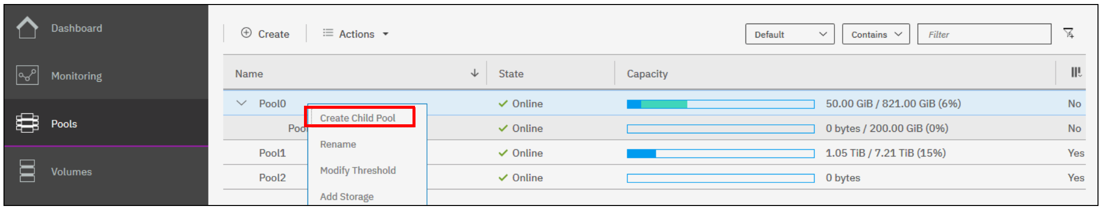


Figure 6-17 Creating a child pool

When the dialog window opens, enter the name and capacity of the child pool and click **Create**, as shown in Figure 6-18.

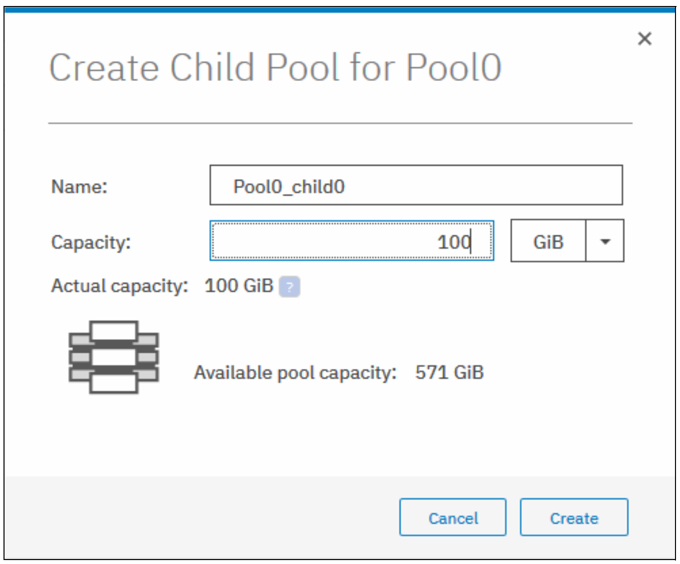


Figure 6-18 Defining a child pool

After the child pool is created, it is listed in the Pools pane 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 6-19.

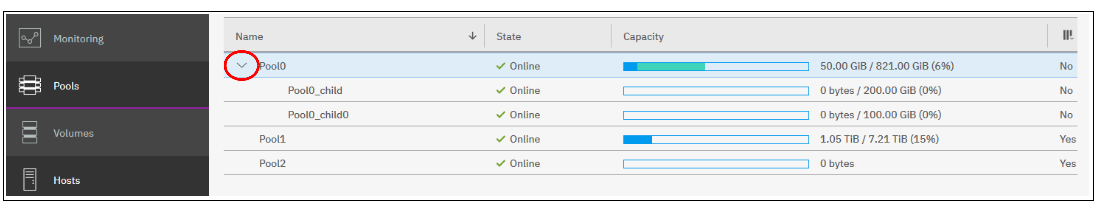


Figure 6-19 Listing parent and child pools

Creation of a child pool within a child pool is not possible.

In the CLI, the same **mkmdiskgrp** command that was used for parent pools is used to create child pools. You need to specify the parent pool for your new child pool and its size, as shown in Example 6-9. Size is given in MB and should be equal to parent's pool extent size multiplied by an integer. In this case, it is 100 * 1024MB = 100 GB.

Example 6-9 mkmdiskgrp for child pools

```
IBM_Storwize:ITSOV7K:superuser>mkmdiskgrp -parentmdiskgrp Pool0 -size 102400 -name
Pool0_child0
```

```
MDisk Group, id [4], successfully created
```

Actions on child storage pools

You can **Rename**, **Resize**, and **Delete** child pool. Also, it is possible to modify the warning threshold for it and edit the pool throttle. To select an action, right-click the child storage pool, as shown in Figure 6-20. Alternatively, select the storage pool and click **Actions**.

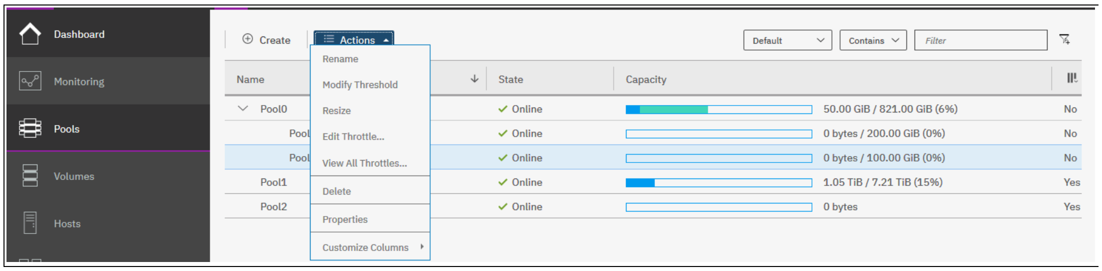


Figure 6-20 Actions on child pools

Select **Resize** to increase or decrease the capacity of the child storage pool, as shown in Figure 6-21. Enter the new pool capacity and click **Resize**.

Note: You cannot shrink a child pool below its real capacity. Thus, the new size of a child pool needs to be larger than the capacity used by its volumes.

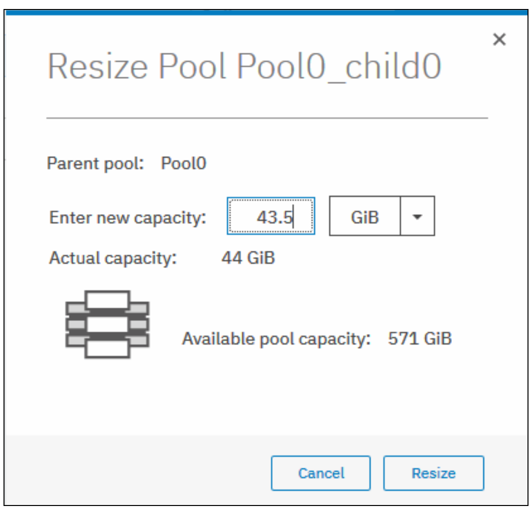


Figure 6-21 Resizing a child pool

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, use **chmdiskgrp**. Example 6-10 on page 206 renames child pool Pool0_child0 to Pool0_child_new and reduces its size to 44 GB. If successful, the command returns no feedback.

Example 6-10 chmdiskgrp to rename child pool

```
IBM_Storwize:ITSOV7K:superuser>chmdiskgrp -name Pool0_child_new -size 45056
Pool0_child0
IBM_Storwize:ITSOV7K:superuser>
```

Deleting a child pool is a task similar to deleting a parent pool. As with a parent pool, the **Delete** action is disabled if the child pool contains volumes, as shown in Figure 6-22.

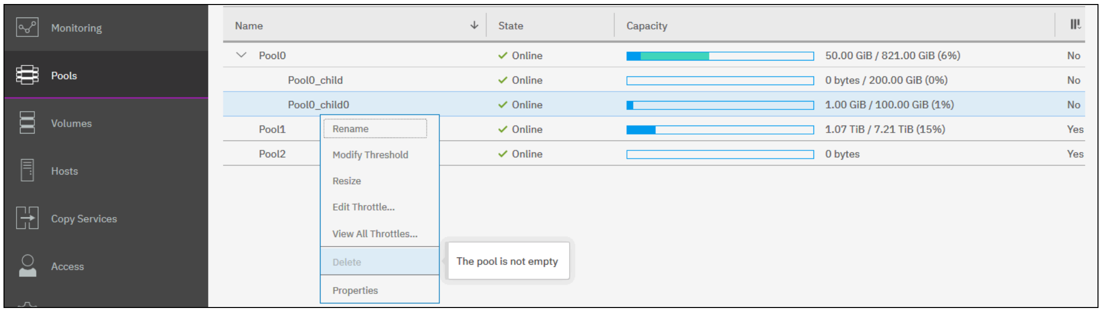


Figure 6-22 Deleting a child pool

After you delete a child pool, the extents that it occupied return to the parent pool as free capacity.

To delete a child pool with the CLI, use **`rmmdiskgrp`**.

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 you use them for parent pools. For more information about volume migration and volume mirroring, see Chapter 7, “Volumes” on page 241.

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. The volume can then 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 same parent), there is no data movement but only extent reassignments.

Volume migration between a child storage pool and its parent storage pool can be performed in the **Volumes** menu, on the **Volumes by Pool** page. Right-click a volume and choose to migrate it into a suitable pool.

In the example on Figure 6-23 on page 207, volume `child_volume` has been created in child pool `Pool0_child0`. Notice that child pools appear exactly like parent pools in the Volumes by Pool pane.

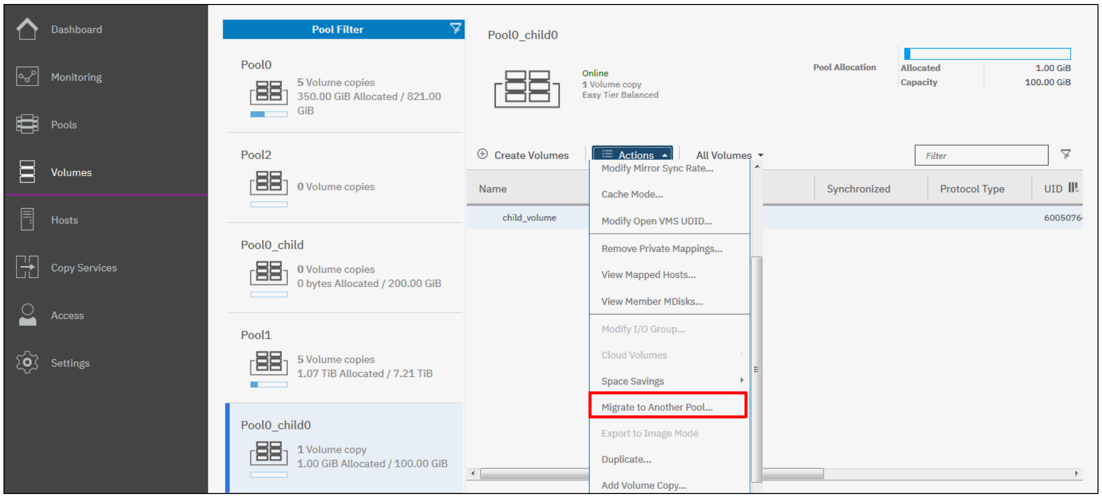


Figure 6-23 Actions menu in Volumes by pool

For CLI commands for migrating volumes to and from child pools, see Chapter 7, “Volumes” on page 241.

6.1.5 Encrypted storage pools

Storwize V7000 supports two types of encryption: hardware encryption and software encryption.

Hardware encryption is implemented at an array level, whereas software encryption is implemented at a storage pool level. For more information about encryption and encrypted storage pools, see Chapter 12, “Encryption” on page 603.

6.2 Working with internal drives and arrays

An array is a type of MDisk that is made up of disk drives (or flash drive modules); these drives are members of the array. A Redundant Array of Independent Disks (RAID) is a method of configuring member drives to create high availability and high-performance groups. The system supports nondistributed (traditional) and distributed array configurations.

6.2.1 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 pane for managing all internal drives. To access the Internal Storage pane, navigate to **Pools** → **Internal Storage**, as shown in Figure 6-24 on page 208.

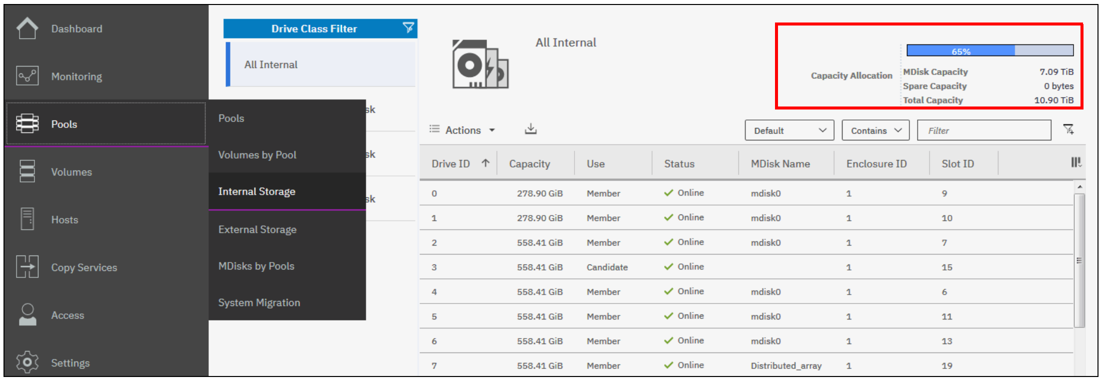


Figure 6-24 Internal storage pane

The pane gives an overview of the internal drives in the system. Select **All Internal** in the drive class filter to display all drives that are managed in the system, including all I/O groups and expansion enclosures.

Select **All Internal** under the Drive Class Filter to display all the drives that are installed in the managed system, including attached expansion enclosures. Alternatively, you can filter the drives by their type or class. For example, you can choose to show only Enterprise drive class (serial-attached Small Computer System Interface (SCSI) or (SAS)), Nearline SAS, or Flash drives. Select the class of the drives on the left side of the pane to filter the list and display only the drives of the selected class.

You can find information on the capacity allocation of each drive class in the upper right corner, as shown in Figure 6-24:

- **MDisk Capacity** shows the storage capacity of the selected drive class that is assigned to MDisks.
- **Spare Capacity** shows the storage capacity of the selected drive class that is used for spare drives.
- **Total Capacity** shows the overall capacity of the selected drive class.

If **All Internal** 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 capacity is allocated to MDisks and spare drives. MDisk capacity is represented by dark blue and spare capacity by light blue.

To list all available in the system internal drives, use CLI command **`lsdrive`**. If needed, you can filter output to list only drives that belong to particular enclosure, only that have specific capacity, or by another attributes. For an example, see Example 6-11.

Example 6-11 lsdrive output (some lines and columns are not shown)

```
IBM_Storwize:ITSOV7K:superuser>lsdrive
id status error_sequence_number use       tech_type       capacity mdisk_id
0  online                       member    tier_enterprise 278.9GB  0
1  online                       member    tier_enterprise 278.9GB  0
<...>
7  online                       member    tier_enterprise 558.4GB  16
8  online                       member    tier_enterprise 558.4GB  16
<...>
14 online                       spare     tier_enterprise 558.4GB
15 online                       candidate tier_enterprise 558.4GB
```

```
16 online                     candidate tier_enterprise 558.4GB
<...>
```

Drive list shows **Status** of each drive. It can be **Online, which** means that drive can communicate with disk enclosure where it is installed. Drive status **Offline** indicates that there is no connection with it. Its status might be failed or physically removed.

Drive **Use** defines its current role. The following use roles are available:

- **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 is a safe state for newly added hardware.
- **Candidate**: The drive is owned by the system, and is not currently part of the RAID configuration. It is available to be used in an array MDisk.
- **Spare**: The drive is a hot spare. It is protecting RAID arrays. If any of the traditional array members fails, it becomes array *Member*.
- **Member**: The drive is part of a (T)RAID or DRAID array.
- **Failed**: The drive is not a part of the array configuration and is waiting for a service action.

The use that can be assigned depends on the current drive use. These dependencies are shown in Figure 6-25.

		To				
		Unused	Candidate	Failed	Member	Spare
From	Unused	allowed	allowed	no option		not allowed
	Candidate	allowed	allowed			allowed
	Failed	allowed	allowed			not allowed
	Member	No change on member drives				
	Spare	not allowed	allowed	no option		allowed

Figure 6-25   Allowed usage changes for internal drives

> **Note:** To start configuring arrays in a new system, all *Unused* drives need to be configured as *Candidates*. Initial setup or Assign storage GUI 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 6-26 on page 210. Alternatively, select the drives and click **Actions**.

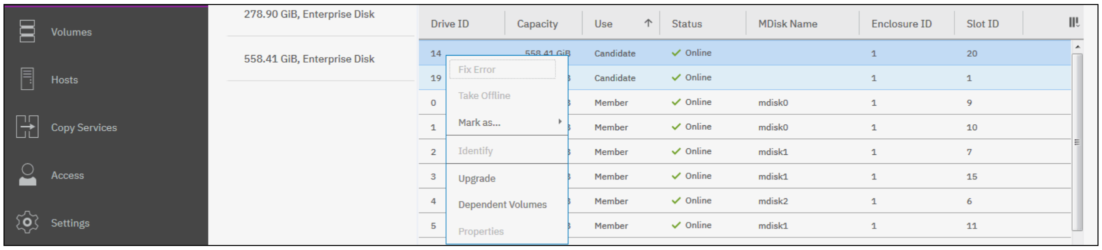


Figure 6-26 Actions on internal storage

The actions available depend on the status of the drive or drives selected. Some actions can be run only on a set of them, and some are possible only for individual drives.

Action: Fix error

This action is only available if the drive 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, “RAS, monitoring, and troubleshooting” on page 673.

Action: Take offline

Select **Take Offline** to take a drive offline. You must confirm the action, as shown in Figure 6-27.

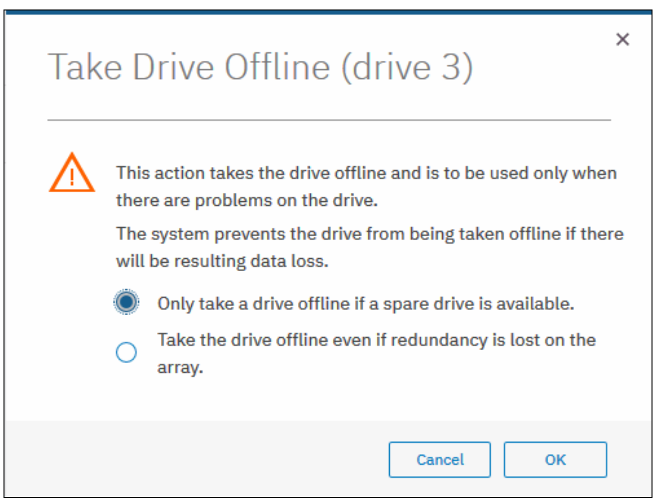


Figure 6-27 Taking a drive offline

If a spare drive is available and the drive is taken offline, the MDisk of which the failed drive is a member remains *Online*. The spare is automatically reassigned. If no spare drive is available and the drive is taken offline, the status of the array of which the failed drive is a member becomes *Degraded*. Therefore, the status of the storage pool to which the MDisk belongs becomes *Degraded* as well.

The system prevents you from taking the drive offline if one of the following conditions is true:

- The first option was selected and no suitable spares are available.
- Losing another drive in the array results in data loss.

A drive that is taken offline is considered *Failed*, as shown in Figure 6-28 on page 211.

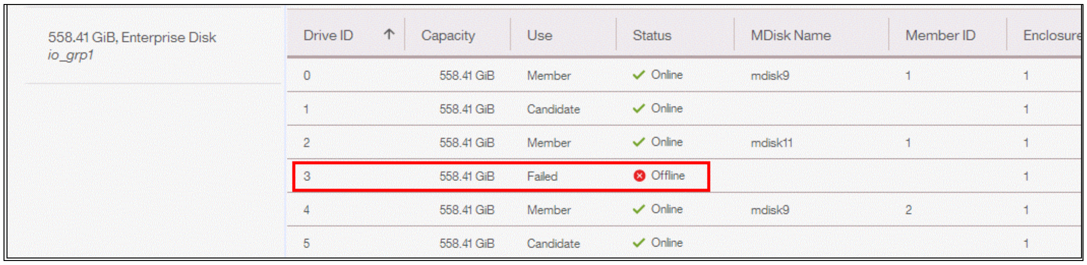


Figure 6-28 An offline drive is marked as failed

To set the drive offline with the CLI, use **chdrive**. See Example 6-12. This command returns no feedback.

Example 6-12 Setting drive offline with CLI

```
IBM_Storwize:ITSOV7K::superuser>chdrive -use failed 3
IBM_Storwize:ITSOV7K::superuser>
```

Action: Mark as

Select **Mark as** to change the usage that is assigned to the drive, as shown in Figure 6-29. A list of available use options depends on current drive state and allowed state changes, as was shown in Figure 6-25 on page 209.

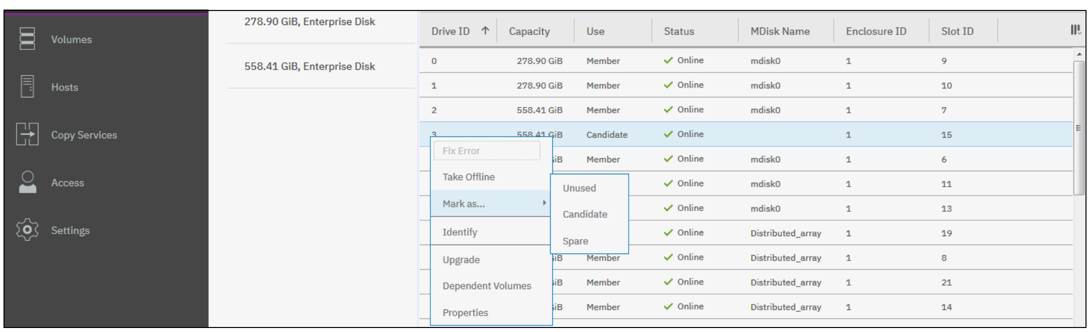


Figure 6-29 A drive can be marked as Unused, Candidate, or Spare

To change the drive role with the CLI, use **chdrive**. See Example 6-13. It shows the drive that was set offline with a previous command is set to spare. Notice that it can't go from *Failed* to *Spare* use in one step. It needs to be assigned to a *Candidate* role before.

Example 6-13 Changing drive role with CLI

```
IBM_Storwize:ITSOV7K:superuser>lsdrive -filtervalue status=offline
id status  error_sequence_number use    tech_type       capacity mdisk_id
3  offline                       failed tier_enterprise 558.4GB
IBM_Storwize:ITSOV7K: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_Storwize:ITSOV7K:superuser>chdrive -use candidate 3
IBM_Storwize:ITSOV7K:superuser>chdrive -use spare 3
IBM_Storwize:ITSOV7K:superuser>
```

Action: Identify

Select **Identify** to turn on the LED light that identifies a drive that must be replaced or that you want to troubleshoot. Select this action to access a dialog box like the one shown in Figure 6-30.

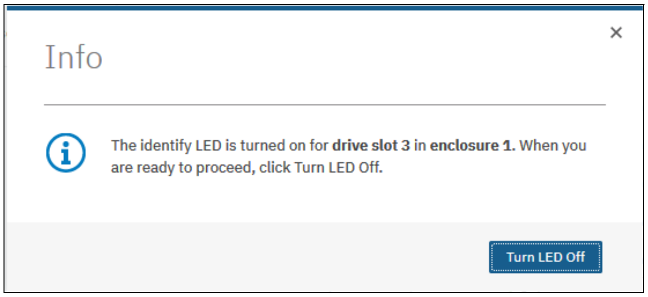


Figure 6-30 Identifying an internal drive

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.

In the CLI, this action requires a command that operates with the drive enclosure. The LED does not belong to a drive itself, but to the slot of the enclosure, so the command is **`chenclosureslot`**. See Example 6-14 for commands to turn on and off identification LED on slot 3 of enclosure 1, populated with drive id 21.

Example 6-14 Changing slot LED to identification mode with CLI

```
IBM_Storwize:ITSOV7K:superuser>lsenclosureslot -slot 3 1
enclosure_id 1
slot_id 3
fault_LED off
powered yes
drive_present yes
drive_id 21
IBM_Storwize:ITSOV7K:superuser>chenclosureslot -identify yes -slot 3 1
IBM_Storwize:ITSOV7K:superuser>lsenclosureslot -slot 3 1
enclosure_id 1
slot_id 3
fault_LED slow_flashing
powered yes
drive_present yes
drive_id 21
IBM_Storwize:ITSOV7K:superuser>chenclosureslot -identify no -slot 3 1
```

Action: Upgrade

Select **Upgrade** to update the drive firmware as shown in Figure 6-31 on page 213. You can choose to update an individual drive, selected drives or all the drives in the system.

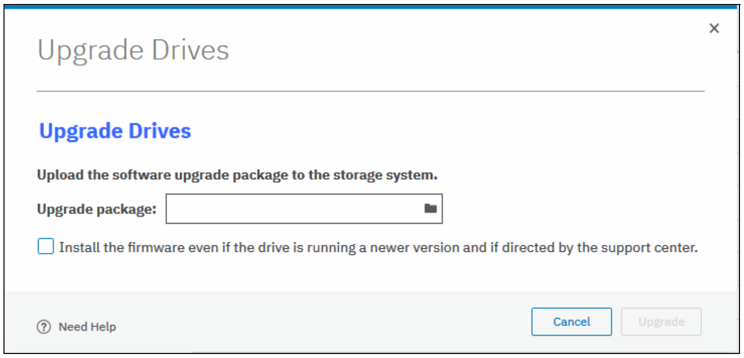


Figure 6-31 Upgrading a drive or a set of drives

For information about updating drive firmware, see Chapter 13, “RAS, monitoring, and troubleshooting” on page 673.

Action: Show dependent volumes

Select **Show Dependent Volumes** to list the volumes that are dependent on the selected drive. A volume is dependent on a drive or a set of drives when removal or failure of that drive or set of drives causes the volume to become unavailable. Use this option before you do maintenance operations, to confirm which volumes will be affected.

Figure 6-32 shows the list of volumes dependent on a set of three drives that belong to the same MDisk. This configuration means that all listed volumes will go offline if all selected drives go offline. If only one drive goes offline, there is no volume dependency.

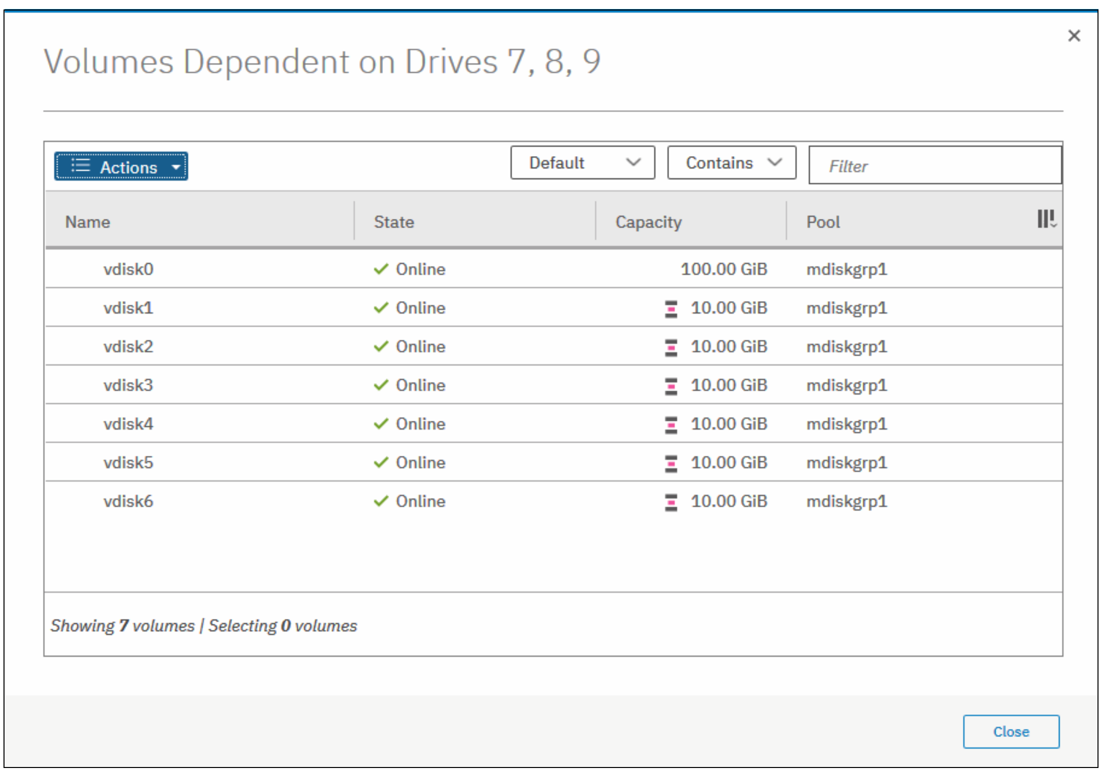


Figure 6-32 List of volumes dependent on disks 7, 8, 9

Note: A lack of dependent volumes does not imply that there are no volumes that use the drive. Volume dependency actually shows the list of volumes that would become unavailable if the drive or the group of selected drive becomes unavailable.

You can get the same information with the CLI command `lsdependentvdisks`. Use the parameter **`-drive`** with a list of drive IDs that you are checking, separated with a colon (**`:`**).

Action: Properties

Select **Properties** to view more information about the drive, as shown in Figure 6-33.

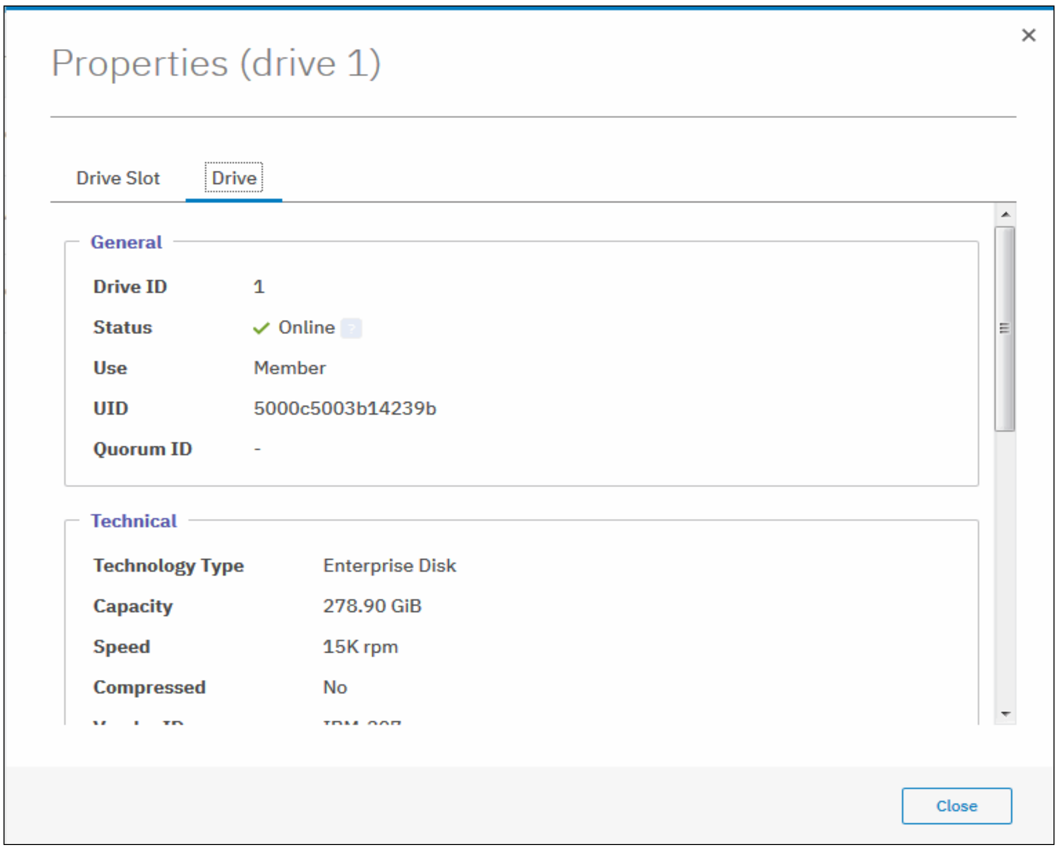


Figure 6-33 Drive properties

You can find a short description of each drive property by hovering on it and clicking **[?]**. You can also display drive slot details by changing to the **Drive Slot** tab.

To get all available information about the particular drive, use the CLI command `lsdrive` with drive ID as the parameter. To get slot information, see the `lsenclosureslot` command.

6.2.2 RAID and DRAID

To use internal Storwize V7000 disks in storage pools, you must join them into RAID arrays to form array 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, high availability for the data, or a combination of both. RAID levels define a trade-off between high availability, performance, and cost.

When multiple physical disks are set up to use the RAID technology, they are in a *RAID array*. The IBM Storwize V7000 provides multiple, traditional RAID levels:

- RAID 0
- RAID 1
- RAID 5
- RAID 6
- RAID 10

In a traditional RAID approach, whether it is RAID10, RAID5, or RAID6, data is spread among drives in an array. However, the spare space is constituted by spare drives, which are global and sit outside of the array. When one of the drives within the array fails the system responds as follows:

- For RAID10, all data is read from the mirrored copy.
- For RAID5 or RAID6, data is calculated from remaining data stripes and parity.

Then, the data is written to a spare drive.

In a disk failure, traditional RAID writes the data to a single spare drive. With increasing capacity, the rebuild time is also increased. And the probability of a second failure during the rebuild process also becomes more likely. In addition, the spares, when they are not being used, are idle, which wastes resources.

Distributed RAID (DRAID) addresses those points.

Distributed RAID

In distributed RAID, all drives are active, which improves performance. Spare capacity is used instead of the idle spare drives from traditional RAID. Because no drives are idling, all drives contribute to performance. The spare capacity is spread across the disk drives, so the write rebuild load is distributed across multiple drives and the bottleneck of one drive is removed.

Distributed RAID reduces the recovery time and the probability of a second failure during rebuild. Like traditional RAID, a distributed RAID 5 array can lose one physical drive and survive. If another drive fails in the same array before the bad drive is recovered, the MDisk and the storage pool go offline as they are supposed to. So, distributed RAID does not change the general RAID behavior.

It is available for the IBM Storwize V7000 in two types:

- Distributed RAID 5 (DRAID 5)
- Distributed RAID 6 (DRAID 6)

Figure 6-34 on page 216 shows an example of a Distributed RAID6 with 10 disks. The physical disk drives are divided into multiple packs. The reserved spare capacity (which is marked in yellow) is equivalent to two spare drives, but the capacity is distributed across all of the physical disk drives. The data is distributed across a single row.

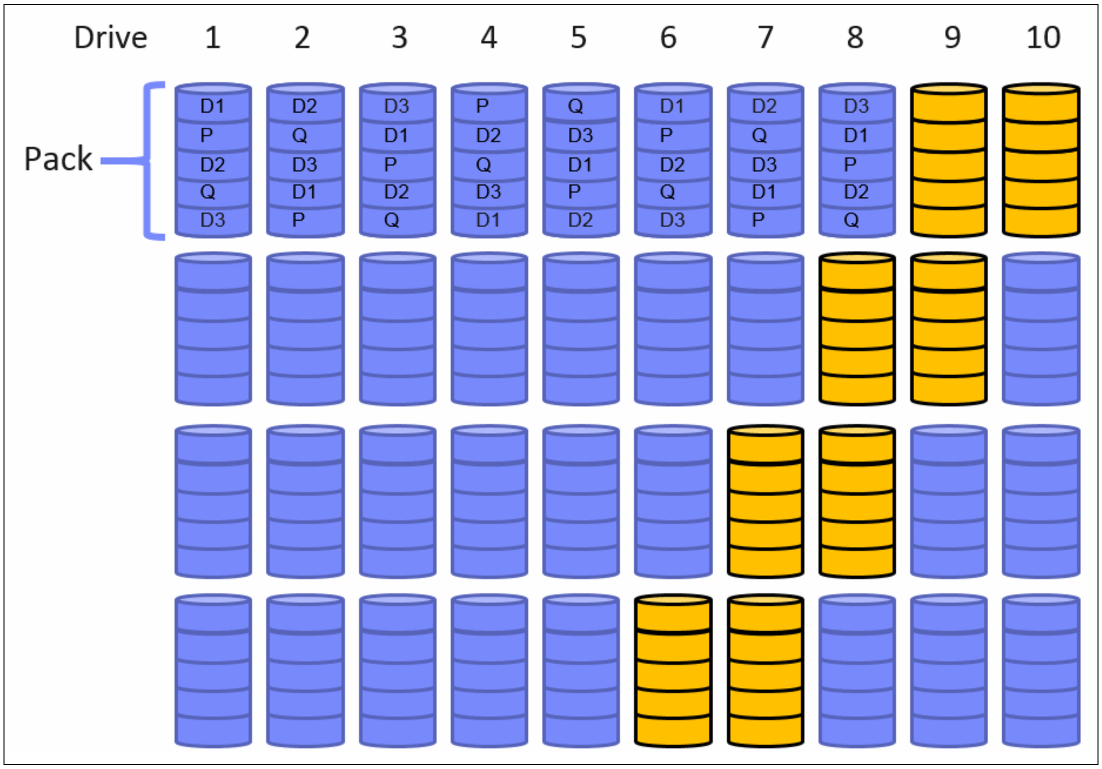


Figure 6-34 Distributed RAID 6 (for simplification, not all packs are shown)

Figure 6-35 on page 217 shows a single drive failure in this DRAID6 environment. Physical disk 3 failed and the RAID 6 algorithm is using the spare capacity for a single spare drive in each pack for rebuild (which is marked in green). All disk drives are involved in the rebuild process, which significantly reduces the rebuild time.

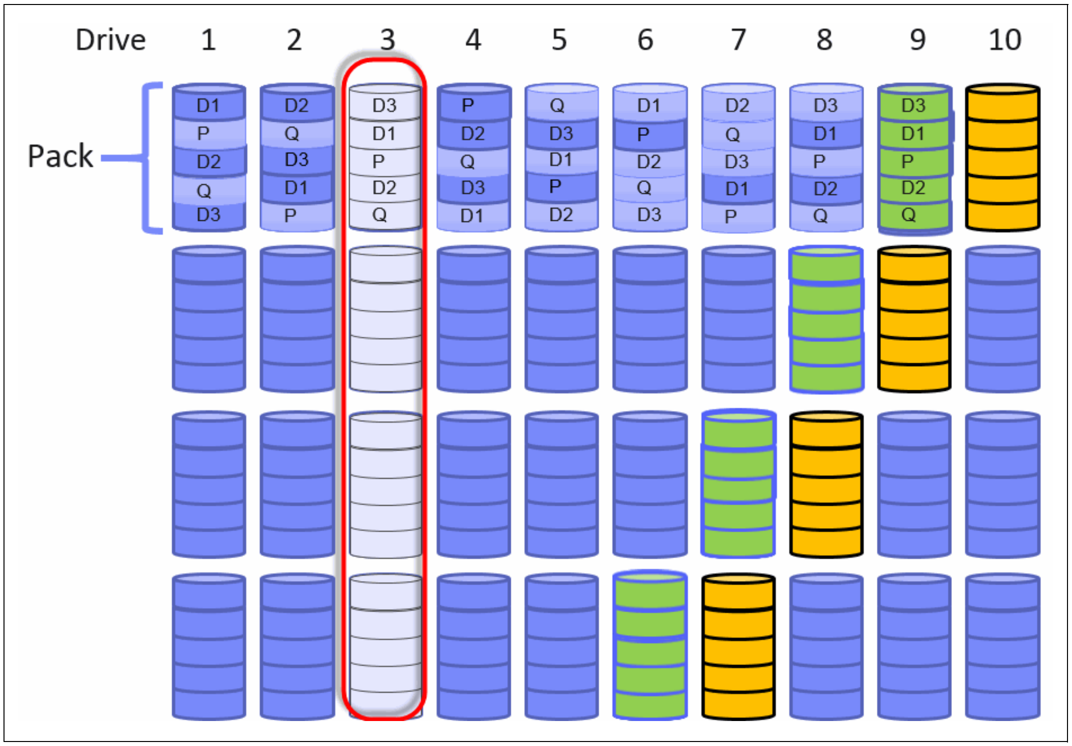


Figure 6-35 Single drive failure with DRAID 6 (for simplification, not all packs are shown)

This model addresses the main disadvantages of traditional RAID:

- In case of a drive failure, data is read from many drives and written to many drives. This process minimizes the effect on performance during the rebuild process. Also, it significantly reduces rebuild time. (Depending on the distributed array configuration and drive sizes the rebuild process can be up to 10 times faster.)
- Spare space is distributed throughout the array. So, more drives are processing I/O, and no spare drives are idling.

IBM Storwize V7000 DRAID implementation has the following additional advantages:

- In case of a drive failure, only the actual data is rebuilt. Space that is not allocated to volumes is not re-created to the spare regions of the array.
- Arrays can be much larger than before and can span many more drives, therefore improving the performance of the array. The maximum number of drives a Distributed RAID can contain is 128.

The following are the minimum number of drives that are needed to build a Distributed Array:

- Six drives for a Distributed RAID6 array
- Four drives for a Distributed RAID5 array

6.2.3 Creating arrays

Only MDisks (RAID arrays of disks) can be added to pools. It is not possible to add JBOD or a single drive. Then, when assigning storage to a pool, the system first needs to create one or many MDisks. You cannot create MDisks out of internal drives without assigning them to a pool.

> **Note**: Although Traditional RAID is still supported and might be suggested as the default choice in the GUI, it is recommended to use Distributed RAID 6 whenever possible. DRAID technology dramatically reduces rebuild times and decreases the exposure volumes have to the extra load of recovering redundancy.

To create and assign array type MDisk, right-click target storage pool and select **Add Storage**, as shown in Figure 6-36.

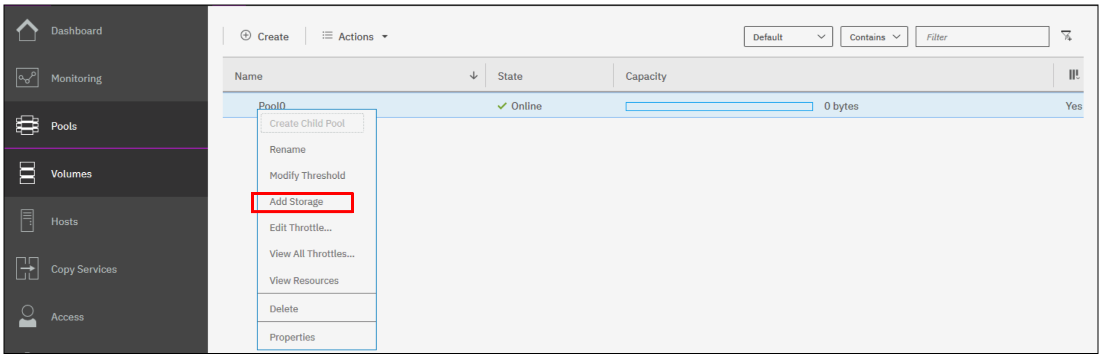


Figure 6-36 Adding storage to a pool

This action starts the configuration wizard that is shown in Figure 6-37. If any of the drives are found with *Unused* role, it is suggested that you reconfigure them as *Candidates* to be included into configuration.

If **Internal** or **Internal Custom** is chosen, then the system guides you through array MDisk creation. If **External** is selected, the system guides you through the selection of external storage.

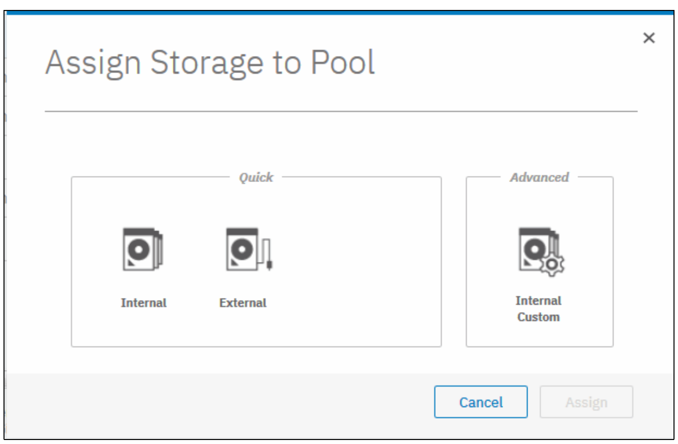


Figure 6-37 Assigning storage to a pool

Add Storage dialog provides the following two options to configure internal arrays:

- **Automatic**, which relies on the system to choose the RAID level automatically for the given set of drives
- **Custom**, which provides more flexibility and lets you choose RAID parameters manually

Automatic internal configuration

By selecting **Internal**, you rely on automatically chosen parameters such as RAID type (traditional or distributed), stripe width, number of spares (for traditional RAID), number of rebuild areas (for distributed RAID), and number of drives of each class. The number of drives is the only value that can be adjusted when the array is created. Depending on the number and the type of drives that are selected for the new array, the RAID level automatically adjusts. Also, the system automatically assigns spare disks.

> **Note:** It is not possible to change RAID level or drive count of the existing array. If you need to change these properties, you must delete array MDisk and re-create it with the required settings.

In the example given in Figure 6-38, the dialog box suggests that you to assign 9 drives out of the given 10 to RAID6 array, the remaining one is assigned as a hot-spare.

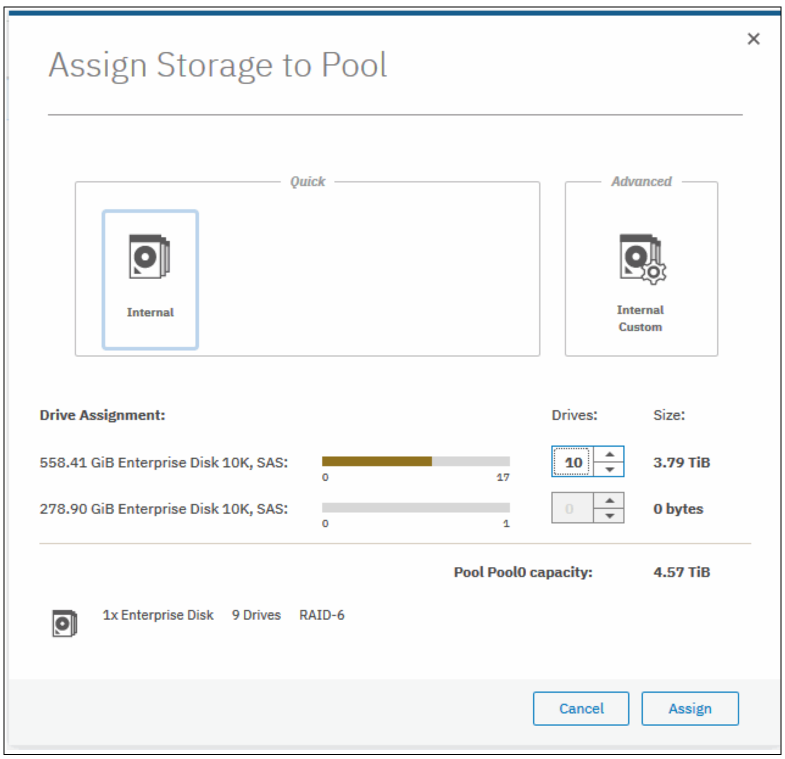


Figure 6-38 Assign Storage to Pool dialog

If you select two drives only, the system automatically creates a RAID-10 array, with no spare drive. For more control of the array creation steps, you can select the **Internal Custom** option.

By default, if sufficient candidate drives are present, the system recommends traditional arrays for most new configurations of MDisks. However, switch to Distributed RAID when possible, with the **Internal Custom** option.

If the system has multiple drive classes (like Flash and Enterprise disks, for example), the default option is as follows:

1. Create multiple arrays of different tiers.
2. Assign them to the pool to take advantage of the Easy Tier functionality.

However, you can adjust this configuration by setting the number of drives of different classes to zero. For information about Easy Tier, see Chapter 10, “Advanced features for storage efficiency” on page 403.

If you are adding storage to a pool with storage already assigned, the existing storage is taken into consideration. Some properties are inherited from existing arrays for a specific drive class. Systems aim to achieve an optimally balanced configuration, so it is not possible to add significantly different MDisks to one pool with the given GUI dialog.

For example, if the pool has an array MDisk made of 16 drives of DRAID6, you cannot add two drives of RAID1 to the same pool. Otherwise, an imbalanced storage pool is created.

You can still add any array of any configuration to an existing pool by using the CLI.

When you are satisfied with the configuration presented, click **Assign**. The RAID arrays, or MDisks, are then created and initialized in the background. You can monitor the progress of the initialization process by selecting the corresponding task under **Running Tasks** in the upper-right corner of GUI screen, as shown in Figure 6-39. The array is available for I/O during this process.

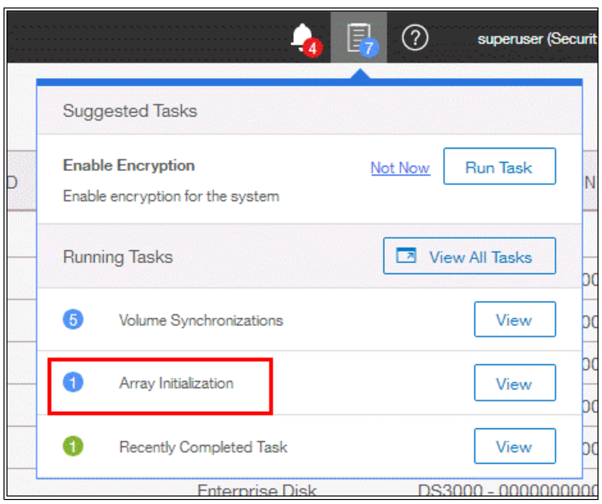


Figure 6-39   Array Initialization task

Click **View** in the Running tasks list to see the initialization progress and the time that remains, as shown in Figure 6-40. Notice that array creation depends on the type of drives it consists of. For example, an array of Flash drives is much quicker to initialize than NL-SAS drives.

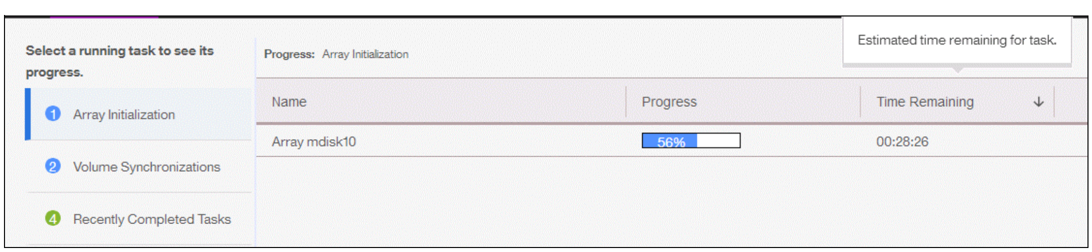


Figure 6-40   Array initialization task progress information

Custom configuration

Select **Internal Custom** to customize the configuration of MDisks made out of internal drives. The following values can be customized:

- RAID level
- Number of spares (or spare areas)
- Array width
- Stripe width
- Number of drives of each class

Figure 6-41 shows an example with nine drives ready to be configured as DRAID 6, with the equivalent of one drive capacity of spare (distributed over the nine disks).

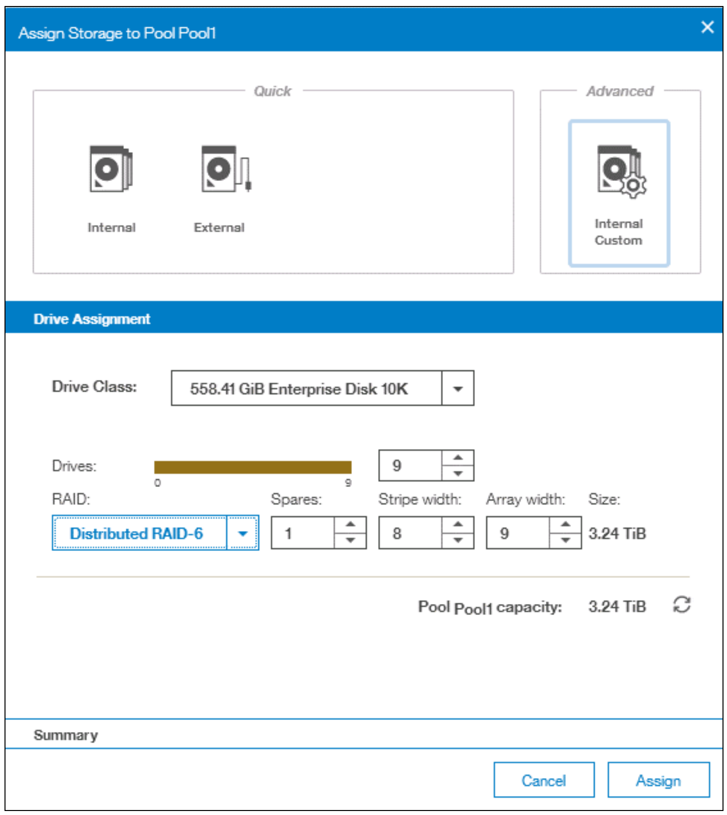


Figure 6-41 Adding internal storage to a pool by using the Advanced option

To return to the default settings, click the **Refresh** button next to the pool capacity. To create and assign the arrays, click **Assign**.

Unlike automatic assignment, custom internal configuration does not create multiple arrays. You must create each array separately.

As with automatic assignment, the system does not allow you to add significantly different arrays to an existing and populated pool. The goal is to achieve balance across MDisks inside one pool. You can still add any array of any configuration to an existing pool by using the CLI.

Note: When Internal Custom configuration is performed, spare drives are not assigned when traditional RAID arrays are created. You need to set them up manually.

Configuring arrays with the CLI

When you work with the CLI, use **mkarray** to create traditional RAID and **mkdistributedarray** to create distributed RAID. For this, first retrieve a list of drives that are ready to become array members. See 6.2.1, "Working with drives" on page 207, to find information on how to list all available drives, read their use roles, and change those use roles.

To create a TRAID MDisk, you must specify a list of drives that you want to add as members, its RAID level, and the storage pool name or ID to which you want to add this array. Example 6-15 creates RAID-6 array out of seven drives with IDs 0 to 6, and adds it to Pool0.

Example 6-15 Creating TRAID with mkarray

```
IBM_Storwize:ITSOV7K:superuser>mkarray -level raid6 -drive 0:1:2:3:4:5:6 Pool0
MDisk, id [0], successfully created
```

Notice that the storage pool should already exist. See the creation instructions in 6.1.1, "Creating storage pools" on page 193. Also, the required amount of spare drives must be assigned manually.

To create a DRAID MDisk, specify the RAID level, number of drives, drive class, and target storage pool. Drive class depends on drive technology type, connectivity, speed, and capacity. The system automatically chooses drives for the array from the available drives in the class. The number of rebuild areas is also set automatically, but you can adjust it. Example 6-16 creates DRAID6 array out of seven drives of class 1 and adds it to Pool1.

Example 6-16 Creating DRAID with mkdistributedarray (some columns are not shown)

```
IBM_Storwize:ITSOV7K:superuser>lsdriveclass
id RPM   capacity tech_type       block_size candidate_count superior_count
0  15000 278.9GB tier_enterprise 512        3               3
1  10000 558.4GB tier_enterprise 512        11              11
2  10000 278.9GB tier_enterprise 512        1               22
IBM_Storwize:ITSOV7K:superuser>mkdistributedarray -level raid6 -driveclass 1
-drivecount 7 -stripewidth 6 Pool1
MDisk, id [0], successfully created
```

Notice that in this example it was required to specify the stripe width, because for DRAID6 it is 12 by default. The drive count value must equal or be greater than the combined value of the stripe width and rebuild areas count.

To check array initialization progress with the CLI, use **lsarrayinitprogress**.

6.2.4 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 traditional RAID arrays are not supported on distributed RAID arrays, and vice versa.

To choose an action, open **Pools** → **MDisks by Pools,** select the array (MDisk), and click **Actions**. Alternatively, right-click the array, as shown in Figure 6-42 on page 223.

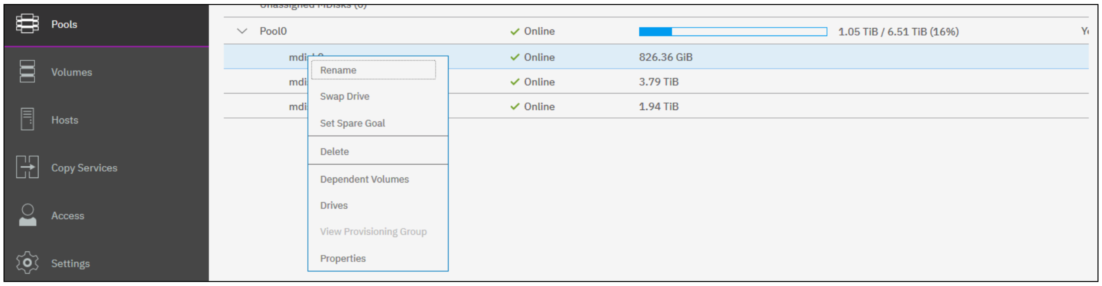


Figure 6-42 Actions on arrays

Rename

Select this option change MDisk object name.

The CLI command for this operation is **charray**. See Example 6-17. No feedback is returned.

Example 6-17 Renaming array MDisk with charray

```
IBM_Storwize:ITSOV7K:superuser>charray -name Distributed_array mdisk1
IBM_Storwize:ITSOV7K:superuser>
```

Swap drive

Select **Swap Drive** to replace a drive in the array with another drive. The other drive needs to have use of `Candidate` or `Spare`. Use this action to perform proactive drive replacement, to replace a drive that is not failed but is expected to fail soon, for example, as indicated by an error message in the event log.

Figure 6-43 shows the dialog box that is displayed. Select the member drive that you want to replace and the replacement drive, and click **Swap**.

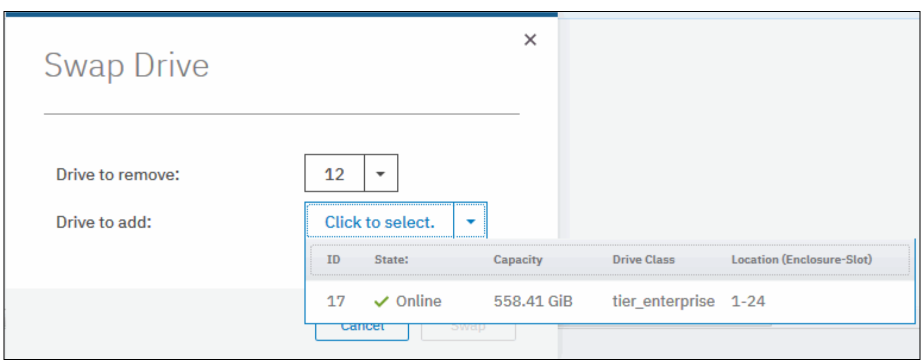


Figure 6-43 Swapping array member with another candidate or spare drive

The exchange of the drives begins to run in the background. The volumes on the affected MDisk remain accessible during the process.

You can use the CLI command **charraymember** to do this task. Example 6-18 shows replacement of array member id 7, that was assigned to drive id 12, with drive id 17.

Example 6-18 Replacing array member with CLI (some columns are not shown)

```
IBM_Storwize:ITSOV7K:superuser>lsarraymember 16
mdisk_id mdisk_name        member_id drive_id new_drive_id spare_protection
16       Distributed_array 6         18                     1
16       Distributed_array 7         12                     1
```

```
16       Distributed_array 8        15                    1
<...>
IBM_Storwize:ITSOV7K: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_Storwize:ITSOV7K:superuser>charraymember -immediate -member 7 -newdrive 17
Distributed_array
IBM_Storwize:ITSOV7K:superuser>
```

Set spare goal or Set Rebuild Areas Goal

Select this option to set the number of spare drives (on TRAID) or rebuild areas (on DRAID) that are required to protect the array from drive failures.

If the number of rebuild areas available does not meet the configured goal, an error is logged in the event log as shown in Figure 6-44. This error can be fixed by replacing failed drives in the DRAID array.

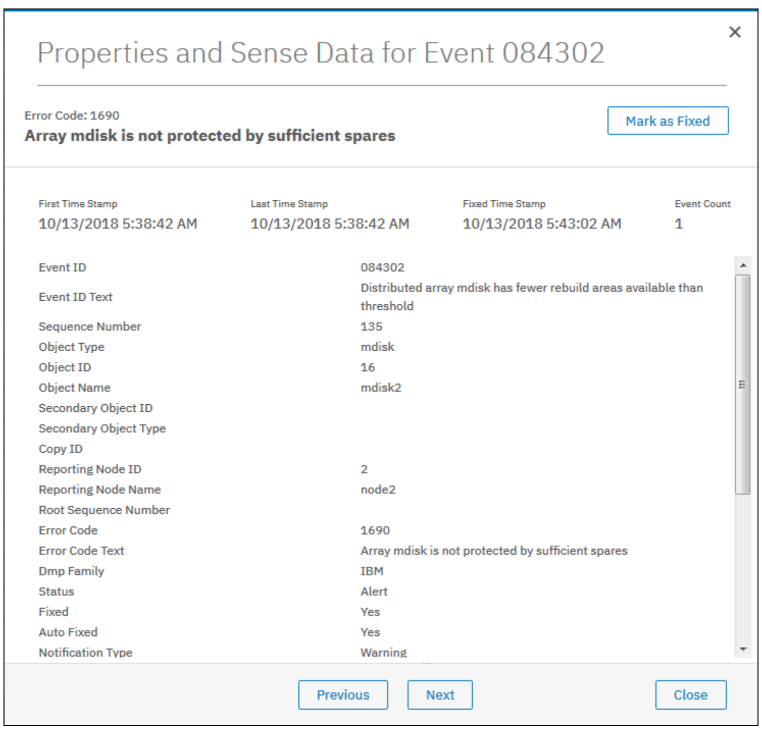


Figure 6-44 If there are insufficient rebuild areas available, an error 1690 is logged

Note: This option does not change the internal DRAID configuration. It specifies only the level at which a warning event is generated. Setting the goal to 0 does not prevent array from rebuilding to a spare drive.

In the CLI, this task is performed with **charray** (see Example 6-19).

Example 6-19 Adjusting array goals with charray (some columns are not shown)

```
IBM_Storwize:ITSOV7K: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_Storwize:ITSOV7K:superuser>charray -sparegoal 2 mdisk0
IBM_Storwize:ITSOV7K:superuser>charray -rebuildareasgoal 0 Distributed_array
```

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 return into `Candidate`.

If there are no volumes that use extents from this array, the command runs immediately without additional confirmation. If there are volumes that use extents from this array, you are prompted to confirm the action, as shown in Figure 6-45.

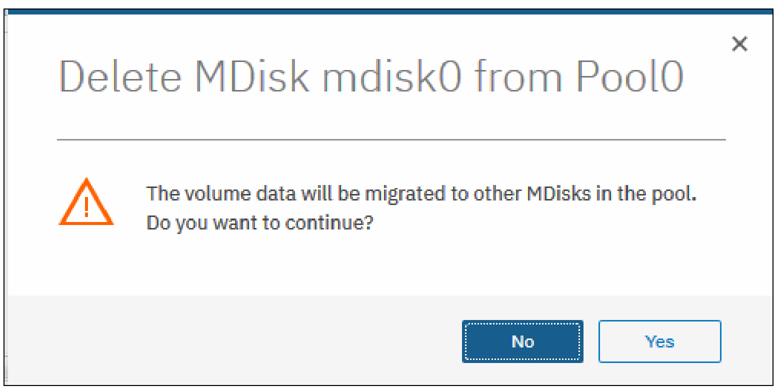


Figure 6-45 Deleting an array from a non-empty storage pool

Your confirmation of the action triggers the migration of the volumes to extents from other MDisks that remain in the pool. After the action is completed, the array is removed from the storage pool and deleted.

> **Note:** The command fails if you don't have enough available capacity remaining in the storage pool to allocate the data that is being migrated from the removed array.

To delete the array with the CLI, use **rmarray**. The **-force** parameter is required if volume extents need to be migrated to other MDisks in a storage pool.

Dependent Volumes

Volumes are entities that are made of extents from a storage pool. The extents of the storage pool come from various MDisks. A volume can then be spread over multiple MDisks, and MDisks can serve multiple volumes. Clicking the Dependent Volumes **Action** menu of an MDisk lists the volumes that are depending on that MDisk as shown in Figure 6-46 on page 226.

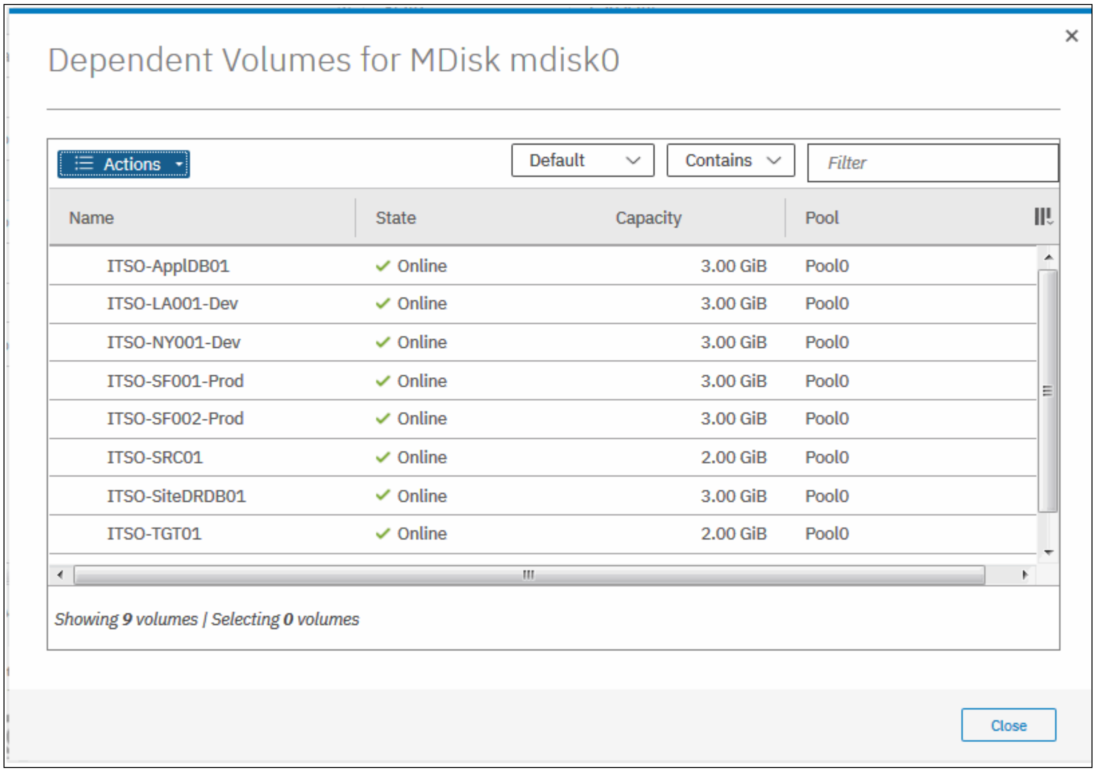


Figure 6-46 Dependent volumes for MDisk mdisk0

You can get the same information with the CLI command **lsdependentvdisks** (see Example 6-20).

Example 6-20 Listing vdisks dependent on MDisk with CLI

```
IBM_Storwize:ITSOV7K:superuser>>lsdependentvdisks -mdisk mdisk0
vdisk_id vdisk_name
0        ITSO-SRC01
1        ITSO-TGT01
2        ITSO-ApplDB01
<...>
```

Drives

Select **Drives** to see information about the drives that are included in the array as shown in Figure 6-47 on page 227.

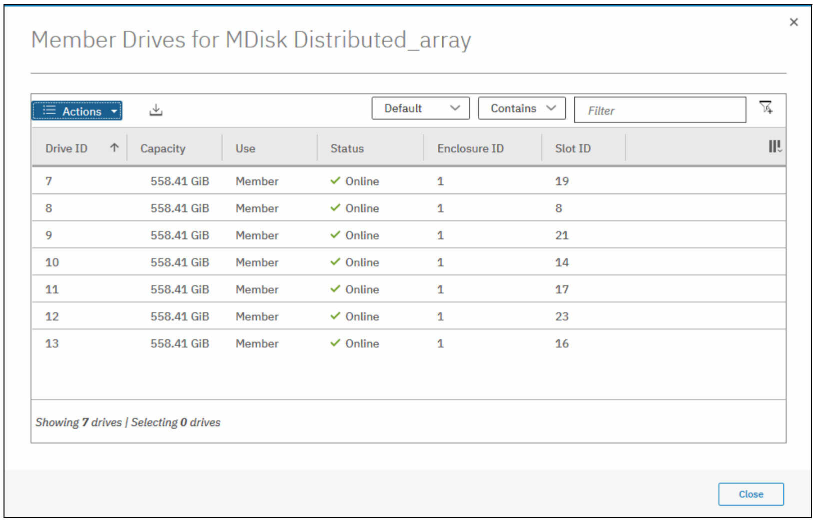


Figure 6-47 List of drives in an array

You can use the CLI command **lsarraymember** to get the same information with the CLI. Provide an array name or ID as the parameter to filter output by the array. If run without arguments, the command lists all members of all configured arrays.

Properties

This section shows all available array MDisk parameters: its state, capacity, RAID level, and others.

Use the CLI command **lsarray** to get a list of all configured arrays. Use **lsarray** with array name or ID as the parameter to get extended information about the selected one, as shown in Example 6-21.

Example 6-21 lsarray output (truncated)

```
IBM_Storwize:ITSOV7K:superuser>lsarray
mdisk_id mdisk_name        status mdisk_grp_id mdisk_grp_name capacity
0        mdisk0            online 0            mdiskgrp0      1.3TB
16       Distributed_array online 1            mdiskgrp1      2.2TB
IBM_Storwize:ITSOV7K:superuser>lsarray 16
mdisk_id 16
mdisk_name Distributed_array
status online
mode array
mdisk_grp_id 1
mdisk_grp_name mdiskgrp1
capacity 2.2TB
<...>
```

6.3 Working with external controllers and MDisks

In IBM Spectrum Virtualize terminology, *Controllers* are external storage systems that provide resources to be used as MDisks. Storwize V7000 supports external storage controllers that are attached through iSCSI and through Fibre Channel.

6.3.1 External storage controllers

You can manage both FC and iSCSI storage controllers through the External Storage pane. To access the External Storage pane, browse to **Pools** → **External Storage**, as shown in Figure 6-48.

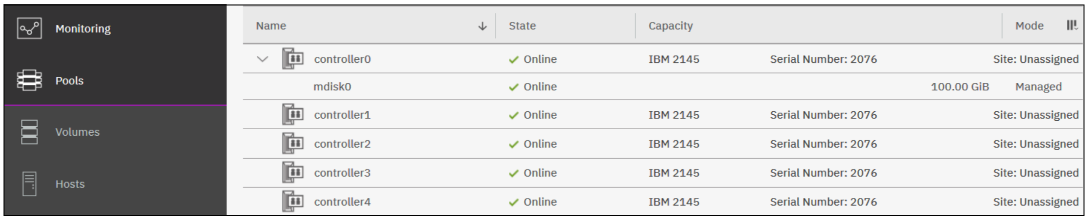


Figure 6-48 External Storage pane

> **Note:** A controller that is connected through Fibre Channel 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 **External Storage** does not appear in **Pools**, virtualization licenses are not configured. To use Storwize V7000 virtualization functions, you must order the correct External Virtualization licenses. You can configure them by clicking **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** pane 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 UID’s, or external storage system WWNNs and LU number, to identify each device.

To list all visible external storage controllers with CLI, use `lscontroller`, as shown in Example 6-22.

Example 6-22 Listing controllers with the CLI (some columns are not shown)

```
IBM_Storwize:ITSOV7K: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
<...>
```

System layers

System layers must be configured if the external controller is a Storwize system. By default, IBM Storwize V7000 is set to *storage layer*. With it, it cannot virtualize another IBM Storwize system, configured to the same layer. To do that, the system must be set to *replication layer*.

Then, the system is able to see external Storwize systems as storage controllers and can virtualize LUs that are provided by them.

You also get a warning that your Storwize is in a *storage layer* if you try to add an iSCSI storage controller with the GUI. The system prompts you to convert the system to *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 family system.
> - No system partnerships can be defined.
> - No Storwize family system can be visible on the SAN fabric.

To switch IBM Storwize V7000 system layer, you can also use **chsystem** CLI command, as shown in Example 6-23. If executed successfully, it returns no output.

Example 6-23 Changing system layer

```
IBM_Storwize:ITSOV7K:superuser>lssystem |grep layer
layer storage
IBM_Storwize:ITSOV7K:superuser>chsystem -layer replication
IBM_Storwize:ITSOV7K:superuser>
```

For more information about layers and how to change them, see IBM Knowledge Center.

Fibre Channel external storage controllers

A controller connected through Fibre Channel is detected automatically by the system, provided that the cabling, the zoning, and the system layer are configured correctly.

If the external controller is not detected, ensure that the Storwize V7000 is cabled and zoned into the same storage area network (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 family.

After the problem is corrected, use **Actions** → **Discover Storage** button found in **Pools** → **External Storage**, as shown in Figure 6-49, to rescan the Fibre Channel network.

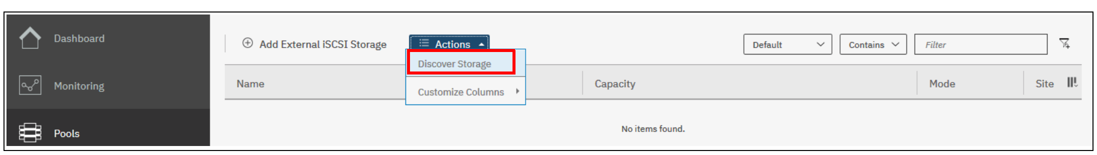


Figure 6-49 Discover storage menu

This action executes CLI command **detectmdisk**. It returns no output. Although it might appear that the command has completed, some extra time might be required for it to run. **detectmdisk** is asynchronous and returns a prompt while the command continues to run in the background.

iSCSI external storage controllers

Unlike Fibre Channel connections, you must manually configure iSCSI connections between the Storwize V7000 and the external storage controller. Until you do this, the controller is not listed in the External Storage pane.

To start virtualizing the iSCSI backend controller, you need to follow the documentation in `IBM Knowledge Center` to perform the configuration steps that are specific to your backend storage controller.

Detailed instructions on configuring Storwize V7000 to virtualize back-end storage controller with iSCSI are available in the following publication: *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327.

6.3.2 Actions on external storage controllers

You can perform a number of actions on external storage controllers. Some actions are available for external iSCSI controllers only.

To select any action, right-click the controller, as shown in Figure 6-50. Alternatively, select the controller and click **Actions**.

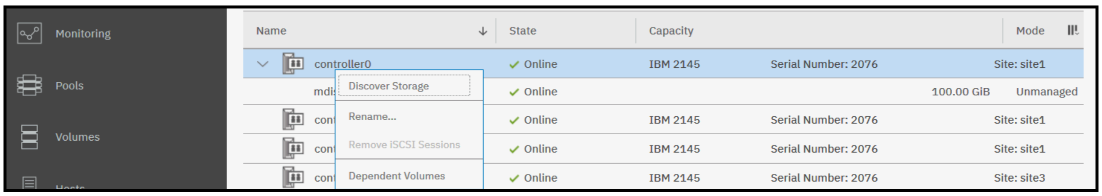


Figure 6-50   Actions on external storage

Discover storage

When you create or remove LUs on an external storage system, the change might not be detected automatically. In this case, select **Discover Storage** for the system to rescan the Fibre Channel 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 executes the CLI command **`detectmdisk`**.

Rename

Select **Rename** to modify the name of an external controller to simplify administration tasks. Naming rules are the same as for storage pools and can be found in 6.1.1, “Creating storage pools” on page 193.

To rename storage controller with the CLI, use **`chcontroller`**. Use case is provided in Example 6-24 on page 231.

Remove iSCSI sessions

This action is available only for external controllers that are attached with iSCSI. Right-click the session and select **Remove** to remove the iSCSI session that is established between the source and target port.

For CLI commands and detailed instructions, see *iSCSI Implementation and Best Practices on IBM Storwize Storage Systems*, SG24-8327.

Modify site

This action is available only for systems that use HyperSwap. Select **Modify Site** to change the site with which the external controller is associated, as shown in Figure 6-51.

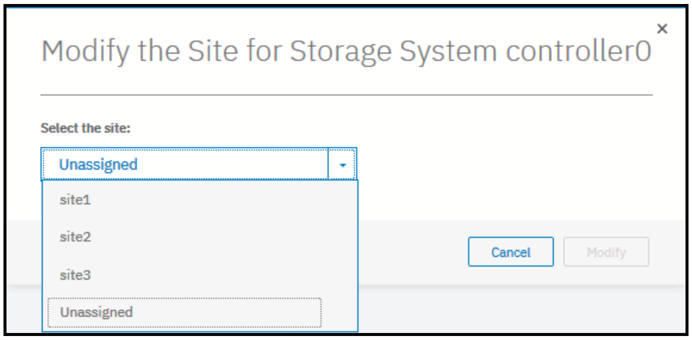


Figure 6-51 Modifying the site of an external controller

To change controller site assignment with CLI, use **chcontroller**. Example 6-24 shows controller0 renamed and reassigned to a different site.

Example 6-24 Changing controller's name and site

```
IBM_Storwize:ITSOV7K:superuser>chcontroller -name site3_controller -site site3 controller0
IBM_Storwize:ITSOV7K:superuser>
```

6.3.3 Working with external MDisks

After an external back-end system is configured, attached to IBM Storwize V7000, and detected as a controller, you can work with LUs that are provisioned from it.

External LUs can have one of the following modes on Storwize V7000:

- *Unmanaged*

  External MDisks are 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 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.

- *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 translation from the MDisk to a volume. This mode is provided to satisfy the following major usage scenarios:

  – Virtualization of external LUs that contain data that was not written through the Storwize V7000.

  – Exporting MDisks from the Storwize V7000 after migration of volumes to image mode MDisks.

Listing external MDisks

External MDisks can be managed through the External Storage pane, which is accessed by clicking **Pools** → **External Storage**, as shown in Figure 6-52.

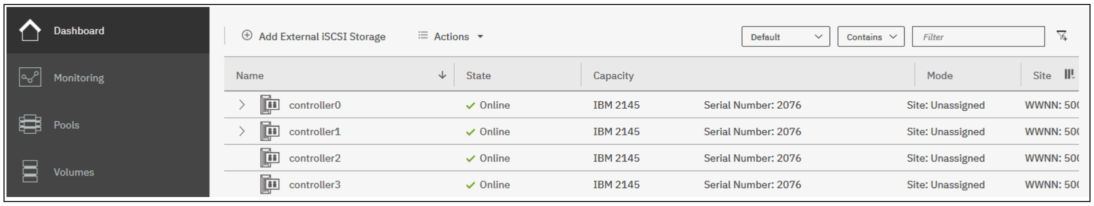


Figure 6-52 External storage view

To list all MDisks that are visible by the system with the CLI, use 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

Add the MDisks to an existing pool or create a new pool to include them. If the storage pool does not exist, follow the procedure that is outlined in 6.1.1, "Creating storage pools" on page 193.

Figure 6-53 shows how to add selected MDisk to an existing storage pool. Click **Assign** under the Actions menu or right-click on the MDisk and select **Assign**.

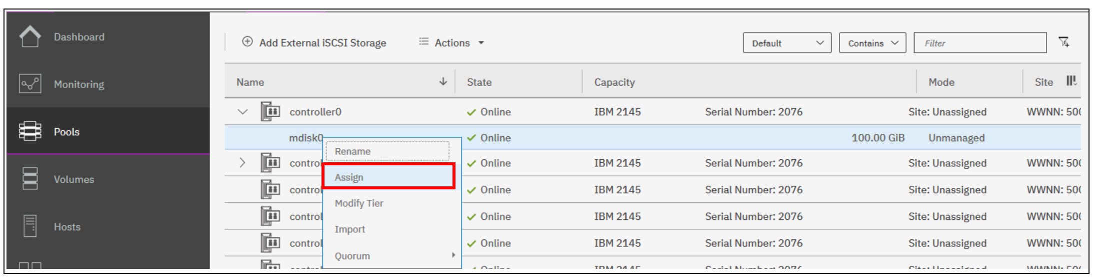


Figure 6-53 Assigning unmanaged MDisk

After you click **Assign**, a dialog window opens, as shown in Figure 6-54. Select the target pool, MDisk storage tier, and external encryption setting.

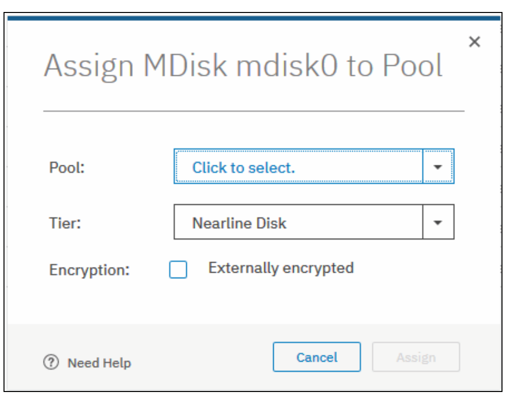


Figure 6-54 Assign MDisk dialog

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. Otherwise, the wrong tier can corrupt Easy Tier algorithm logic and might affect system performance. For more information about storage tiers, see Chapter 10, “Advanced features for storage efficiency” on page 403.

The storage tier setting can be also changed after the MDisk is assigned to the pool.

Check the **Externally encrypted** checkbox if back-end storage currently performs data encryption. For details on IBM Storwize V7000 encryption, see Chapter 12, “Encryption” on page 603.

After the task is completed, click **Close.**

> **Note:** If the external storage LUs to virtualize behind the Storwize V7000 contain data and this data needs to be retained, do not use the “Assign to pool” option to manage the MDisks. This option destroys the data on the LU. Instead, use the Import option. For more information, see Chapter 9, “Storage migration” on page 385.

The external MDisks that are assigned to a pool within IBM Storwize V7000 are displayed in **Pools** → **MDisks by Pools**.

When a new MDisk is added to the pool that already has other MDisks and volumes, the system does not immediately redistribute extents that are used by volumes between new and existing MDisks. This task is performed by the Easy Tier feature. It automatically balances volume extents between the MDisks to provide the best performance to the volumes.

The system decision to turn on Easy Tier for pool movement of extents—between tiers of storage (inter-tier) or between MDisks within a single tier (intra-tier)—is based on the activity that is monitored. Therefore, when you add an MDisk to a pool, extent migration is not performed immediately. No migration of extents occurs until there is sufficient activity to trigger it.

If Easy Tier is turned off, no extent migration is performed. Only newly allocated extents can be written to a new MDisk.

For more information about IBM Easy Tier feature, see Chapter 10, “Advanced features for storage efficiency” on page 403.

To assign an external MDisk to a storage pool with CLI, use the `addmdisk` command. You need to specify the MDisk name or ID, MDisk tier, and target storage pool, as shown in Example 6-25. The command returns no feedback.

Example 6-25 addmdisk example

```
IBM_Storwize:ITSOV7K:superuser>addmdisk -mdisk mdisk3 -tier enterprise Pool0
IBM_Storwize:ITSOV7K:superuser>
```

6.3.4 Actions on external MDisks

External MDisks support specific actions that are not supported on arrays. Some actions are supported only on unmanaged external MDisks, and some are supported only on managed external MDisks.

To choose an action, select the external MDisk in **Pools** → **External Storage** or **Pools** → **MDisks by Pools,** and click **Actions**, as shown in Figure 6-55. Alternatively, right-click the external MDisk.

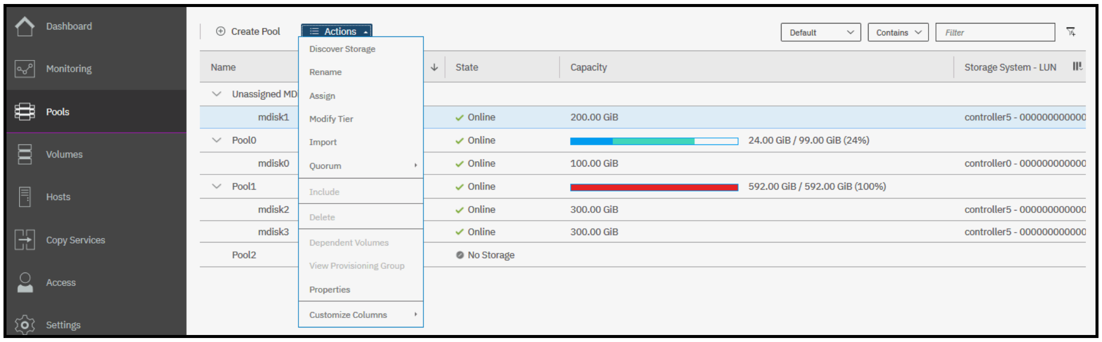


Figure 6-55   Actions on MDisks

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 managed disks (MDisks) that might have been added.
- Rebalance MDisk access across all available controller device ports.

This action executes the CLI command **`detectmdisk`**.

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 232.

Modify tier

Select **Modify Tier** to modify the tier to which the external MDisk is assigned, as shown in Figure 6-56. This setting is adjustable because the system cannot always detect the tiers that are associated with external storage automatically, unlike with internal arrays.

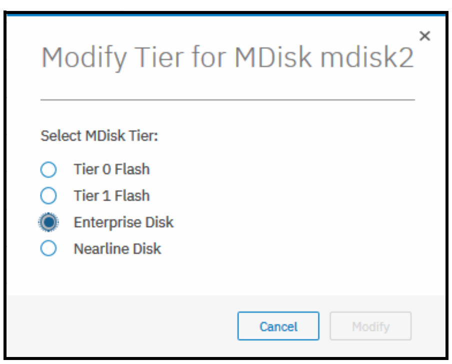


Figure 6-56   Modifying an external MDisk tier

For information about storage tiers and their importance, see Chapter 10, “Advanced features for storage efficiency” on page 403.

To change the external MDisk storage tier, use **chmdisk**. Example 6-26 shows setting the new tier to the mdisk2. No feedback is returned.

Example 6-26 Changing tier setting with CLI

```
IBM_Storwize:ITSOV7K:superuser>chmdisk -tier tier1_flash mdisk2
IBM_Storwize:ITSOV7K:superuser>
```

Modify encryption

Select **Modify Encryption** to modify the encryption setting for the MDisk. 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 information about encryption, encrypted storage pools, and self-encrypting MDisks, see Chapter 12, “Encryption” on page 603.

To perform this task with CLI, use **chmdisk**. See Example 6-27.

Example 6-27 Using chmdisk to modify encryption

```
IBM_Storwize:ITSOV7K:superuser>chmdisk -encrypt yes mdisk5
IBM_Storwize:ITSOV7K:superuser>
```

Import

This action is available only for unmanaged MDisks. Importing an unmanaged MDisk allows the user to preserve the 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. IBM Storwize V7000 provides a migration wizard to help with this process. It is discussed in Chapter 9, “Storage migration” on page 385.

> **Note:** The wizard is the preferred method to migrate data from legacy storage to the Storwize V7000. When an MDisk presents as 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.

Select **Import** to choose one of the following migration methods:

- **Import to temporary pool as image-mode volume** does not migrate data from the source MDisk. It creates an *image-mode volume* that has a direct block-for-block translation of the MDisk. The existing data is preserved on the external storage system, but it is also accessible from the Storwize V7000 system.

  In this method, the image-mode volume is created in a temporary migration pool and is presented through the Storwize V7000. Choose the extent size of the temporary pool and click **Import**, as shown in Figure 6-57 on page 236.

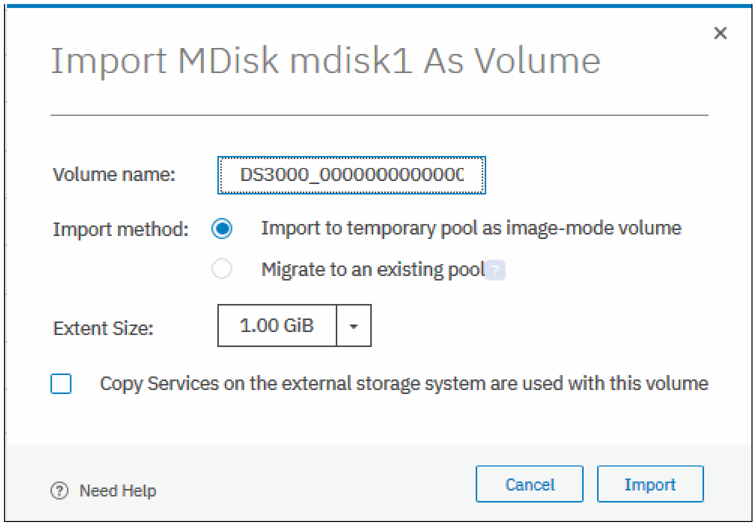


Figure 6-57 Importing an unmanaged MDisk

The MDisk is imported and listed as an `image` mode MDisk in the temporary migration pool, as shown in Figure 6-58.

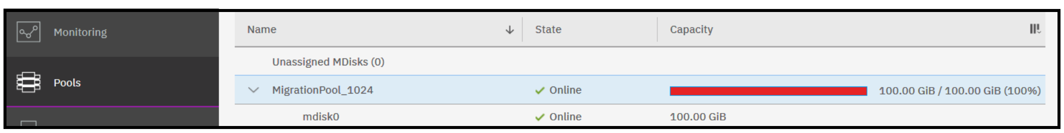


Figure 6-58 Image-mode imported MDisk

A corresponding image-mode volume is now available in the same migration pool, as shown in Figure 6-59.

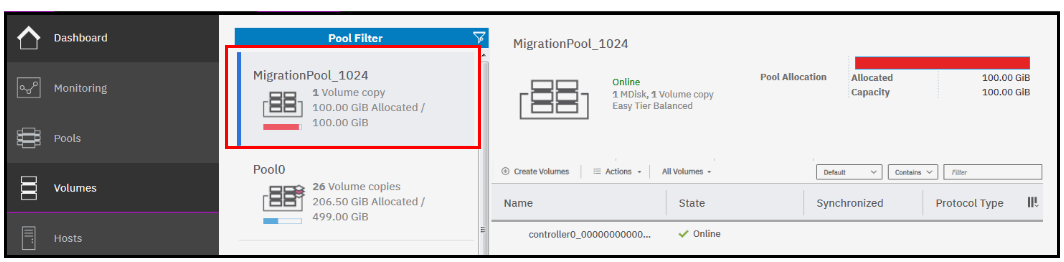


Figure 6-59 Image-mode Volume

The image-mode volume can then be mapped to the original host. The data is still physically present on the physical disk of the original external storage controller system and no automatic migration process is currently running. The original host sees no difference and the applications can continue to run. The image-mode Volume can now be handled by Spectrum Virtualize. If needed, it can be migrated to another storage pool later. You can use the **Volume Migration** wizard or manually. That action converts the image-mode volume to virtualized and fully managed by the system.

- **Migrate to an existing pool** starts by creating an image-mode volume, like the first method. However, it then automatically migrates the data from the image-mode volume onto the virtualized volume in the selected storage pool. After the migration process is completed, the image-mode volume and temporary migration pool are deleted. Free extents must be available in the selected pool so that 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 6-60.

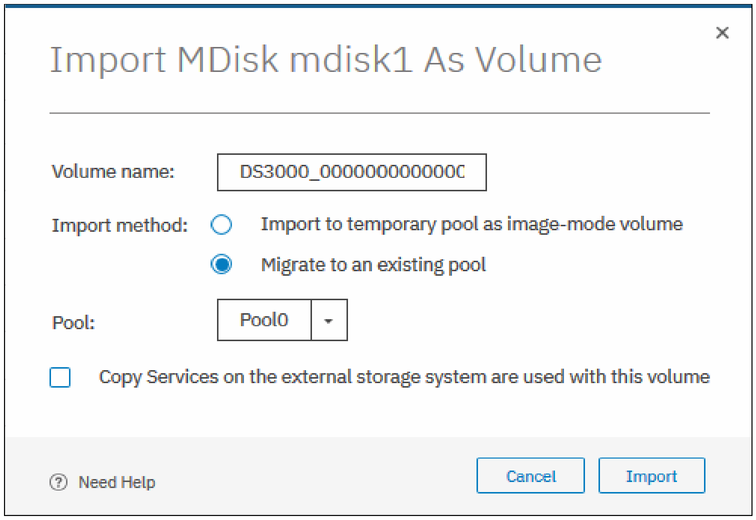


Figure 6-60 Migrating an MDisk to an existing pool

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 6-61.

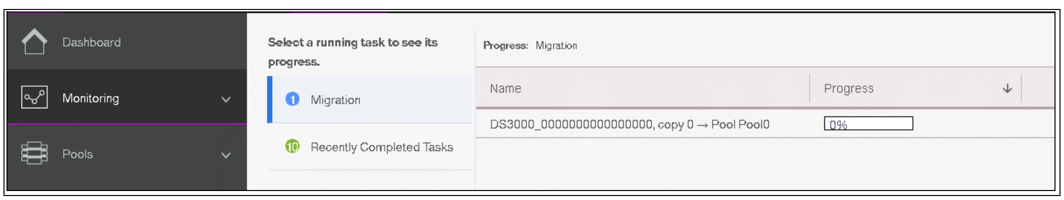


Figure 6-61 MDisk migration in the tasks pane

After the migration is completed, 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.

At this point, all data has been migrated off the source MDisk, and the MDisk has switched its mode, as shown in Figure 6-62.

The MDisk can be removed from the migration pool. It returns in the list of external MDisks and can be used as a regular MDisk to host volumes, or the legacy storage device can be decommissioned.

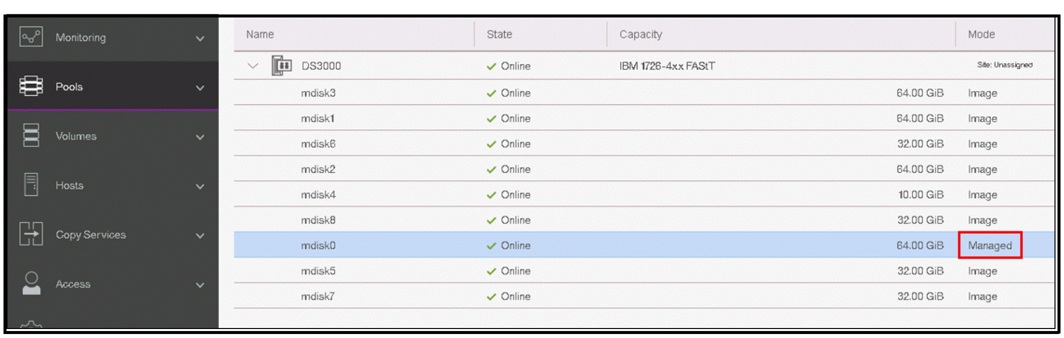


Figure 6-62 Imported MDisks appear as "Managed"

Alternatively, import and migration of external MDisks to another pool can be done by selecting **Pools** → **System Migration** to start system migration wizard. For more information, see Chapter 9, “Storage migration” on page 385.

Include

The system can exclude an MDisk that has multiple I/O failures or has persistent connection errors from its storage pool. Exclusion ensures that these errors do not interfere with data access. If an MDisk has been automatically excluded, run the fix procedures 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 has been fixed, select **Include** to add the excluded MDisk back into the storage pool.

The CLI command `includemdisk` performs same task. It needs the MDisk name or ID to be provided as a parameter. See Example 6-28.

Example 6-28 Including degraded MDisk with CLI

```
IBM_Storwize:ITSOV7K:superuser>includemdisk mdisk3
IBM_Storwize:ITSOV7K:superuser>
```

Remove

In some cases, you might want to remove external MDisks from storage pools to reorganize your storage allocation. Select **Remove** to remove the MDisk from the storage pool. After the MDisk is removed, it goes back to `unmanaged`. If there are no volumes in the storage pool to which this MDisk is allocated, the command runs immediately without additional confirmation. If there are volumes in the pool, you are prompted to confirm the action, as shown in Figure 6-63.

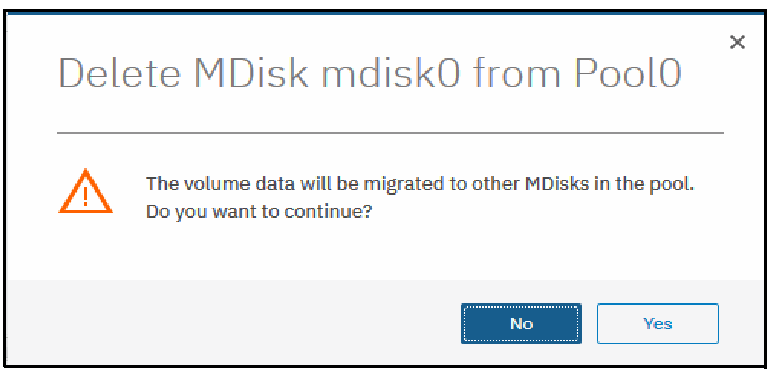


Figure 6-63 Removing an MDisk from a pool

Confirming the action starts the migration of the volumes to extents from other MDisks that remain in the pool. When the action is completed, the MDisk is removed from the storage pool and returns to `Unmanaged`.

Ensure that you have enough available capacity remaining in the storage pool to allocate the data being migrated from the removed MDisk. Otherwise, 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 with the CLI, use `rmmdisk`. You must use the `-force` parameter if you need to migrate volume extents to other MDisks in a storage pool.

The command fails if you don't have enough available capacity remaining in the storage pool to allocate the data that you are migrating from the removed array.

Dependent Volumes

Volumes are entities that are made of extents from a storage pool. The extents of the storage pool come from various MDisks. A volume can then be spread over multiple MDisks, and MDisks can serve multiple volumes. Clicking the **Dependent Volumes** menu item of an MDisk lists volumes that use extents that are stored on this particular MDisk.

You can get the same information with the CLI command `lsdependentvdisks`.

7

Volumes

In IBM Spectrum Virtualize, a *volume* is an amount of storage space that is provisioned out of a storage pool and presented to a host as a 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 available, and available volume customization options.

The second part of this chapter describes how to create volumes by using the GUI and shows you how to map these volumes to hosts.

The third part of this chapter provides an introduction to volume manipulation from the command line interface (CLI).

This chapter includes the following topics:

7.1 Introduction to volumes

A volume is a logical disk that the system presents to attached hosts. 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 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 made out of extents that are allocated from a storage pool. Storage pools group managed disks (MDisks), which are Redundant Arrays of Independent Disks (RAIDs) from internal storage, or logical units that are presented to and virtualized by IBM Spectrum Virtualize system. Each MDisk is divided into sequentially numbered extents (0 based indexing). The extent size is a property of a storage pool, and is used for all MDisks comprising the storage pool.

MDisks are not directly visible to or used by host systems.

A volume is presented to hosts by an I/O Group, and within that group has a preferred node; that is, a node that by default serves I/O requests to that volume.

IBM Spectrum Virtualize uses 512 bytes block size for volumes that are presented to hosts.

7.1.1 Operations on volumes

The following operations can be performed on a volume:

- Volumes can be created and deleted.
- Volumes can be resized (expanded or shrunk).
- Volumes can be migrated at run time to another MDisk or storage pool.
- Volumes can be mirrored, or mirrored volumes can be split.
- Point-in-time volume can be created by using FlashCopy. Multiple snapshots and quick restore from snapshots (reverse FlashCopy) are supported.

Note: With V7.4 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 section 7.8.9, “Volume delete protection” on page 303 and the “Enabling volume protection” topic in IBM Knowledge Center.

7.1.2 Volume characteristics

Volumes have the following characteristics or attributes:

- Volumes can be configured as managed or in image mode.
- Volumes can have extents allocated in striped or sequential mode.
- Volumes can be created as fully allocated or thin-provisioned. A conversion from a fully allocated to a thin-provisioned volume and vice versa can be done at run time.
- Volumes can be configured to have single data copy, or be mirrored to make them resistant to disk subsystem failures or to improve the read performance.

- Volumes can be compressed.
- Volumes can be configured as VMware vSphere Virtual Volumes, sometimes referred to as *VVols*, which allows VMware vCenter to manage storage objects. The storage system administrator can create these objects and assign ownership to VMware administrators to allow them to manage these objects.

  Volumes have two major modes: Managed mode and image mode. For managed mode volumes, the sequential policy and the striped policy define how the extents of the volume are allocated from the storage pool.
- Volumes can be replicated synchronously or asynchronously. An IBM Spectrum Virtualize system can maintain active volume mirrors with a maximum of three other IBM Spectrum Virtualize systems, but not from the same source volume.

These volume characteristics are described next.

7.1.3 I/O operations data flow

Volumes can have one or two copies. 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.

When host requests an I/O operation to any IBM Spectrum Virtualize volume type, the multipath driver on the host identifies the preferred node for the volume and by default uses only paths to this node to communicate I/O requests.

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 is available and synchronized, reads from the volume are directed to that copy. The user selects which copy is primary when the volume is created, but can change this setting 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; therefore, no choice exists for read source, and all reads are directed to the single volume copy.

Write I/O operations data flow

For write I/O operations to a mirrored volume, the host sends the I/O request to the preferred node, which is responsible for destaging the data from cache. Figure 7-1 on page 244 shows the data flow for this scenario.

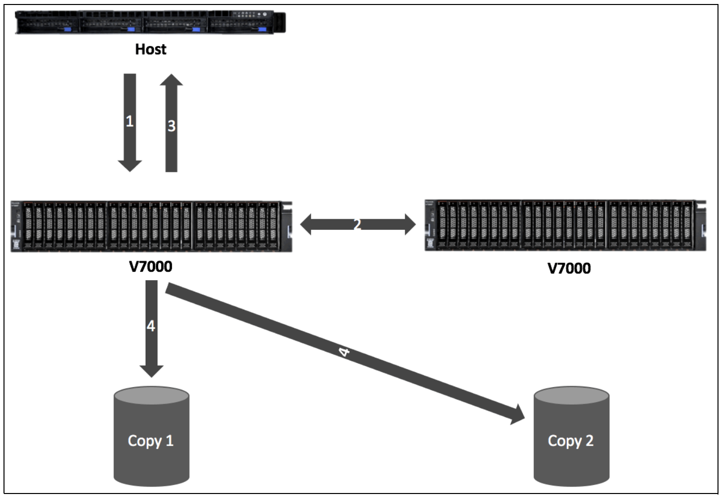


Figure 7-1   Data flow for write I/O processing in a mirrored volume

As shown in Figure 7-1, 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 7-1 shows a case with a destage to a mirrored volume; that is, one with two physical data copies.

With V7.3, the cache architecture changed from an upper-cache design to a two-layer cache design. With this change, the data is only written once, and is then directly destaged from the controller to the disk system.

Figure 7-2 shows the data flow in a stretched environment.

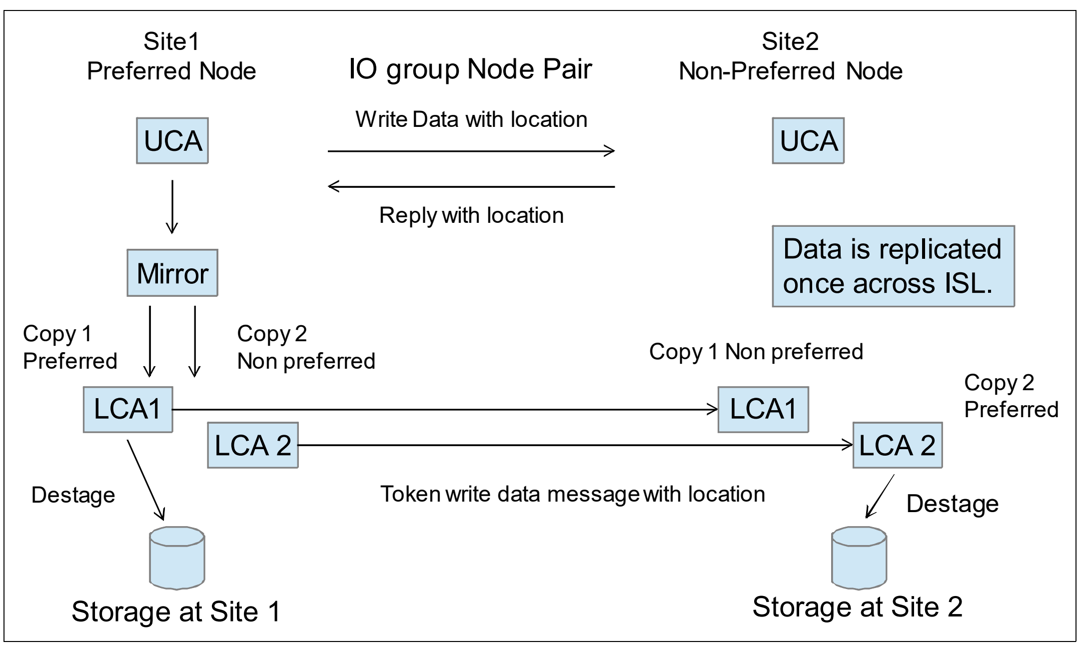


Figure 7-2   Design of an Enhanced Stretched Cluster

7.1.4 Managed mode and image mode volumes

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 7-3 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.

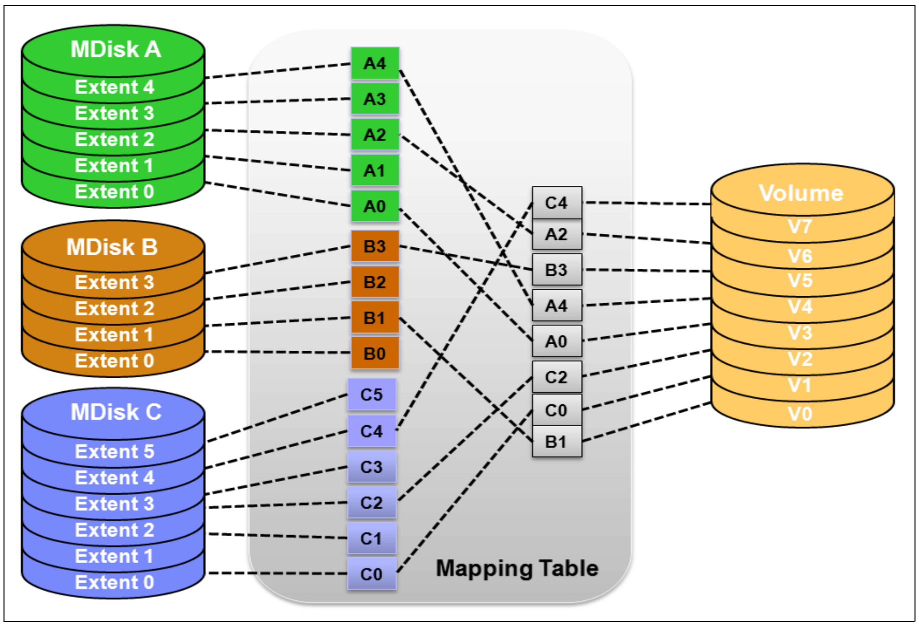


Figure 7-3 Simple view of block virtualization

Several of the MDisk extents are unused; that is, no volume extent maps to them. These unused extents are available for use in creating volumes, migration, expansion, and so on.

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 (by default, all MDisks belonging to a storage pool) and can be subject of the full set of virtualization functions.

In particular, they offer full flexibility in mapping between logical volume representation (logical blocks) and physical storage used to store these blocks. This requires that physical storage (MDisks) are fully managed by IBM Spectrum Virtualize, which means that the LUs that are presented to the IBM Spectrum Virtualize by the back-end storage systems does not contain any data when it is added to the storage pool.

Image mode volumes enable IBM Spectrum Virtualize to work with LUs that were previously directly mapped to hosts. This is often required when IBM Spectrum Virtualize is introduced into a storage environment and image mode volumes are used a method enabling seamless migration of data and smooth transition to virtualized storage.

Image mode creates one-to-one mapping of logical block addresses (LBAs) between a volume and an MDisk (LU presented by the virtualized storage). Image mode volumes have the 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; however, all of 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 7-4.

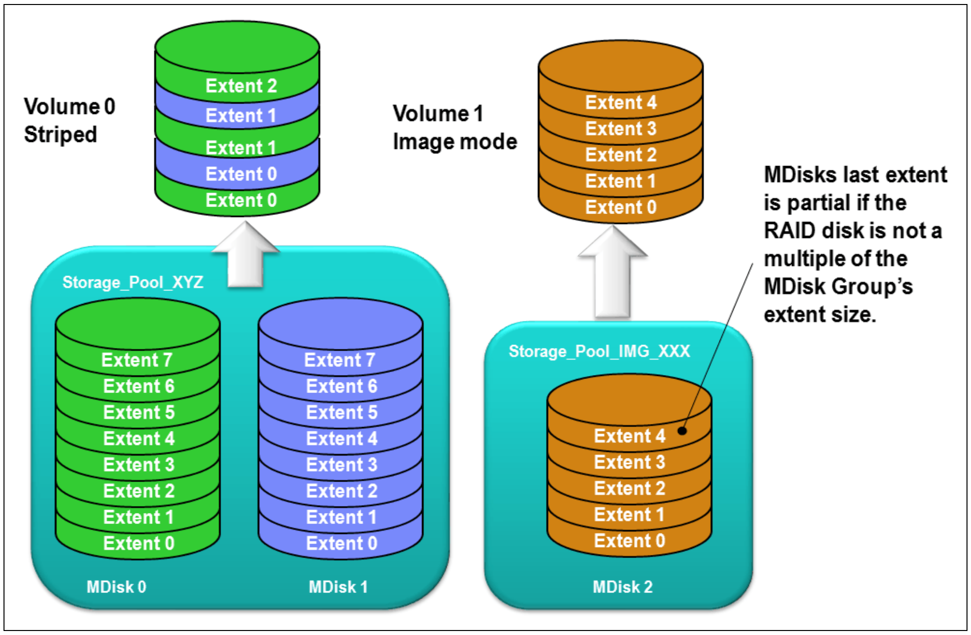


Figure 7-4 Image mode volume versus striped volume

An image mode volume is mapped to one, and only one, image mode MDisks and is mapped to the entirety of the MDisk. Therefore, the image mode volume capacity must be 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, the specified MDisk 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 (`Storage Pool_IMG_xxx`).

IBM Spectrum Virtualize also supports the reverse process, in which a managed mode volume can be migrated to an image mode volume. The extent size that is chosen for this specific storage pool must be the same as the extent size of the storage pool into which you plan to migrate the data off the image mode volumes. If a volume is migrated to another MDisk, it is represented as being in managed mode during the migration. Its mode changes to "image" only after the process completes.

It is a preferred practice to put image mode MDisks in a dedicated storage pool and use a special name for it (for example, `Storage Pool_IMG_xxx`).

7.1.5 Striped and sequential volumes

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 by using round-robin algorithm from all MDisks in the storage pool. 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 if sufficient free space is available to create a striped volume copy, select one of the following options:
>
> - Check the free space on each MDisk in the storage pool by using the **lsfreeextents** command.
> - Allow the system to automatically create the volume copy by not supplying a specific stripe set.

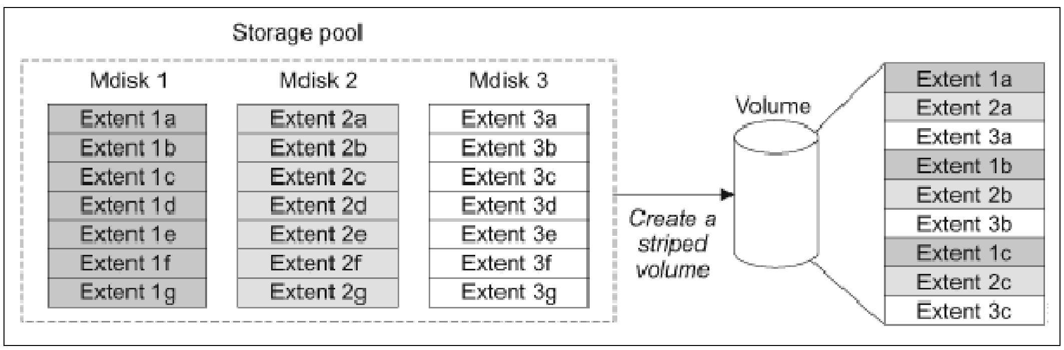


Figure 7-5 Striped extent allocation

- A *sequential* volume contains a volume copy that has extents allocated sequentially on one MDisk.
- *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/O, or that create multiple sequential streams.
- All subsequent extents (if required) are allocated from the MDisk set by using 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, many I/O to file system metadata disk regions exist. For example, for 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 successful completion of the file write.

If multiple volumes (each with its 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 first MDisk for new volume extent allocation minimizes probability, that file systems that are created on these volumes places their metadata regions on the same physical MDisk.

7.1.6 Mirrored volumes

IBM Spectrum Virtualize offers the capability to mirror volumes, which means that a single volume presented to a host can have two physical copies. Each volume copy can belong to a different pool, and each copy has the same virtual capacity as the volume.

When a server writes to a mirrored volume, the storage system writes the data to both copies. When a server reads a mirrored volume, and the volume copies are synchronized, the system reads the data from the primary copy (in the management GUI the primary volume copy is marked by an asterisk [*]).

If one of the mirrored volume copies is temporarily unavailable (for example, because the storage system that provides the pool is unavailable), the volume remains accessible to servers. The system remembers which areas of the volume were modified after loss of access to a volume copy and resynchronizes only these areas when both copies are available.

The use of mirrored volumes has the following consequences:

- Improves availability of volumes by protecting them from a single storage system failure.
- Enables concurrent maintenance of a storage system that does not natively support concurrent maintenance.
- Provides an alternative method of data migration.
- Enables conversion between fully allocated volumes and thin-provisioned volumes.
- Enables space reclamation of thin-provisioned volumes.

Note: Volume mirroring is not a true 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 two copies of a mirrored volume typically are allocated from separate storage pools backed by different physical hardware. 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 copy of a volume.

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; that is, all of its copies are expanded or shrunk when the volume is resized. The volume can participate in FlashCopy and remote copy relationships, is serviced by an I/O Group, and includes a preferred node.

An overview of volume mirroring is shown in Figure 7-6.

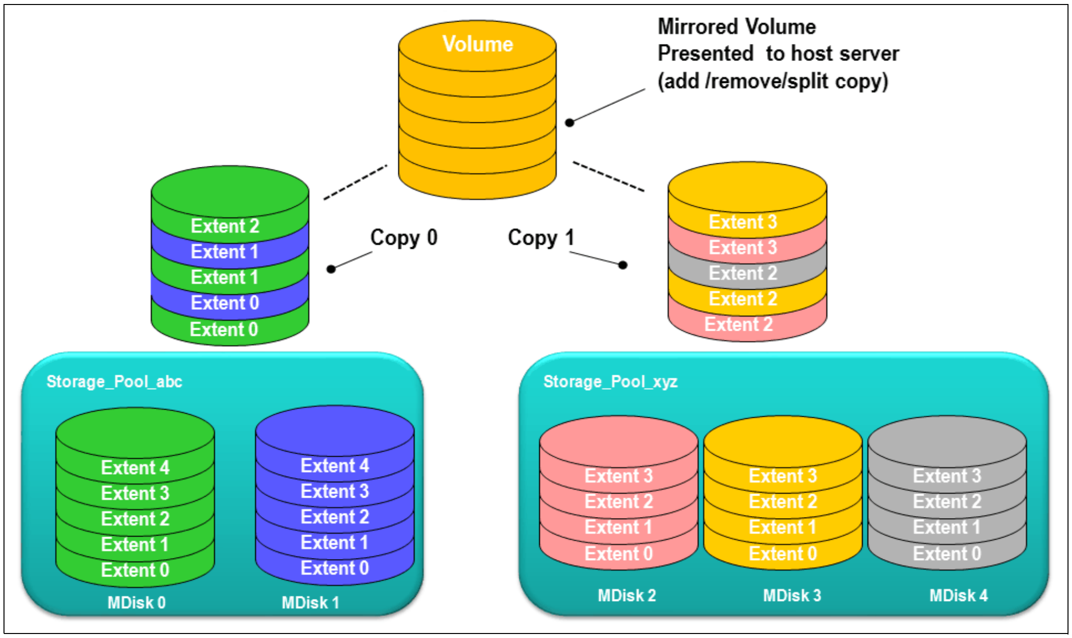


Figure 7-6 Volume mirroring overview

A second copy can be added to a volume with a single copy or removed from a volume with two copies. Checks that are in place 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. This update is done at the *synchronization rate* that is defined when the volume is created (this volume parameter can be modified after volume creation). The synchronization status for mirrored volumes is recorded on the 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 selected that declares the copies as `synchronized` even when they are not.

If the mirrored volume needs resynchronization, the system copies to the out-of-sync volume copy only these 256 kibibyte (KiB) grains that were written to (changed) since the synchronization was lost. This approach is known as an *incremental synchronization* and it minimizes the time required to synchronize the volume copies.

Note: An unmirrored volume can be migrated from one location to another by adding a second volume copy at the required destination, waiting for the two copies to synchronize, and then removing the original copy. This operation can be stopped at any time. The two copies can be in different storage pools with different extent sizes. For more information, see 7.7.2, “Volume migration by adding a volume copy” on page 285.

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, it is repaired by using data from the other copy. This consistency check is performed asynchronously with host I/O.

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.

Mirrored volumes use bitmap space at a rate of 1 bit per 256 KiB grain, which means that 1 MiB of bitmap space supports up to 2 TiB of mirrored volumes. The default size of bitmap space is 20 MiB, which enables 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 7-1 lists the bitmap space configuration options.

Table 7-1 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
Remote copy[b]	0	20 MiB	512 MiB	40 TiB of remote mirroring volume capacity
FlashCopy[c]	0	20 MiB	2 GiB	► 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	► 80 TiB array capacity 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 settings, such as grain size and strip size. RAID is subject to a 15% margin of error.
b. Remote copy includes Metro Mirror, Global Mirror, and active-active relationships.
c. FlashCopy includes the FlashCopy function, Global Mirror with change volumes, and active-active relationships.

The sum of all bitmap memory allocation for all functions except FlashCopy must not exceed 552 MiB.

7.1.7 Volume cache mode

Another adjustable 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 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 is required.

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 is the default cache mode for all volumes.
- *readonly*: Only read I/O operations that are performed on the volume are stored in cache.
- *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. The use of IBM Spectrum Virtualize Copy Services rather than the underlying disk controller copy services gives better results.

Note: Disabling volume cache is a prerequisite for the use of 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.

7.1.8 Fully allocated and thin-provisioned volumes

For each volume the following parameters describe its capacity:

- The real physical capacity that is allocated to the volume from the storage pool; determines how many MDisk extents are initially allocated to the volume.
- The virtual capacity that is reported to the host and to IBM Spectrum Virtualize components (for example, FlashCopy, cache, and remote copy).

In a *fully allocated* volume, these two values are the same. In a *thin-provisioned* volume, the real capacity can be as little as a few percent of virtual capacity. The *real capacity* is used to store the user data and in the case of thin-provisioned volumes, metadata of the volume. The real capacity can be specified as an absolute value, or as a percentage of the virtual 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 fully or thin-provisioned volume.

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 logical address in the volume to a *grain* on a physically allocated extent.

When a write request comes from a host, the block address for which the write is requested is checked against the mapping table. If a previous write to a block on the same grain exists as the incoming request, physical storage was allocated for this logical block address and can be used to service the request; otherwise, new physical grain is allocated to store the data, and the mapping table is updated to record that allocation.

The grain size is defined when the volume is created. The grain size can be 32 KiB, 64 KiB, 128 KiB, or 256 KiB. The grain size cannot be changed after the thin-provisioned volume is created. 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:

- 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.
- 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.
- Grain size affects maximum size of the thin-provisioned volume. For 32 KiB size, the volume size cannot exceed 260 TiB.

Figure 7-7 shows the thin-provisioning concept.

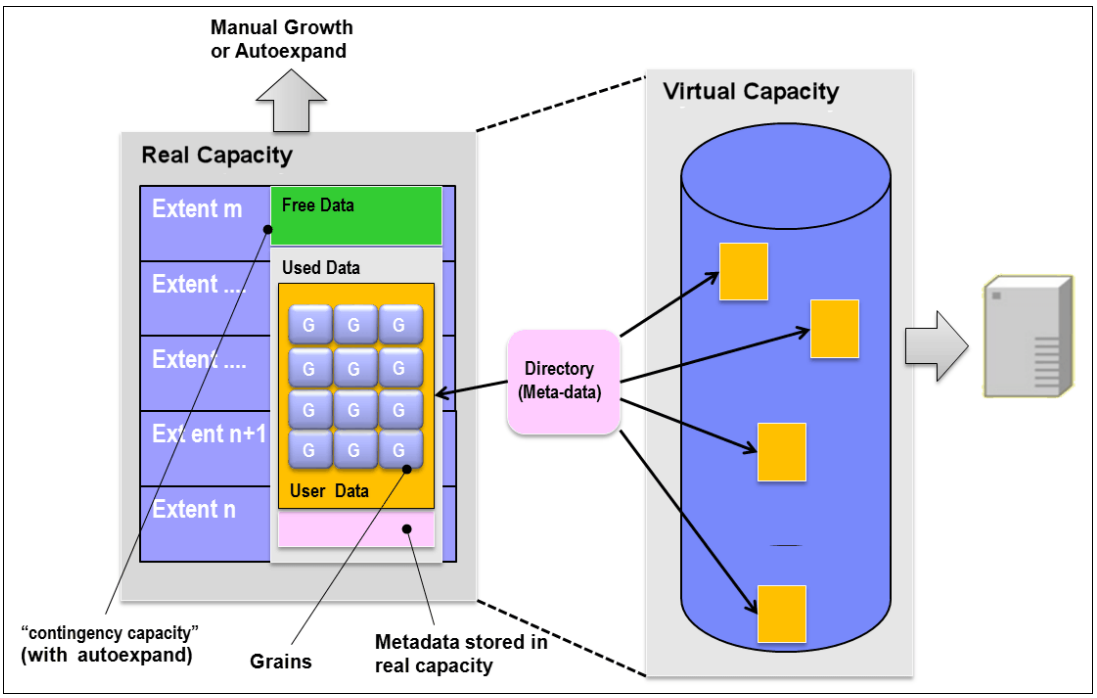


Figure 7-7 Conceptual diagram of thin-provisioned volume

Thin-provisioned volumes store user data and metadata, 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 fully allocated volumes.

The metadata storage 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 not allocated data space, returns zeros. When a write I/O causes space to be allocated, the grain is "zeroed" before use.

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.

Note: Thin-provisioning is supported in standard pools and data reduction pools. However, data reduction pools support reclaiming capacity when it is no longer used by host data and then can redistribute it automatically for other uses. For more information, see 7.1.11, “Capacity reclamation” on page 254.

A thin-provisioned volume can be configured to *autoexpand*. This feature causes the 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 virtual 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 therefore has a zero contingency capacity, goes offline when the real capacity is used and it must expand.

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 fully allocated volume (or vice versa) by using the volume mirroring function. You can create a thin-provisioned copy to a fully allocated primary volume and then remove the fully allocated copy from the volume after they are 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.

7.1.9 Compressed volumes

This custom type of volume is where data is compressed as it is written to disk, which saves more space. To use the compression function, you must obtain the IBM Real-time Compression license.

7.1.10 Deduplicated volumes

Deduplication is a specialized data compression technique for eliminating duplicate copies of data. However, in contrast to the standard file-compression tools that work on single files or sets of files, deduplication is a technique applied on a larger scale, such as a file system or volume. In IBM Spectrum Virtualize deduplication can be enabled for thin provisioned and compressed volumes created in data reduction pools.

Deduplication works by identifying repeating chunks in the data written to the storage system. Spectrum Virtualize calculates a signature for each data chunk (by using a hash function), and checks if that signature is already present in the deduplication database.

If 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 and backup servers. In both cases, it is expected, that multiple copies of identical files exist, such as components of the standard operating system or applications that are used in the organization. Conversely, data that is encrypted or compressed at the file system level do not benefit from deduplication because these operations remove redundancy of the data patterns.

Although deduplication (as with other features of IBM Spectrum Virtualize) is transparent to users and applications, it needs to 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 chances of data corruption by way of a random event, the copies are deduplicated and thus 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 data reduction pools.
- Code level V8.1.3 or higher is needed for deduplication.
- Nodes must have at least 32 GB memory to support deduplication. Nodes that have more than 64 GB memory can use a bigger deduplication fingerprint database, which might lead to better deduplication.
- You must run supported IBM Storwize V7000 hardware. For more information about the current table that specifies valid hardware and features combinations, see this website.

> **Note:** You can use the management GUI or the command-line interface 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.

7.1.11 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. This improves file system performance by reducing the number of I/O operations on physical storage required to perform file deletion.

However, this approach affects the management of 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 SCSI unmap command, which can be issued after file deletion, and informs the storage system that physical blocks that are used by the removed file should be marked as no longer in use and can be freed up. Modern operating systems issue SCSI unmap commands only to storage that advertises support for this feature.

V8.1.0 and later releases support the SCSI unmap command on Spectrum Virtualize systems (including SAN Volume Controller, Storwize V5000, Storwize V7000, and FlashSystem V9000). This enables hosts to notify the storage controller of capacity that is no longer required and may be reused or de-allocated, which might improve capacity savings.

> **Note:** For volumes that are outside data reduction pools, the complete stack from the operating system 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:

- V8.1.2 can also reclaim capacity in data reduction pools when a host issues SCSI unmap commands.
- By default, V8.2.1 does not advertise to hosts support for SCSI unmap.

Before enabling SCSI unmap, see this IBM Support web page.

Analyze your storage stack to optimally balance advantages and costs of data reclamation.

7.1.12 Virtual volumes

IBM Spectrum Virtualize V7.6 introduced support for *virtual volumes*. These volumes enable support for VMware vSphere Virtual Volumes (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 VVol with IBM Spectrum Virtualize, see *Configuring VMware Virtual Volumes for Systems Powered by IBM Spectrum Virtualize*, SG24-8328.

7.1.13 Volumes in multi-site topologies

IBM Spectrum Virtualize system 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 purpose of defining the storage topology, 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.

The sites can be in the same data center room or across rooms in the data center, in buildings on the same campus, or buildings in different cities, depending on the type and scale of a failure the solution must survive.

The following topologies are available for IBM Spectrum Virtualize that is running on Storwize V7000:

- *Standard* topology, which is intended for single-site configurations that does not allow site definition and assumes all components of the solution to be at a single site. Global Mirror or Metro Mirror can be used to maintain a copy of a volume on a different system at a remote site, which can be used for disaster recovery.

- HyperSwap topology, which is a three-site 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 via the other site.

> **Note:** Multi-site topologies of IBM Spectrum Virtualize use two sites as component locations (nodes and back-end storage), and a third site as a location for a tie-breaker component that is used to resolve split-brain scenarios where the storage system components lose communication with each other.
>
> For more information about enhanced stretched cluster and HyperSwap, see the white paper *IBM Spectrum Virtualize HyperSwap configuration*, WP102538.

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.

To start the process of creating a volume, navigate to the Volumes menu and click the **Volumes** option of the IBM Spectrum Virtualize GUI, as shown in Figure 7-8.

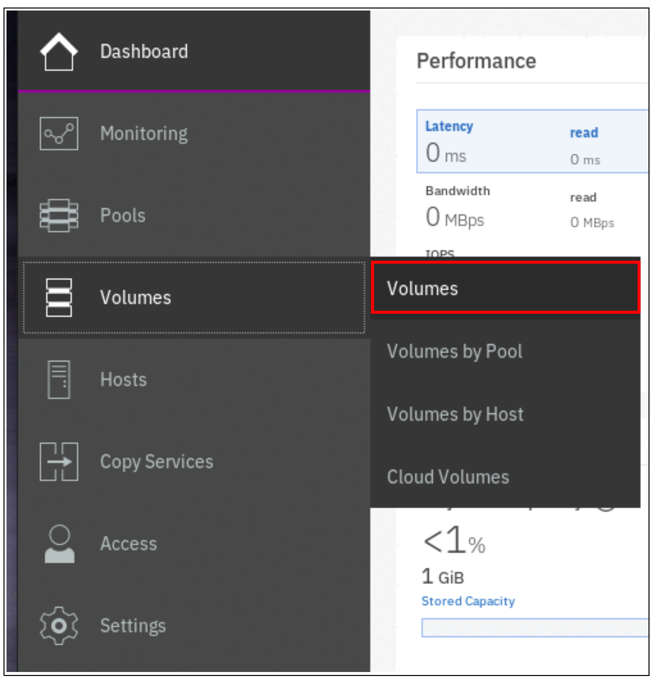


Figure 7-8 Volumes menu

A list of volumes, their state, capacity, and associated storage pools, is displayed.

To create a volume, click **Create Volumes**, as shown in Figure 7-9.

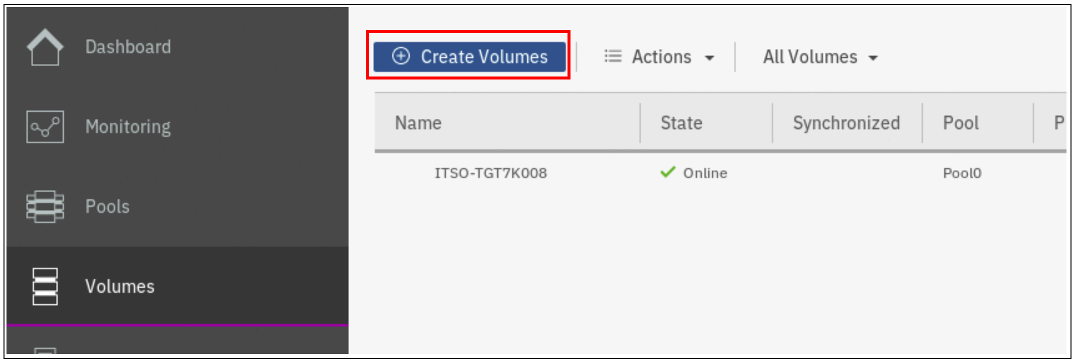


Figure 7-9 Create Volumes button

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 standard topology is shown in Figure 7-10 on page 258.

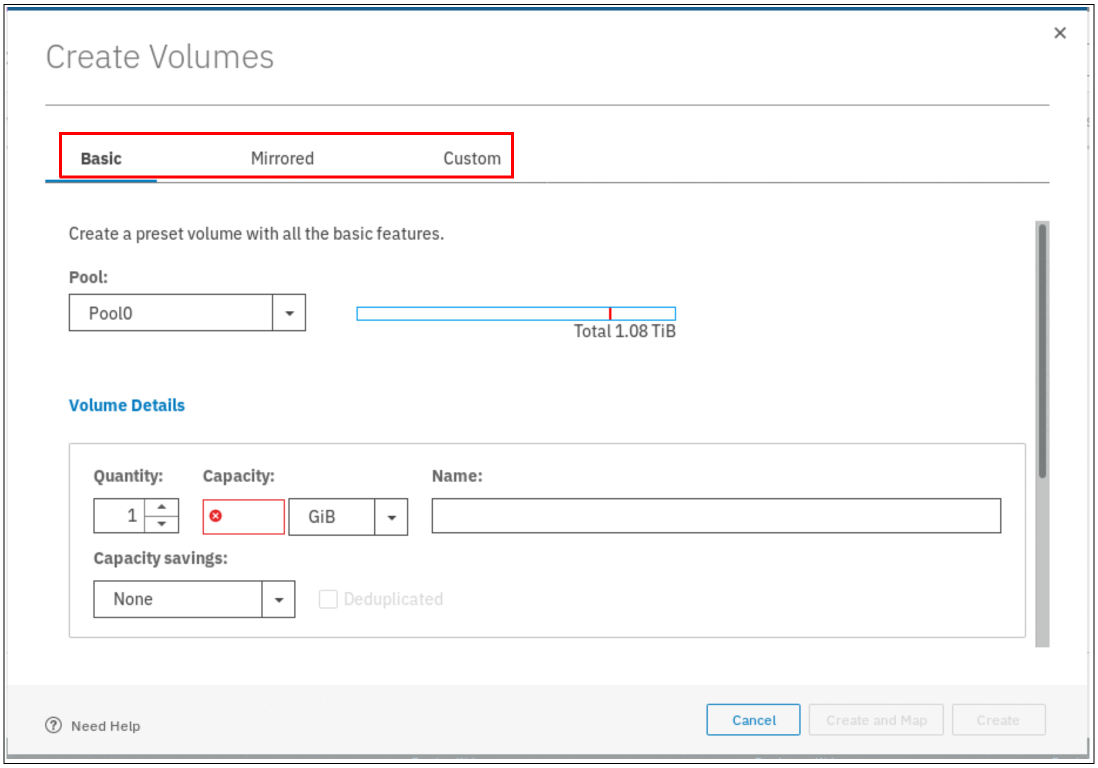


Figure 7-10 Basic, Mirrored, and Custom Volume Creation options

> **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 way of 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 7.2.3, “Capacity savings option” on page 265.

7.2 Creating volumes

This section focuses on the use of the Create Volumes menu to create Basic and Mirrored volumes in a system with standard topology. Volume creation is available on the following volume classes:

- Basic
- Mirrored
- HyperSwap
- Custom

Note: The ability to create HyperSwap volumes by using the GUI simplifies creation and configuration.

To start the process of creating a volume, navigate to the Volumes menu and click the **Volumes** option of the IBM Spectrum Virtualize GUI

7.2.1 Creating basic volumes

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 data reduction pools for added capacity savings.

To create a basic volume, click **Basic** as shown in Figure 7-11 on page 260. This action opens 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 7-11.

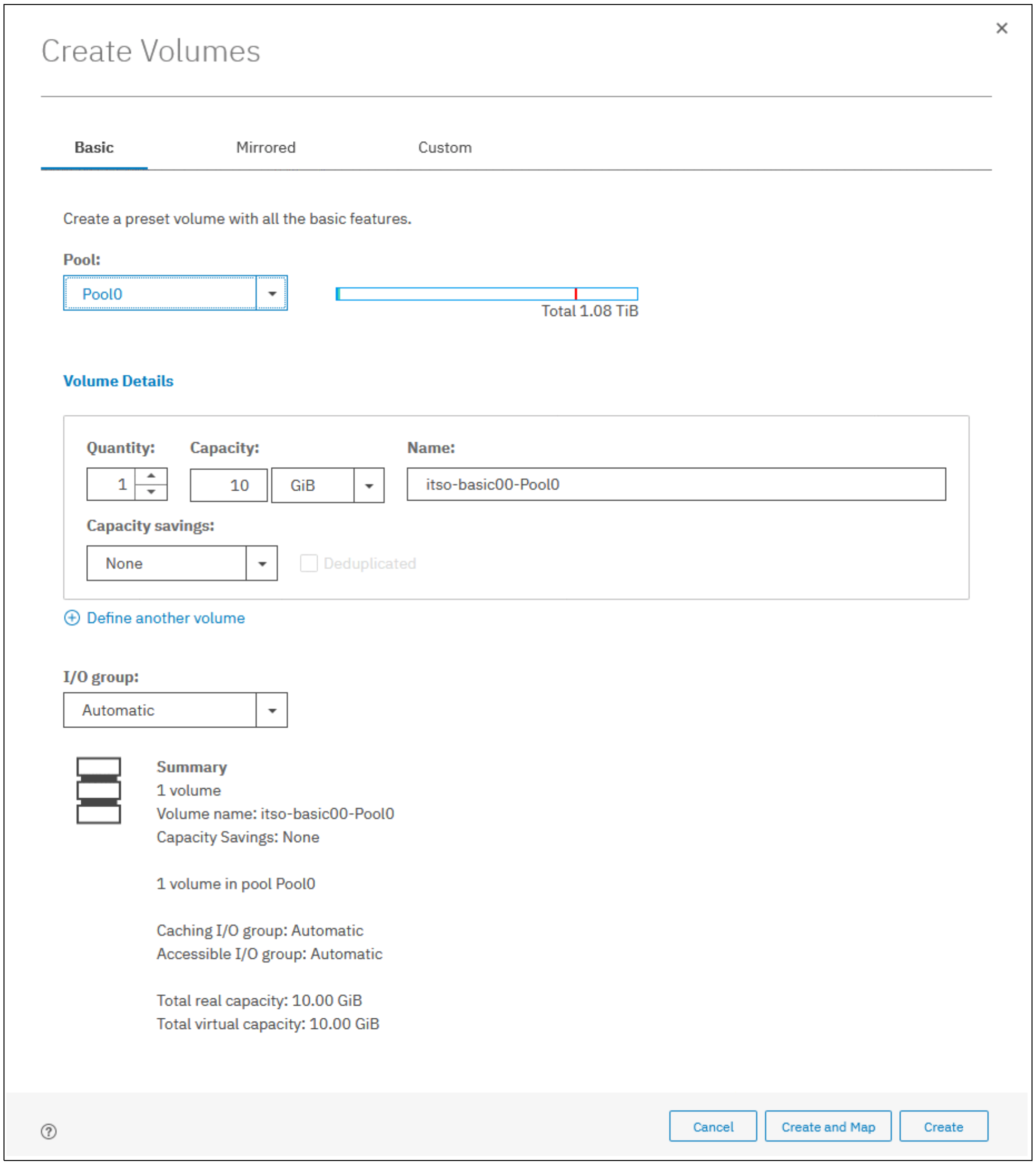


Figure 7-11 Create Volumes window

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 section 7.6, “Mapping a volume to a host” on page 278. When the operation completes, a confirmation window opens, as shown in Figure 7-12.

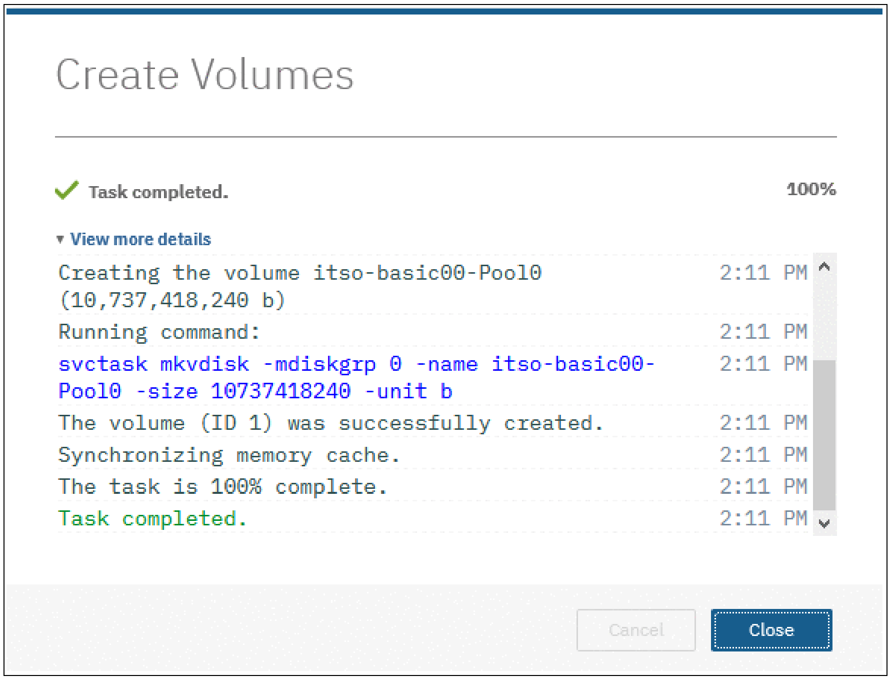


Figure 7-12 Create Volumes Task Completion window

Success is also indicated by the state of the Basic volume being reported as formatting in the Volumes pane, as shown in Figure 7-13.

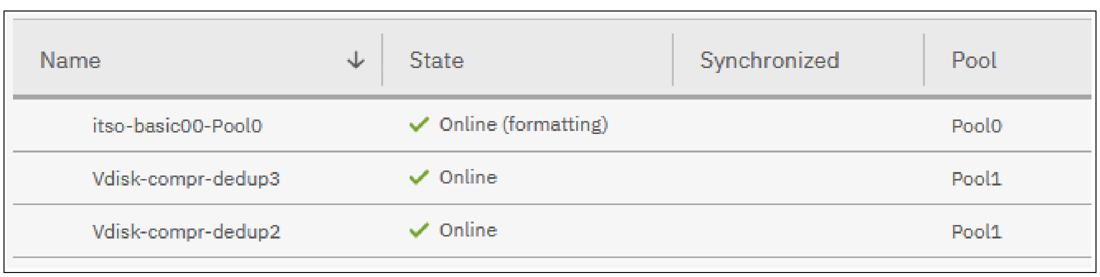


Figure 7-13 Basic volume formatting

Note: Consider the following points:

- Fully allocated volumes are automatically formatted through the quick initialization process after the volume is created. This process makes fully allocated 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 the Fully allocated volumes topic in IBM Knowledge Center.

7.2.2 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 7-14 on page 263).

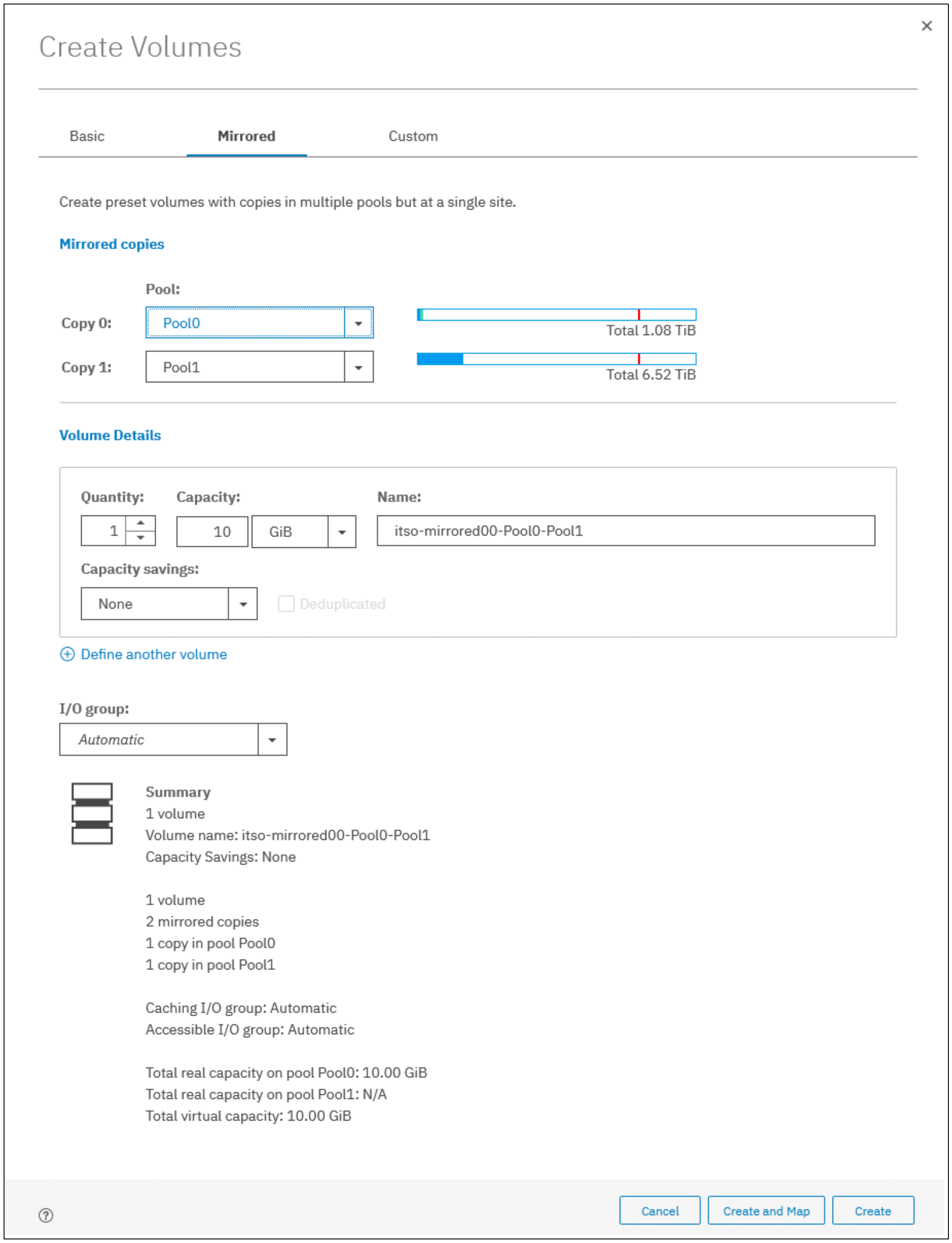


Figure 7-14 Mirrored Volume creation

3. Click **Create** (or **Create and Map**).

The GUI displays the underlying CLI commands that are run to create the mirrored volume and reports completion of the operation, as shown in Figure 7-15.

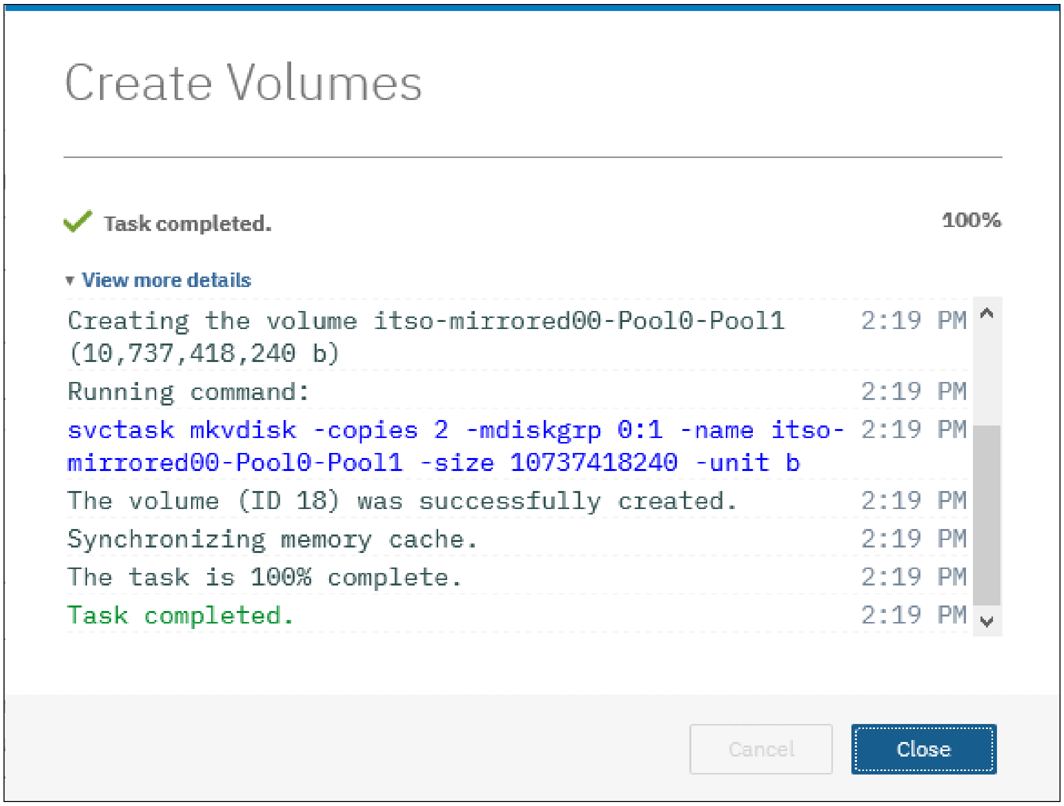


Figure 7-15   Task complete: Created Mirrored Volume

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.

7.2.3 Capacity savings option

When the Basic or Mirrored method of volume creation is sued, the GUI provides a Capacity Savings option, which enables altering the volume provisioning parameters without the use of the Custom volume provisioning method. You can select **Thin-provisioned** or **Compressed** from the drop-down menu, as shown in Figure 7-16, to create thin-provisioned or compressed volumes.

Figure 7-16 Volume Creation with Capacity Saving option

Note: Consider the compression guidelines in Chapter 10, “Advanced features for storage efficiency” on page 403 before creating the first compressed volume copy on a system.

When a thin-provisioned or compressed volume is defined in a data reduction pool, the Deduplicated option becomes available. Select this option to enable deduplication of the volume.

7.3 Creating custom volumes

The Create Volumes window also enables Custom volume creation that expands the set of options for volume creation available to the administrator.

The Custom menu consists of several panes:

- Volume Location: Mandatory, defines the number of volume copies, Pools to be used, and I/O group preferences.
- Volume Details: Mandatory, 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: For configuring for Cache mode and Formatting.
- Summary

Work through these panes to customize your Custom volume as wanted, and then commit these changes by clicking **Create**.

You can mix and match settings on different panes to achieve the final volume configuration that meets your requirements

7.3.1 Volume Location pane

The Volume Location pane is shown in Figure 7-17.

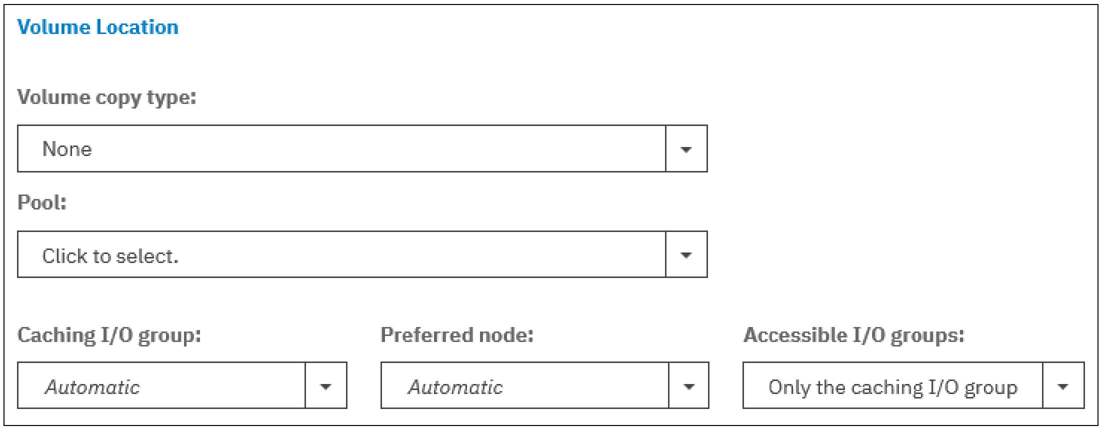


Figure 7-17 Volume Location pane

This pane gives 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 Mirrored volume copy type, and allows you to set the volume copy synchronization rate to a value between 128 KiBps and 64 MiBps.
- Caching I/O group: You can choose between Automatic (allocated by the system) and manually specifying 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.

7.3.2 Volume Details pane

The Volume Details pane is shown in Figure 7-18.

Figure 7-18 Volume Details pane

This pane gives the following options:

- Quantity: Allows you to specify how many volumes are created.
- Capacity: Capacity of the volume.
- Name: Allows you to define the volume name.
- Capacity savings: You can choose between None (fully provisioned volume), Thin-provisioned and Compressed.
- Deduplicated: Thin-provisioned and Compressed volumes created in a data reduction pool 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 7-19.

Figure 7-19 Volume Details pane with two volume subpanes

In this way, you can create volumes with different characteristics in a single invocation of the volume creation wizard.

Thin Provisioning pane

If you choose to create a thin-provisioned volume, a Thin Provisioning pane is displayed, as shown in Figure 7-20.

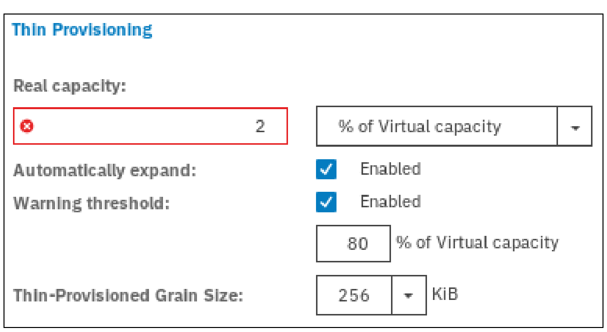


Figure 7-20 Custom volume creation – Thin Provisioning pane

This pane includes 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%.
- Thin-Provisioned Grain Size: Allows you to 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 is dependent upon 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 EasyTier with thin-provisioned volumes, see the IBM Support article *Performance Problem When Using EasyTier With Thin Provisioned Volumes*.

Compressed pane

If you choose to create a compressed volume, a Compressed pane is displayed, as shown in Figure 7-21.

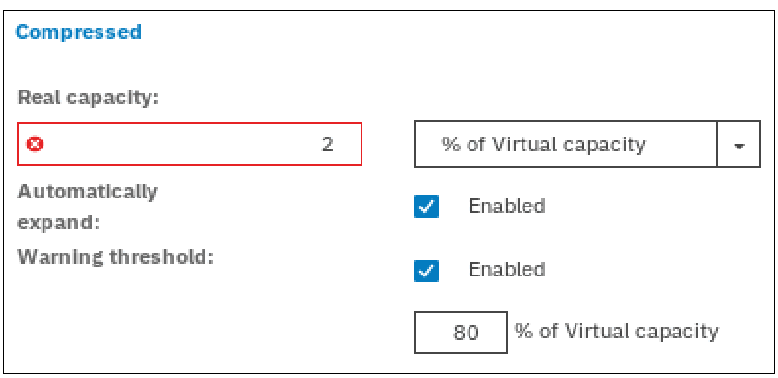


Figure 7-21 Compressed pane

This pane 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 grain size for a compressed volume.

Note: Consider the compression guidelines in Chapter 10, “Advanced features for storage efficiency” on page 403 before creating the first compressed volume copy on a system.

7.3.3 General pane

The General pane is shown in Figure 7-22.

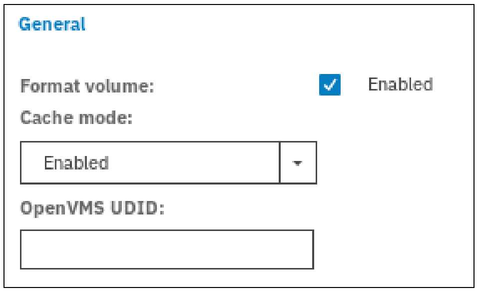


Figure 7-22 Custom volume creation – General pane

This pane gives the following options:

- Format volume: Controls whether the volume is formatted before being made available; defaults to Enabled.
- Cache mode: Controls volume caching; defaults to Enabled. Other available options are Read-only and Disabled.
- OpenVMS UDID: Each OpenVMS Fibre Channel-attached volume requires a user-defined identifier or unit device identifier (UDID). A UDID is a nonnegative integer that is used in the creation of the OpenVMS device name.

7.4 HyperSwap volumes

The HyperSwap function provides highly available volumes accessible through two sites at up to 300 km (186.4 miles) 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 completion is reported to the host. The HyperSwap function automatically optimizes itself to minimize data 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 the foundation of two earlier technologies: Non-Disruptive Volume Move (NDVM) function that was introduced in IBM Spectrum Virtualize V6.4, and the Remote Copy features that include Metro Mirror, Global Mirror, and Global Mirror with Change Volumes.

HyperSwap volume configuration is possible only after the IBM Spectrum Virtualize system is configured in HyperSwap topology. After this topology change, the GUI presents an option to create a HyperSwap volume and creates them by using the **`mkvolume`** command instead of the usual **`mkvdisk`** command. The GUI continues to uses the **`mkvdisk`** command when all other classes of volumes are created.

> **Note:** It is still possible to create HyperSwap volumes in V7.5, as described in this IBM Support white paper.
>
> For more information, see *IBM Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317.
>
> For more information about HyperSwap topology, see this IBM Knowledge Center web page.

From the perspective of a host or a storage administrator, a HyperSwap volume is a single entity; however, it is realized by using four volumes and a set of FlashCopy maps and a remote copy relationship (see Figure 7-23).

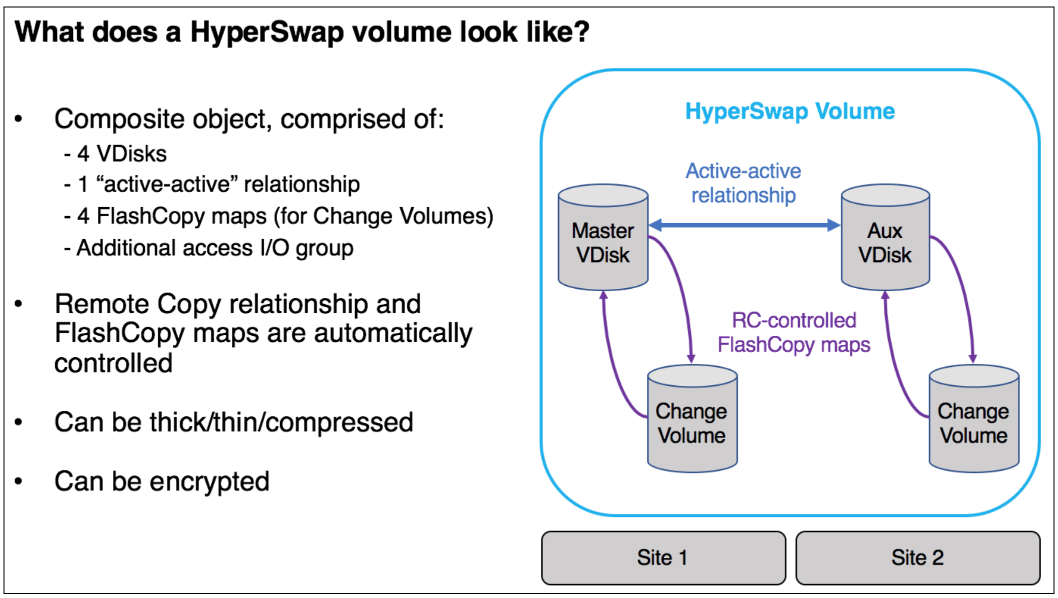


Figure 7-23 What makes up a HyperSwap Volume

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 7-24 on page 271.

The notable difference between a HyperSwap volume and a basic volume creation is that in the HyperSwap Details the system asks for storage pool names at each site. The system uses its topology awareness to map storage pools to sites, which ensures that the data is correctly mirrored across locations.

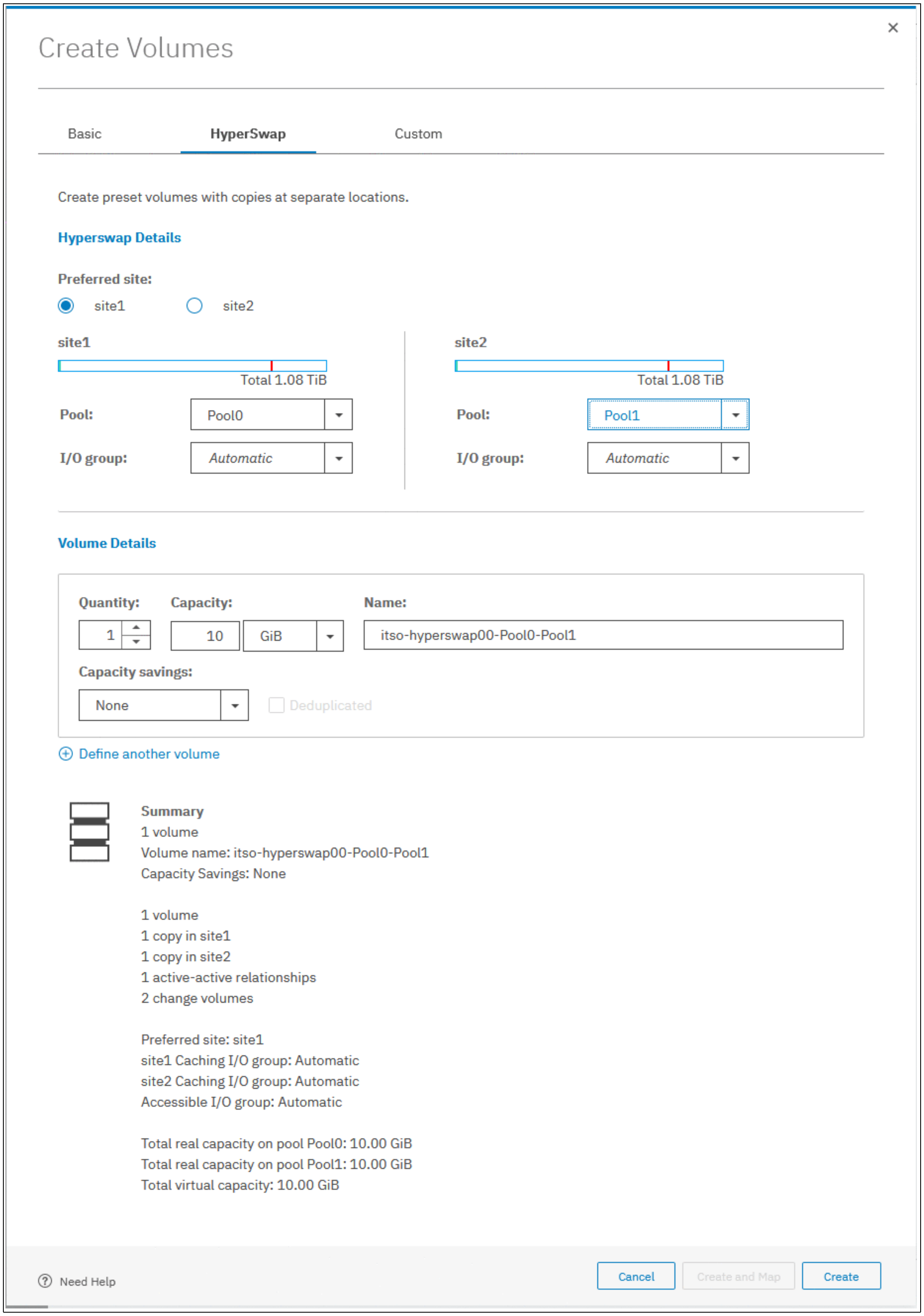


Figure 7-24 HyperSwap Volume creation window

A summary (lower left of the creation window) lists the entities that are instantiated when the Create option is selected. As shown in Figure 7-24 on page 271, a single volume is created, with volume copies in sites `site1` and `site2`. This volume is in an active-active (Metro Mirror) relationship with extra resilience provided by two change volumes.

The command that is issued to create this volume is shown in Figure 7-25.

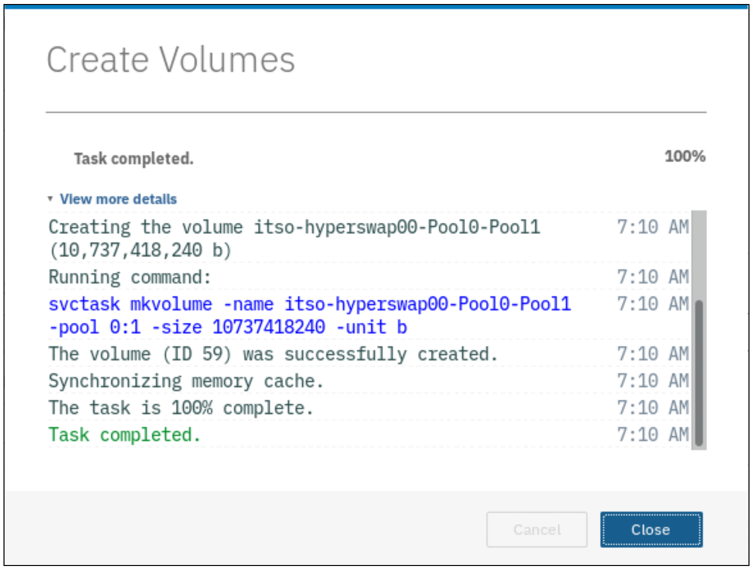


Figure 7-25   Example mkvolume command

After the volume is created, it is visible in the volumes list, as shown in Figure 7-26

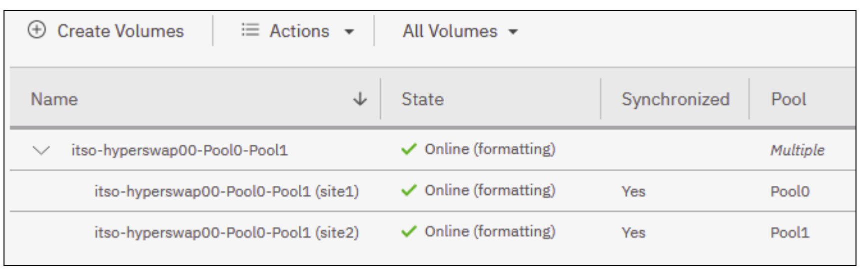


Figure 7-26   A HyperSwap volume visible in the Volumes list

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.

A single **`mkvolume`** command allows creation of a HyperSwap volume. Up to IBM Spectrum Virtualize V7.5, this process required careful planning and the use of the following sequence of commands:

- `mkvdisk master_vdisk`
- `mkvdisk aux_vdisk`
- `mkvdisk master_change_volume`
- `mkvdisk aux_change_volume`
- `mkrcrelationship –activeactive`
- `chrcrelationship -masterchange`
- `chrcrelationship -auxchange`
- `addvdiskacces`

7.5 I/O throttling

You can set a limit on the number 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 I/O operations per second (IOPS) or bandwidth (MBps, GBps, or TBps). By default, I/O throttling is disabled, but each volume can have up to two throttles defined: one for bandwidth and one for IOPS.

When deciding between the use of 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 based on IOPS would be better suited to this use case.

Conversely, a video streaming application generally 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 Metro Mirror or Global Mirror secondary volumes does not affect the rate of data copy from the primary volume.

7.5.1 Defining a volume throttle

To set a volume throttle, complete these steps:

1. From the **Volumes** → **Volumes** menu, select the wanted volume to throttle. From the **Actions** menu, select **Edit Throttle**, as shown in Figure 7-27.

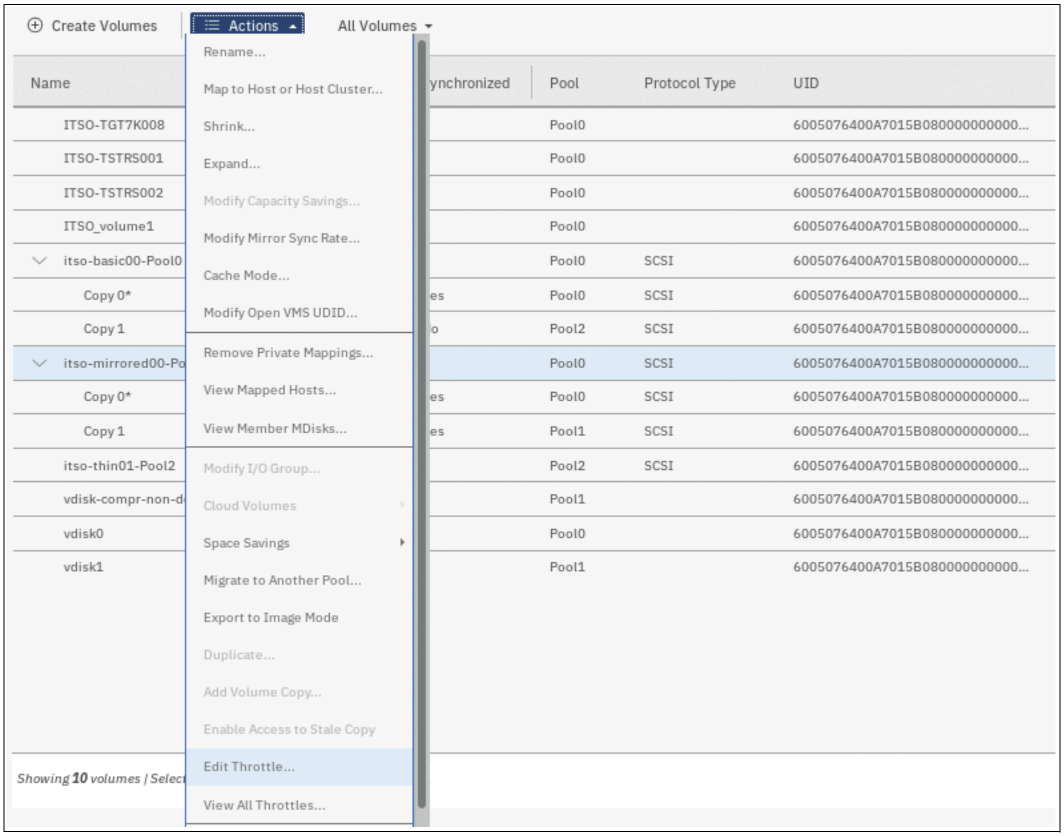


Figure 7-27 Edit throttle

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 7-28. Click **Create**.

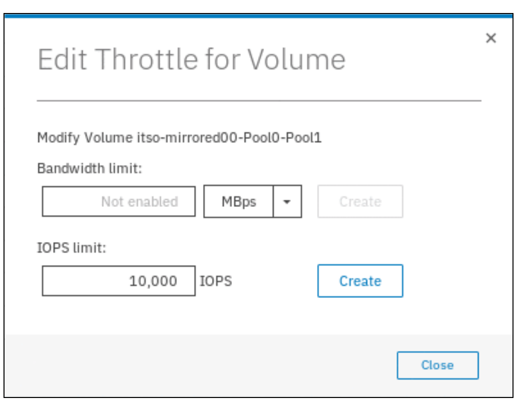


Figure 7-28 IOPS throttle on a volume

3. After the Edit Throttle task completes successfully, the Edit Throttle window is shown again. You can now set throttle based on the different metrics, modify throttle, or close the window without performing any action by clicking **Close**.

7.5.2 Listing volume throttles

To view volume throttles, from the **Volumes** → **Volumes** menu, click the **Actions** menu and select **View All Throttles**, as shown in Figure 7-29.

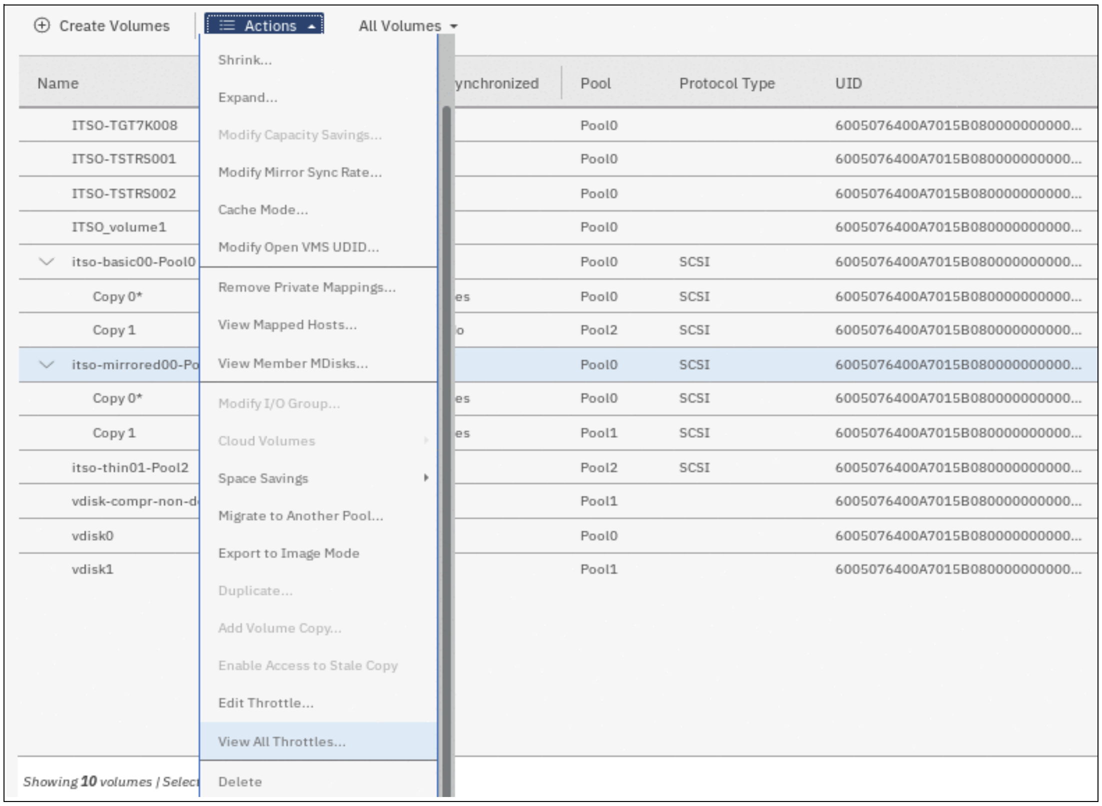


Figure 7-29 View all throttles

The **View All Throttles** menu shows all volume throttles defined in the system, as shown in Figure 7-30.

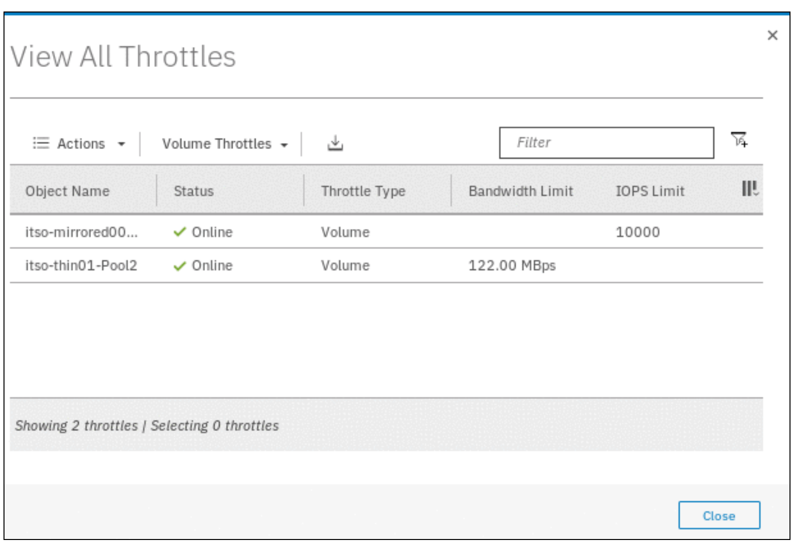


Figure 7-30 View volume throttles

You can view other throttles by selecting a different throttle type in the drop-down menu, as shown in Figure 7-31.

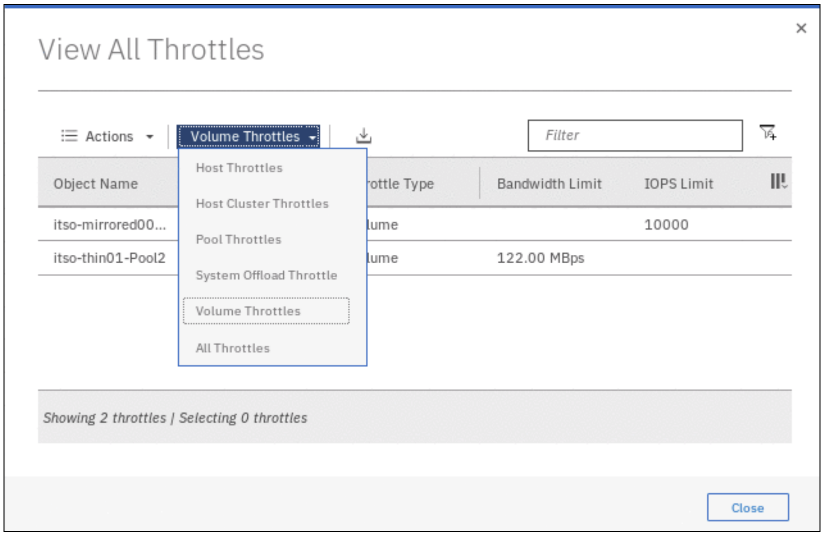


Figure 7-31 Filter throttle type

7.5.3 Modifying or removing a volume throttle

To remove a volume throttle, complete the following steps:

1. From the **Volumes** menu, select the wanted volume to which is attached the throttle that you want to remove. From the **Actions** menu, select **Edit Throttle**, as shown in Figure 7-32 on page 277.

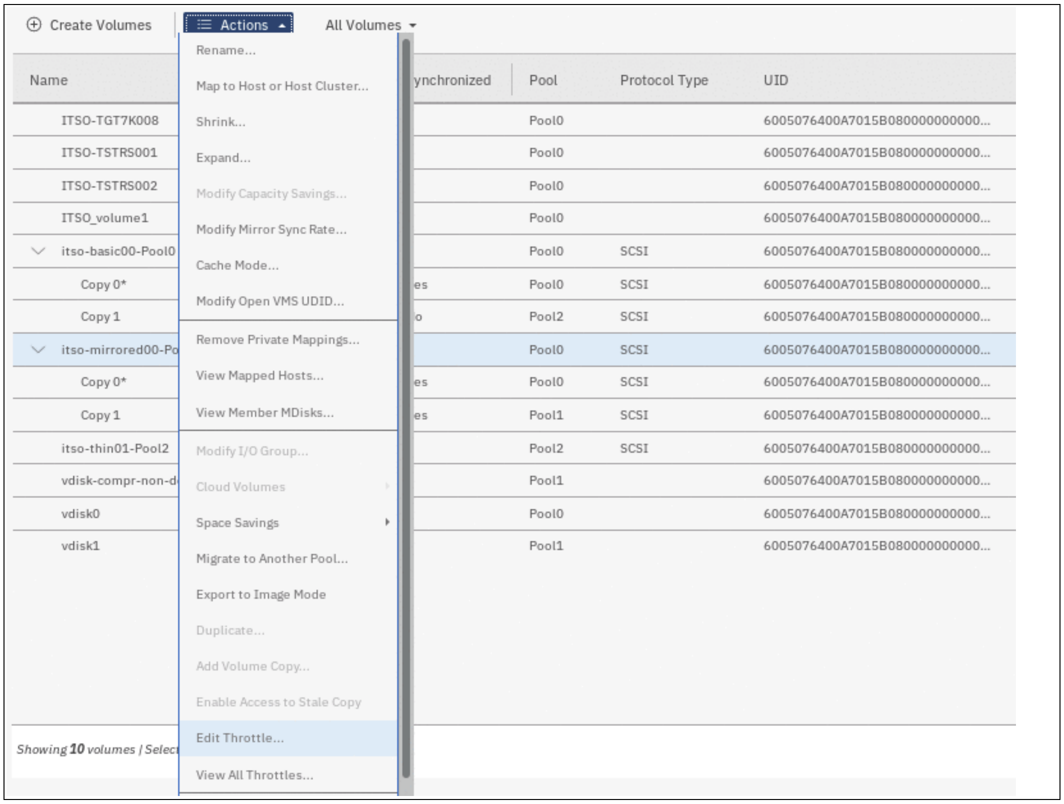


Figure 7-32 Edit throttle

2. In the Edit Throttle window, click **Remove** for the throttle you want to remove. As shown in Figure 7-33, we remove the IOPS throttle from the volume.

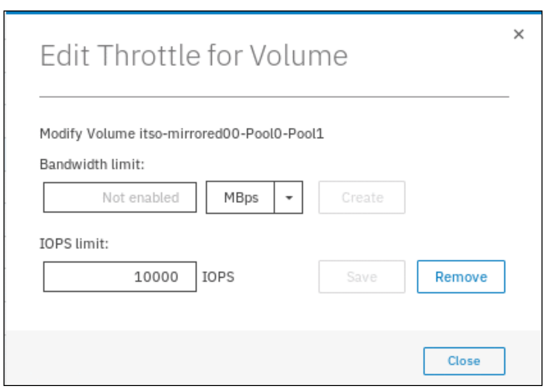


Figure 7-33 Remove throttle

After the Edit Throttle task completes successfully, the Edit Throttle window is shown again. You can now set throttle based on the different metrics, modify throttle, or close the window without performing any action by clicking **Close**.

7.6 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. In the **Volumes** view (see Figure 7-34) select the volume for which you want to create a mapping and then, select **Actions** from the menu bar.

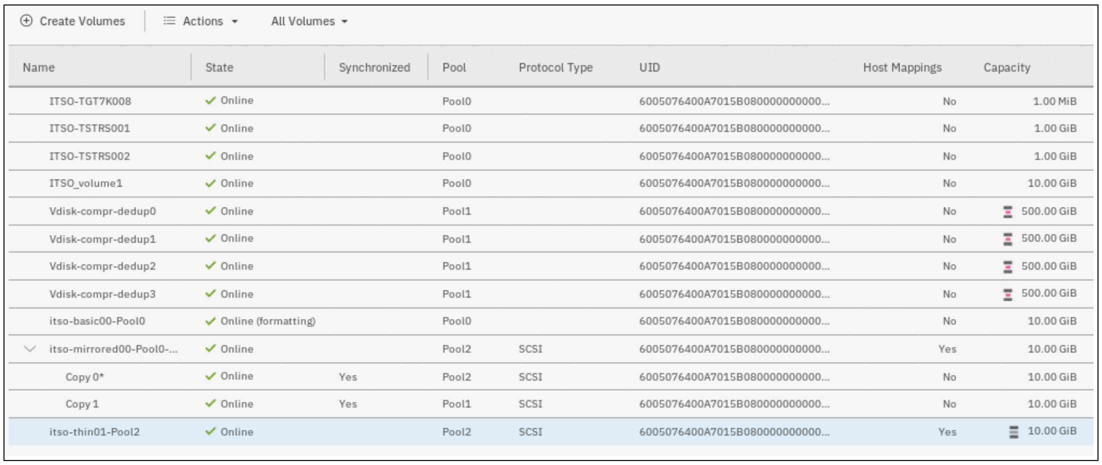


Figure 7-34 Volume list

> **Tip:** An alternative way of opening the Actions menu is to highlight (select) a volume and use the right mouse button.

2. From the **Actions** menu, select the **Map to Host or Host Cluster** option, as shown in Figure 7-35.

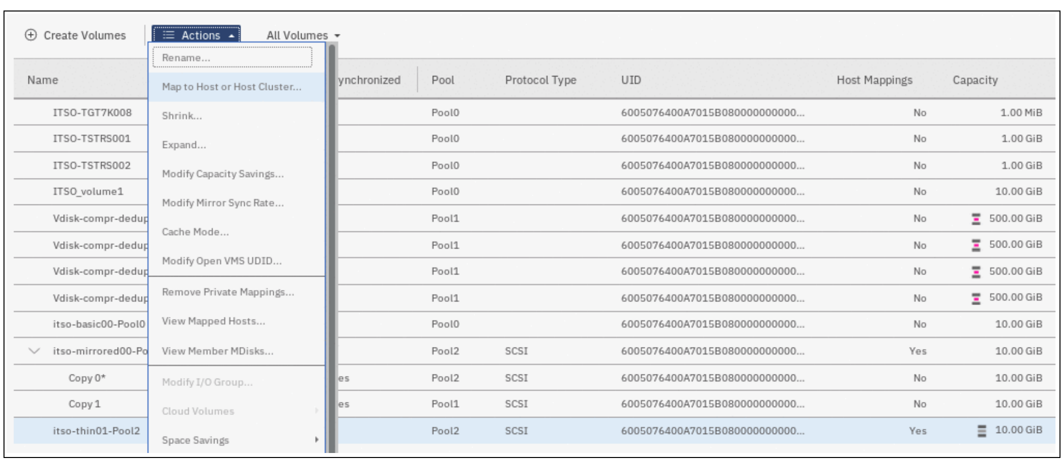


Figure 7-35 Map to Host or Host Cluster

3. This action opens a Create Mapping window. In this window, select whether to create a mapping to a Host or Host Cluster. The list of clusters is displayed. Then, select which volumes to create the mapping for.

You can also select whether you want to assign SCSI LUN ID to the volume (Self-Assign option) or allow IBM Spectrum Virtualize to assign the ID (System Assign option). In the example that is shown in Figure 7-36, a single volume is mapped to a host and the system assigns the SCSI LUN IDs.

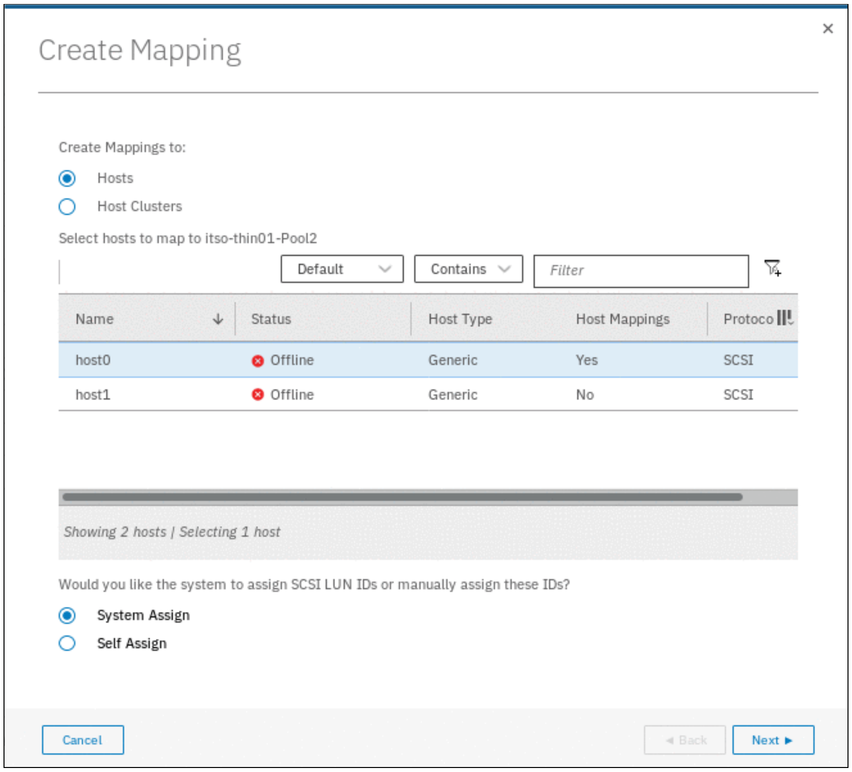


Figure 7-36 Mapping a volume to a host

4. A summary window is displayed showing the volume to be mapped along with existing volumes that are mapped to the host or host cluster, as shown in Figure 7-37. Click **Map Volumes**.

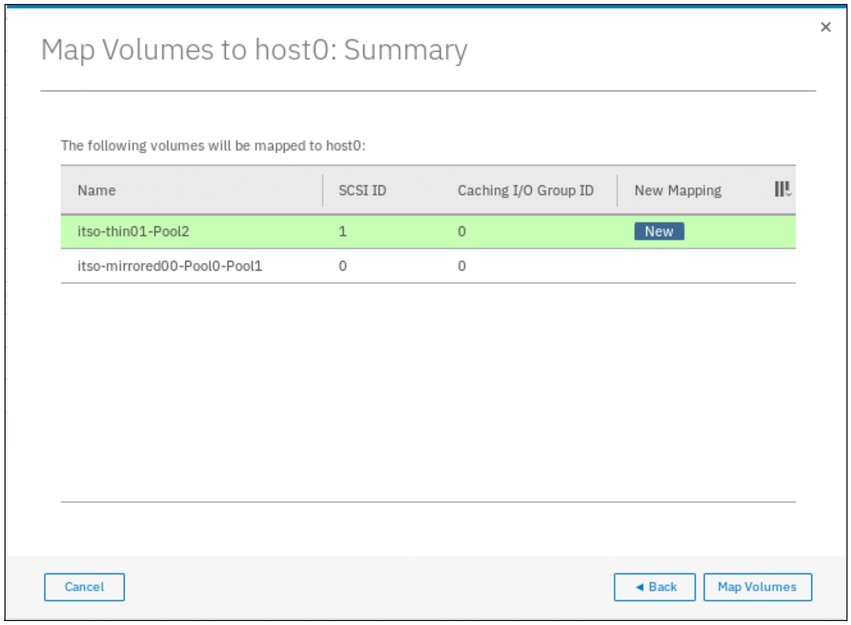


Figure 7-37 Map volume to host cluster summary

The confirmation window shows the result of the volume mapping task, as shown in Figure 7-38.

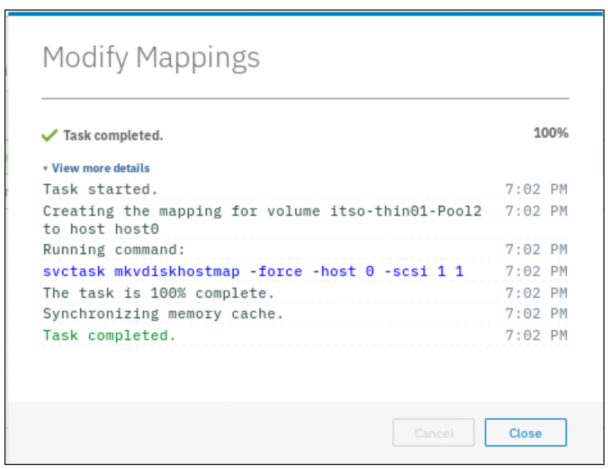


Figure 7-38 Confirmation of volume to host mapping

5. After the task completes, the wizard returns to the Volumes window. You can list volumes mapped to the host by navigating to **Hosts** → **Mappings**, as shown in Figure 7-39

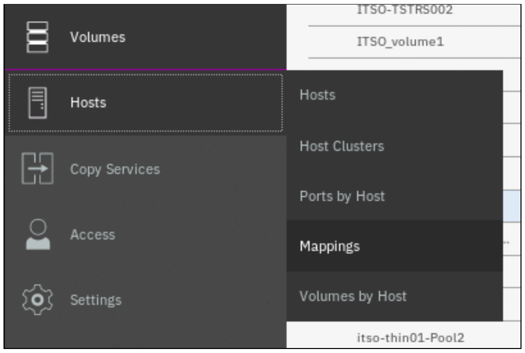
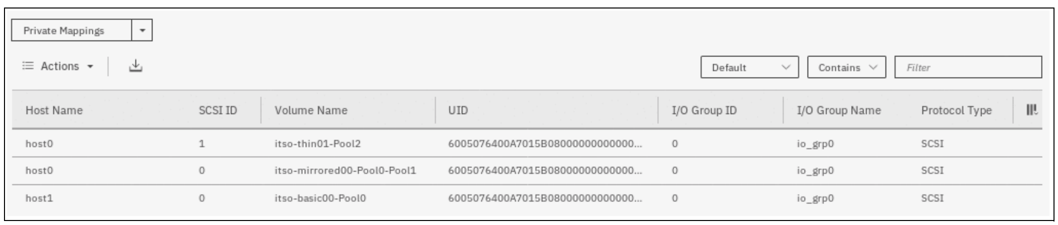


Figure 7-39 Accessing the Hosts Mapping menu

6. List of volumes mapped to all hosts is displayed, as shown in Figure 7-40.



Host Name	SCSI ID	Volume Name	UID	I/O Group ID	I/O Group Name	Protocol Type
host0	1	itso-thin01-Pool2	6005076400A7015B080000000000000...	0	io_grp0	SCSI
host0	0	itso-mirrored00-Pool0-Pool1	6005076400A7015B080000000000000...	0	io_grp0	SCSI
host1	0	itso-basic00-Pool0	6005076400A7015B080000000000000...	0	io_grp0	SCSI

Figure 7-40 List of volumes mapped to hosts

To see volumes mapped to clusters instead of hosts, change the value that is shown in the upper left corner (see Figure 7-41 on page 282) from **Private Mappings** to **Shared Mappings**.

> **Note:** You can use the filter to display only the wanted hosts or volumes.

The host can now access the mapped volume. For more information about discovering the volumes on the host, see Chapter 8, “Hosts” on page 317.

7.7 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. These approaches are described next.

7.7.1 Volume migration using migration feature

The migration process 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 at least one 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 7-41.

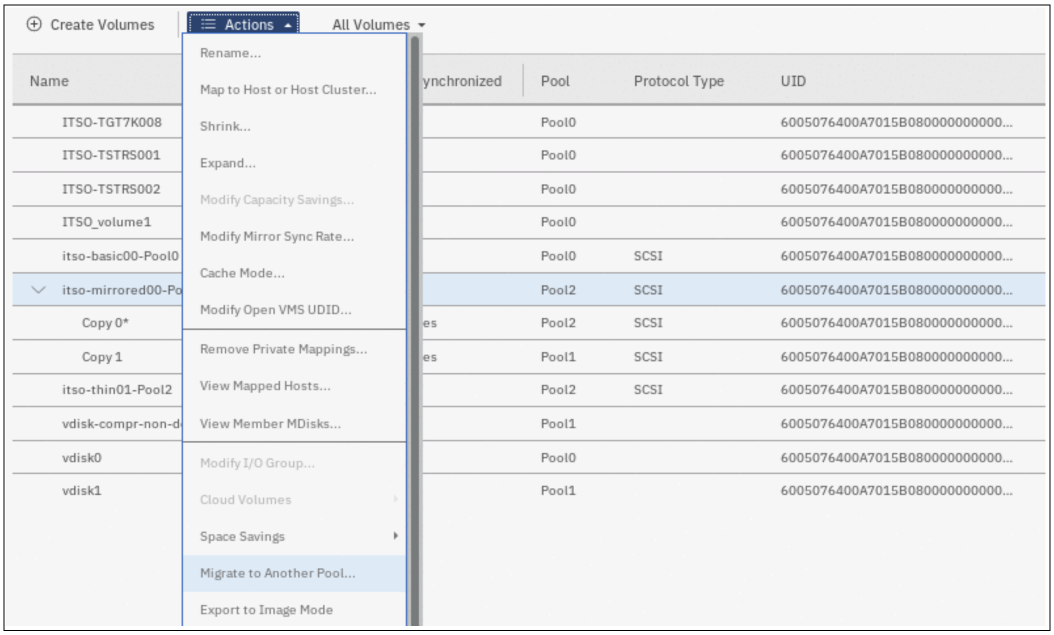


Figure 7-41 Migrate Volume Copy window: Select menu option

2. 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 7-42. If the selected volume consists of one copy, the volume copy selection pane is not displayed.

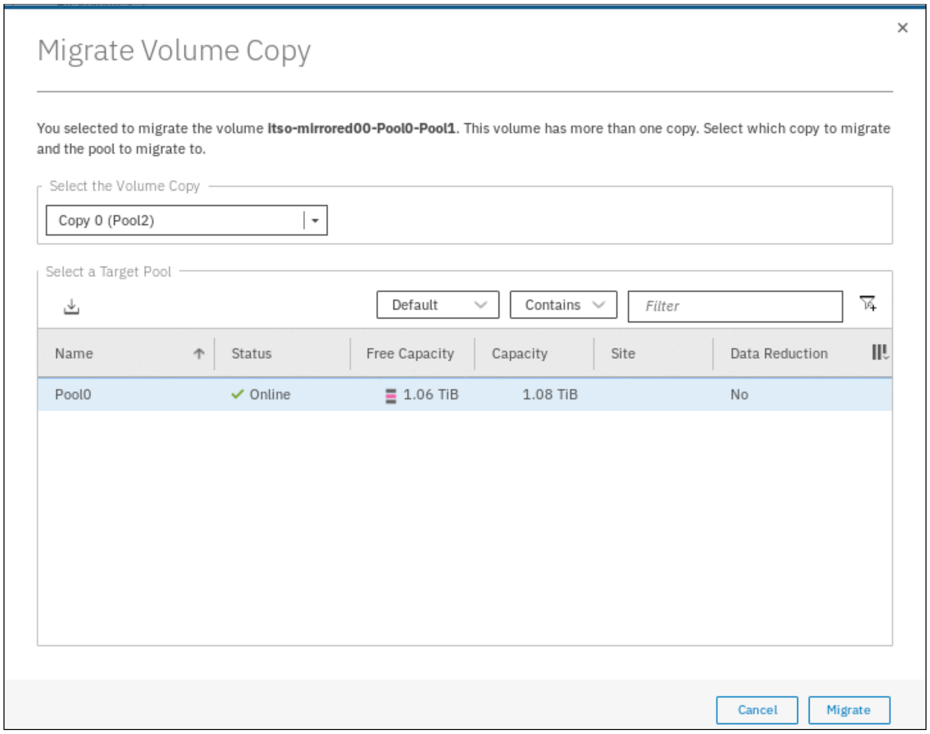


Figure 7-42 Migrate Volume Copy: Selecting the volume copy

Select the new target storage pool and click **Migrate**, as shown in Figure 7-42. The Select a Target Pool pane displays the list of all pools that are a valid migration copy target for the selected volume copy.

3. The volume copy migration process starts, as shown in Figure 7-43. Click **Close** to return to the Volumes window.

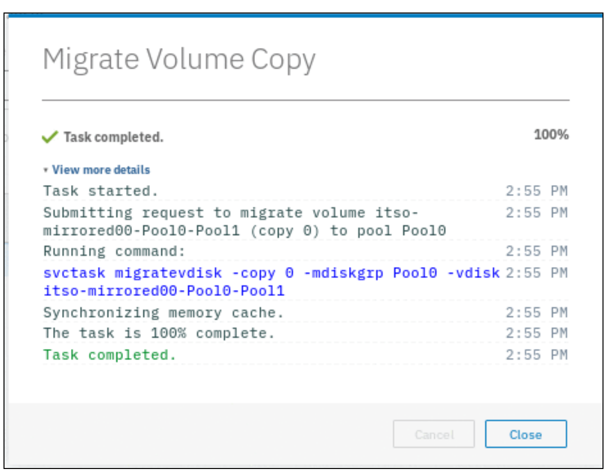


Figure 7-43 Migrate Volume Copy started

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 navigating to **Monitoring** → **Background Tasks**, as shown in Figure 7-44.

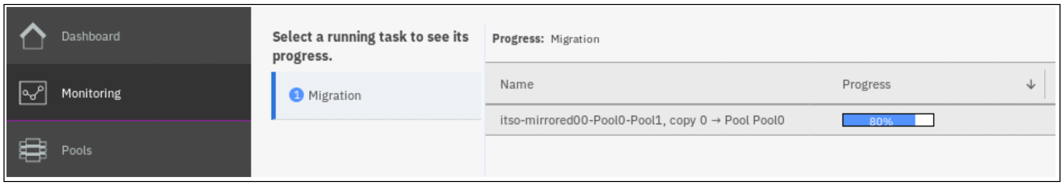


Figure 7-44 Migration progress

After the migration task completes, the Background Tasks menu displays a Recently Completed Task, as shown in Figure 7-45.

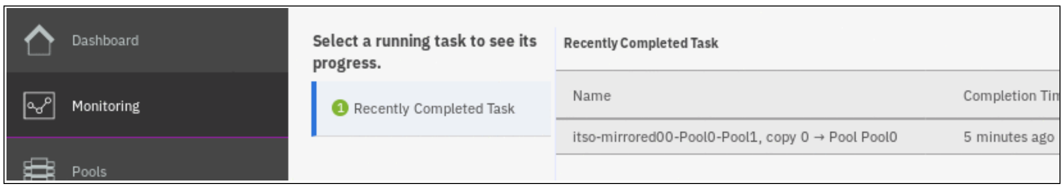


Figure 7-45 Migration complete

In the **Volumes** → **Volumes** menu, the volume copy is now displayed in the target storage pool, as shown in Figure 7-46.

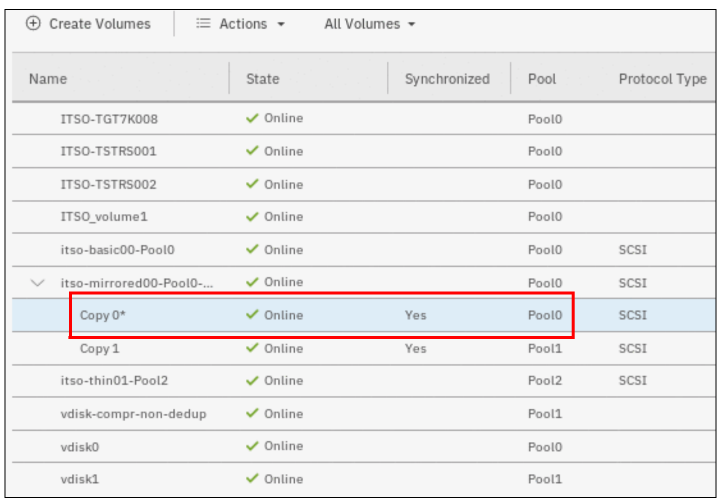


Figure 7-46 Volume copy after migration

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 7.7.2, "Volume migration by adding a volume copy" on page 285.

> **Note:** Migrating a volume between storage pools with different extent sizes is *not* supported. If you need to migrate a volume to a storage pool with a different extent size, use the volume copy feature instead.

7.7.2 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 to other storage pools.

The easiest way to migrate volume copies is to use the migration feature that is described in 7.7, "Migrating a volume to another storage pool" on page 281. 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 these of the primary copy. For example, you can make a thin-provisioned copy of a fully 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 using the volume copy feature, complete the following steps:

1. Select the volume that you want to move, and click **Actions** → **Add Volume Copy**, as shown in Figure 7-47.

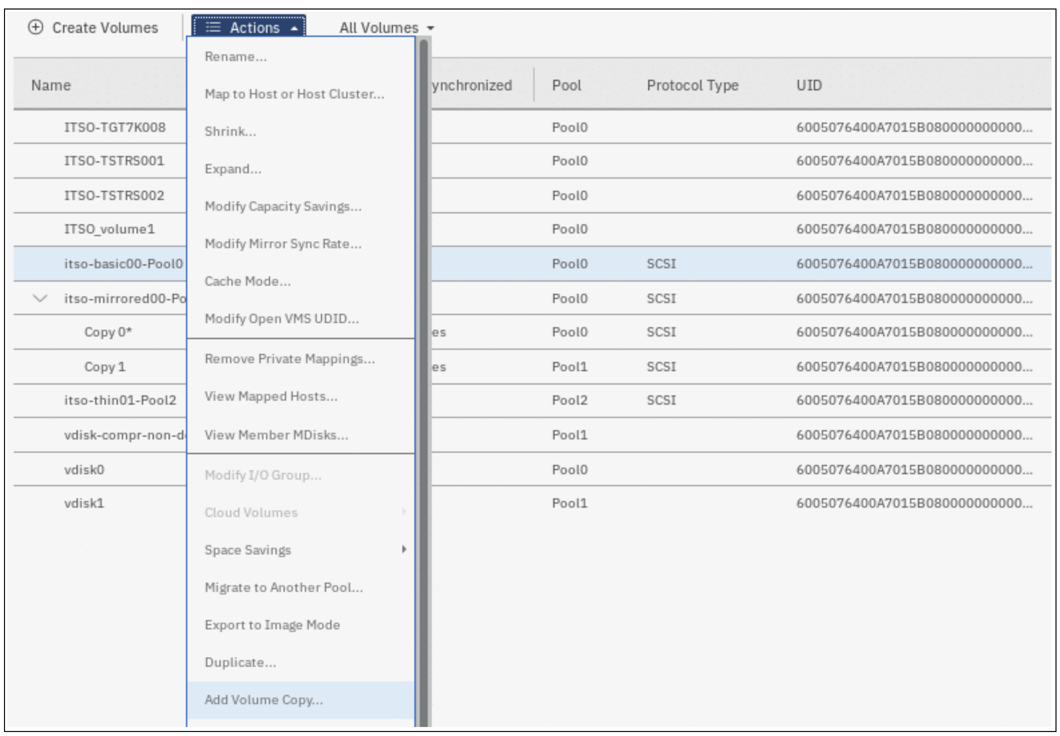


Figure 7-47 Adding the volume copy

2. Create a second copy of your volume in the target storage pool, as shown in Figure 7-48. In our example, a compressed copy of the volume is created in target pool `Pool2`. Click **Add**.

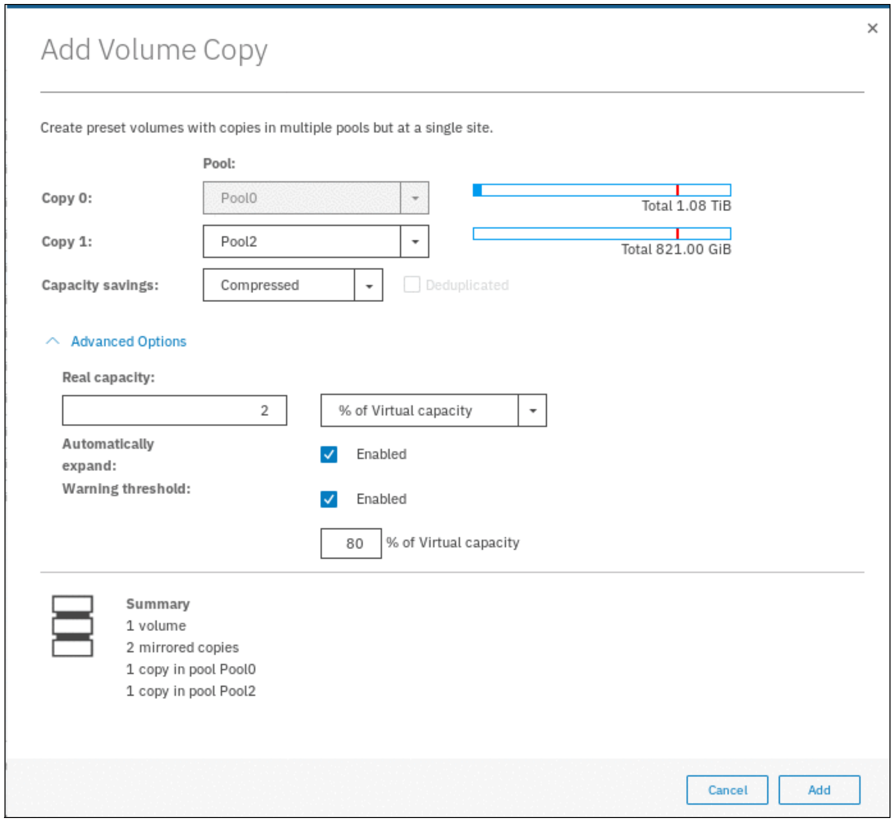


Figure 7-48 Defining the new volume copy

Wait until the copies are synchronized, as shown in Figure 7-49.

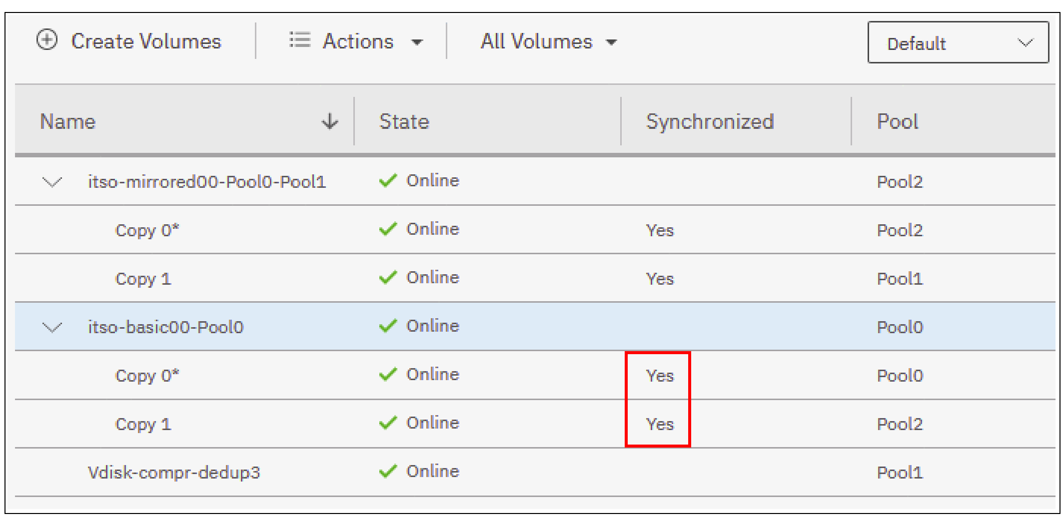


Figure 7-49 Waiting for the volume copies to synchronize

3. Change the roles of the volume copies by making the new copy the primary copy, as shown in Figure 7-50. The current primary copy is displayed with an asterisk next to its name.

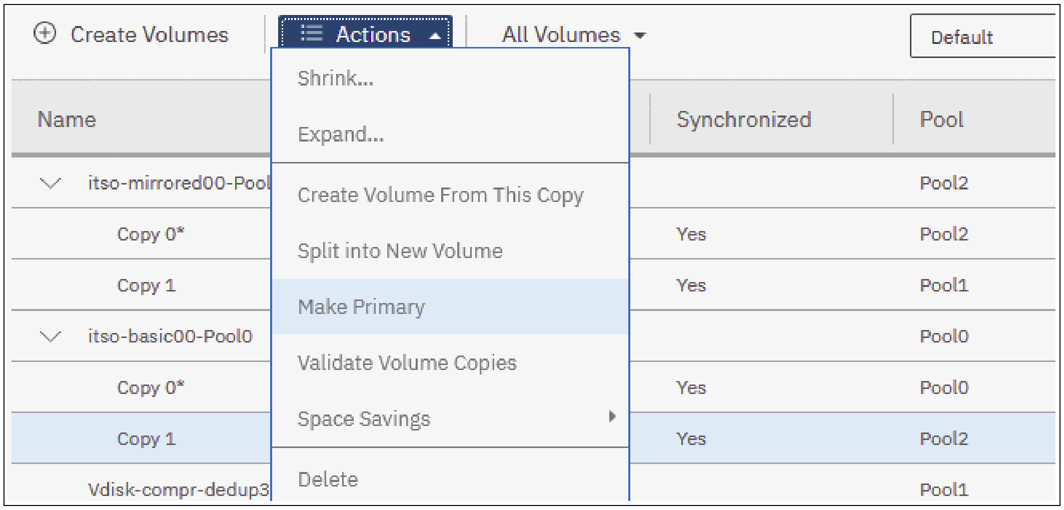


Figure 7-50 Making the new copy in a different storage pool the primary

4. Split or delete the old copy from the volume, as shown in Figure 7-51.

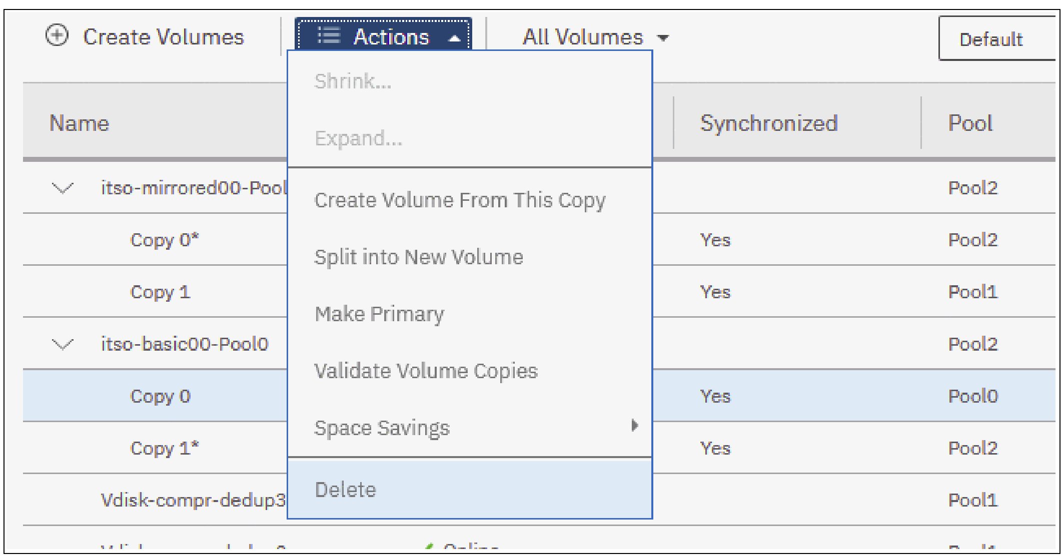


Figure 7-51 Deleting the old volume copy

5. The Volumes view now shows that the volume has a single copy in the wanted pool, as shown in Figure 7-52.

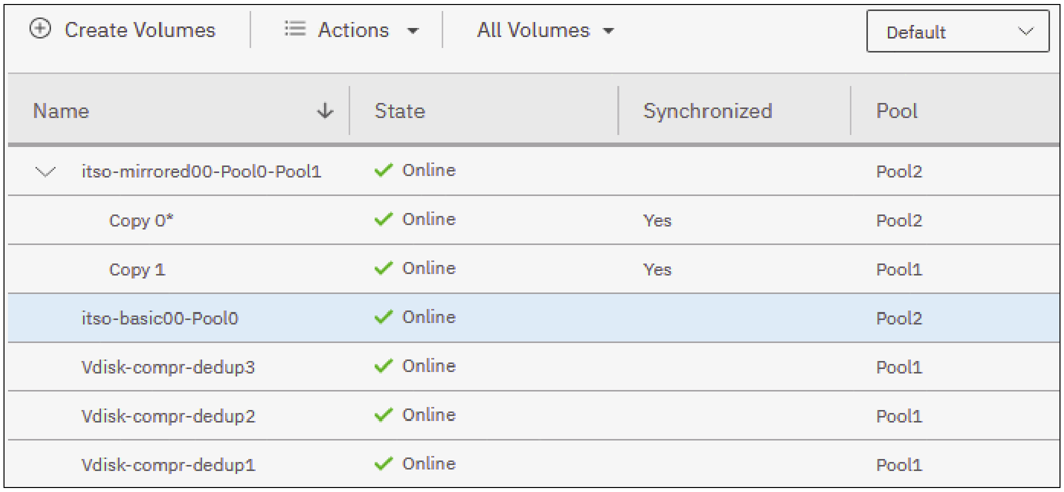


Figure 7-52 Volume copy in the target storage pool

The migration of 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 only done from the primary copy.

Now, the roles of the volume copies can be switched such that the new copy is the primary (step 4) to test the performance of the new pool. If the performance is acceptable, the volume copy in tier 1 can be split or deleted. If the testing of the tier 2 pool shows unsatisfactory performance, the old copy can be made the primary again to switch volume reads back to the tier 1 copy.

7.8 Volume operations by using the CLI

This section describes how to perform various volume configuration and administrative tasks using the CLI. For more information, see the Command-line interface section in IBM Knowledge Center.

For more information about how to set up CLI access, see Appendix B, “CLI setup” on page 755.

7.8.1 Displaying volume information

Use the **lsvdisk** command to display information about all volumes that are defined within the IBM Spectrum Virtualize environment. To display more d 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 7-1.

Example 7-1 The lsvdisk command

```
IBM_Storwize:ITSO: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 Pool1 15.00GB striped
6005076400F580049800000000000003 0 1 empty 0 no 1 1 Pool1 no yes 1 COMPRESSED_VOL_1
2 vdisk0 0 io_grp0 online 0 Pool0 10.00GB striped     6005076400F580049800000000000004 0 1
empty 0 no 0 0 Pool0 no yes 2 vdisk0
3 THIN_PROVISION_VOL_1 0 io_grp0 online 0 Pool0 100.00GB striped
6005076400F580049800000000000005 0 1 empty 1 no 0 0 Pool0 no yes 3 THIN_PROVISION_VOL_1
4 COMPRESSED_VOL_2 0 io_grp0 online 1 Pool1 30.00GB striped
6005076400F580049800000000000006 0 1 empty 0 no 1 1 Pool1 no yes 4 COMPRESSED_VOL_2
5 COMPRESS_VOL_3 0 io_grp0 online 1 Pool1 30.00GB striped
6005076400F580049800000000000007 0 1 empty 0 no 1 1 Pool1 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 Pool0 10.00GB striped     6005076400F580049800000000000010 0 1
not_empty 0 no 0 0 Pool0 yes yes 8 Tiger
12 vdisk0_restore 0 io_grp0 online 0 Pool0 10.00GB striped
6005076400F58004980000000000000E 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
6005076400F58004980000000000000F 0 1 empty 0 no 0 0 Pool0 no yes 13 vdisk0_restore1
```

7.8.2 Creating a volume

The use of 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, use the **mkvdisk** command. The command shown in Example 7-2 creates a 10 gigabyte (GB) striped volume with volume ID 8 within the storage pool Pool0 and assigns it to the I/O group io_grp0. Its preferred node is node 1.

Example 7-2 The mkvdisk command

```
IBM_Storwize:ITSO:superuser>mkvdisk -mdiskgrp Pool0 -iogrp io_grp0 -size 10 -unit gb -name
Tiger
Virtual Disk, id [8], successfully created
```

To verify the results, use the **lsvdisk** command providing the volume ID as the command parameter, as shown in Example 7-3.

Example 7-3 The lsvdisk command

```
IBM_Storwize:ITSO: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.

7.8.3 Creating a thin-provisioned volume

Example 7-4 shows an example of creating a thin-provisioned volume. This 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 fully allocated.

-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 7-4 Using the command mkvdisk

```
IBM_Storwize:ITSO: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. The volume belongs to the storage pool that is named `Site1_Pool` and is owned by input/output (I/O) Group `io_grp0`. The real capacity automatically expands 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 **disk_size** option to directly specify the real physical capacity specifying its size in the units 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.

7.8.4 Creating a volume in image mode

Use 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 managed disk.

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 **-fmtdisk** 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.

Use of the **mkvdisk** command to create an image mode volume is shown in Example 7-5.

Example 7-5 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 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 volume named `Image_Volume_A` with type image, as shown in Example 7-6.

Example 7-6 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
```

7.8.5 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, use 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 the how to create a new copy of a volume in a different storage pool. As shown in Example 7-7, the volume initially has a single copy with `copy_id 0` that is provisioned in pool `Pool0`.

Example 7-7 The lsvdisk command

```
IBM_Storwize:ITSO: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 Pool0
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 Pool0
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 7-8 shows adding the second volume copy by using the **addvdiskcopy** command.

Example 7-8 The addvdiskcopy command

```
IBM_Storwize:ITSO: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 using the **lsvdisksyncprogress** command.

As shown in Example 7-9 on page 296, the first time that the status is checked, the synchronization progress is at 48%, and the estimated completion time is `161026203918`. The estimated completion time is displayed in the YYMMDDHHMMSS format. In our example, it is 2016, Oct-26 20:39:18. When the command is run again, the progress status is at 100%, and the synchronization is complete.

Example 7-9 Synchronization

```
IBM_Storwize:ITSO:superuser>lsvdisksyncprogress
vdisk_id vdisk_name copy_id progress estimated_completion_time
2        vdisk0     1       0        171018232305
IBM_Storwize:ITSO:superuser>lsvdisksyncprogress
vdisk_id vdisk_name copy_id progress estimated_completion_time
2        vdisk0     1       100
```

As shown in Example 7-10, the new volume copy (`copy_id 1`) was added and appears in the output of the **`lsvdisk`** command.

Example 7-10 The lsvdisk command

```
IBM_Storwize:ITSO: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 6005076400F580049800000000000004
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 fully allocated 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 7-11, the volume name is changed from `VOL_NO_MIRROR` to `VOL_WITH_MIRROR`.

Example 7-11 Volume name changes

```
IBM_Storwize:ITSO:superuser>chvdisk -name VOL_WITH_MIRROR VOL_NO_MIRROR
IBM_Storwize:ITSO:superuser>
```

Using -autodelete flag to migrate a volume

This section shows the use of 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 7-12 shows a shortened **`lsvdisk`** output for an uncompressed volume with a single volume copy.

Example 7-12 An uncompressed volume

```
IBM_Storwize:ITSO: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 7-13 adds a compressed copy with the **-autodelete** flag set.

Example 7-13 Compressed copy

```
IBM_Storwize:ITSO:superuser>addvdiskcopy -autodelete -rsize 2 -mdiskgrp 0 -compressed
UNCOMPRESSED_VOL
Vdisk [9] copy [1] successfully created
```

Example 7-14 shows the **lsvdisk** output with another compressed volume (copy 1) and volume copy 0 being set to **auto_delete yes**.

Example 7-14 The lsvdisk command output

```
IBM_Storwize:ITSO: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 using the **lsvdisksyncprogress** command.

7.8.6 Splitting a mirrored volume

The use of 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 7-15 shows the **splitvdiskcopy** command, which is used to split a mirrored volume. It creates a volume named SPLIT_VOL from copy with ID 1 of the volume named VOLUME_WITH_MIRRORED_COPY.

Example 7-15 Split volume

```
IBM_Storwize:ITSO: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 7-16, the new volume is created as an independent volume.

Example 7-16 The lsvdisk command

```
IBM_Storwize:ITSO: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
```

7.8.7 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 is unable to 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.

7.8.8 Deleting a volume

Use the **rmvdisk** command to delete a volume. 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 remote copy, 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. Use of **-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 using the image-mode volume (crash-consistent file system). If the **-force** flag is used, consistency is not ensured; that is, data 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 7-17 shows how to use the **rmvdisk** command to delete a volume from your IBM Spectrum Virtualize configuration.

Example 7-17 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 7-18.

Example 7-18 The rmvdisk -force command

```
IBM_2145:ITSO_CLUSTER:superuser>rmvdisk -force volume_A
```

7.8.9 Volume delete 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.

Use the **chsystem** command to set the time interval for which the volume must be idle before it can be deleted from the system. This setting affects the following commands:

- `rmvdisk`
- `rmvolume`
- `rmvdiskcopy`
- `rmvdiskhostmap`
- `rmmdiskgrp`
- `rmhostiogrp`
- `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, issue 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 only be deleted if they were inactive for over 60 minutes.

To disable volume protection, issue the following command:

```
svctask chsystem -vdiskprotectionenabled no
```

7.8.10 Expanding a volume

Expanding a volume presents a larger capacity disk to your operating system. Although this expansion can be easily performed by using IBM Spectrum Virtualize, you must ensure that your operating system supports expansion before this function is used.

Assuming that your operating system supports expansion, you can use the **expandvdisksize** command to increase the capacity of a volume, as shown in Example 7-19.

Example 7-19 The expandvdisksize command

```
IBM_2145:ITSO_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 7-20 on page 304. 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 7-20 The lsvdisk command

```
IBM_Storwize:ITSO:superuser>lsvdisk 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:ITSO:superuser>expandvdisksize -rsize 5 -unit gb volume_B
IBM_Storwize:ITSO:superuser>lsvdisk 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.
```

7.8.11 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 specified. 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 highly available 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 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.

7.8.12 Mapping a volume to a host

Use the **mkvdiskhostmap** command to map a volume to a host. 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 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.

Example 7-21 shows how to map volumes `volume_B` and `volume_C` to defined host `Almaden by` using **`mkvdiskhostmap`** command.

Example 7-21 The mkvdiskhostmap command

```
IBM_Storwize:ITSO:superuser>mkvdiskhostmap -host Almaden volume_B
Virtual Disk to Host map, id [0], successfully created
IBM_Storwize:ITSO:superuser>mkvdiskhostmap -host Almaden volume_C
Virtual Disk to Host map, id [1], successfully created
```

Example 7-22 shows the output of the **`lshostvdiskmap`** command, which shows that the volumes are mapped to the host.

Example 7-22 The lshostvdiskmap -delim command

```
IBM_2145:ITSO_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 current volume 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, use the **`mkvolumehostclustermap`** command to map a volume to a host cluster instead (see Example 7-23).

Example 7-23 The mkvolumehostclustermap command

```
BM_Storwize:ITSO:superuser>mkvolumehostclustermap -hostcluster vmware_cluster
UNCOMPRESSED_VOL
Volume to Host Cluster map, id [0], successfully created
```

7.8.13 Listing volumes mapped to the host

Use the **lshostvdiskmap** command to show the volumes that are mapped to the specific host, as shown in Example 7-24.

Example 7-24 The lshostvdiskmap command

```
IBM_2145:ITSO_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 host name is specified as **lshostvdiskmap** command, it returns all defined host-to-volume mappings.

> **Specifying the flag before the host name:** Although the **-delim** flag normally comes at the end of the command string, you must specify this flag before the host name 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 use the **lshostclustervolumemap** command to show the volumes that are mapped to a specific host cluster, as shown in Example 7-25.

Example 7-25 The lshostclustervolumemap command

```
IBM_Storwize:ITSO: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
```

7.8.14 Listing hosts mapped to the volume

To identify the hosts to which a specific volume was mapped, issue the **lsvdiskhostmap** command, as shown in Example 7-26.

Example 7-26 The lsvdiskhostmap command

```
IBM_2145:ITSO_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.

7.8.15 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. Use the **rmvdiskhostmap** command to unmap a volume from a host, as shown in Example 7-27.

Example 7-27 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 use the **rmvolumehostclustermap** command to delete a volume mapping from a host cluster, as shown in Example 7-28.

Example 7-28 The rmvolumehostclustermap command

```
IBM_Storwize:ITSO:superuser>rmvolumehostclustermap -hostcluster vmware_cluster
UNCOMPRESSED_VOL
```

This command unmaps the volume called UNCOMPRESSED_VOL from the host cluster called vmware_cluster.

Note: Removing a volume mapped to the host makes the volume unavailable for I/O operations. Ensure that the host is prepared for this before removing a volume mapping.

7.8.16 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 9, "Storage migration" on page 385.

Important: After migration is started, it continues until completion 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 7-29, 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 7-29 moves volume_C to the storage pool named STGPool_DS5000-1.

Example 7-29 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 7-30.

Example 7-30 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 finished.

7.8.17 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 7-31 shows the use of 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 7-31 The migratetoimage command

```
IBM_2145:ITSO_CLUSTER:superuser>migratetoimage -vdisk volume_A -mdisk mdisk10 -mdiskgrp
STGPool_IMAGE
```

7.8.18 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 units are MBs.

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 assigned to the volume. Use the following parameters to change volume size:

- For a fully 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 extent, or extents from those sets 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 they must first 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 a logical volume manager), 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 issuing the **svcinfo lsvdisk -bytes vdiskname** command.

Shrink the volume by the required amount by issuing the **shrinkvdisksize -size disk_size -unit b | kb | mb | gb | tb | pb vdisk_name | vdisk_id** command.

Example 7-32 shows use of **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 7-32 The shrinkvdisksize command

```
IBM_2145:ITSO_CLUSTER:superuser>shrinkvdisksize -size 44 -unit gb volume_D
```

7.8.19 Listing volumes using the MDisk

Use the **lsmdiskmember** command to identify which volumes use space on the specified MDisk. Example 7-33 displays list of volume IDs of all volume copies that use mdisk8. To correlate the IDs that are displayed in this output to volume names, use the **lsvdisk** command.

Example 7-33 The lsmdiskmember command

```
IBM_2145:ITSO_CLUSTER:superuser>lsmdiskmember mdisk8
id copy_id
24 0
27 0
```

7.8.20 Listing MDisks used by the volume

Use the **lsvdiskmember** command to list MDisks that supply space used by the specified volume. Example 7-34 lists MDisk IDs of all MDisks used by volume with ID 0.

Example 7-34 The lsvdiskmember command

```
IBM_2145:ITSO_CLUSTER:superuser>lsvdiskmember 0
id
4
5
6
7
```

If you want to know more about these MDisks, you can run the **lsmdisk** command providing as a parameter the MDisk ID that is listed in the output of the **lsvdiskmember** command.

7.8.21 Listing volumes defined in the storage pool

Use the **lsvdisk -filtervalue** command to list volumes that are defined in the specified storage pool. Example 7-35 shows how to use the **lsvdisk -filtervalue** command to list all volumes that are defined in the storage pool named Pool0.

Example 7-35 The lsvdisk -filtervalue command: volumes in the pool

```
IBM_Storwize:ITSO:superuser>lsvdisk -filtervalue mdisk_grp_name=Pool0 -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,,,,,6005076400F5800498000000000
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,,,,,6005076400F5800498000000000000010,0,1,e
mpty,0,no,0,0,Pool0,no,yes,8,Tiger,
9,UNCOMPRESSED_VOL,0,io_grp0,online,0,Pool0,10.00GB,striped,,,,,6005076400F5800498000000000
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,,,,,6005076400F5800498000000000
0000F,0,1,empty,0,no,0,0,Pool0,no,yes,13,vdisk0_restore1,
```

7.8.22 Listing storage pools in which a volume has its extents

Use the **lsvdisk** command to show to which storage pool a specific volume belongs, as shown in Example 7-36.

Example 7-36 The lsvdisk command: Storage pool ID and name

```
IBM_Storwize:ITSO: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 6005076400F580049800000000000002
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, use the **lsmdiskgrp** command, as described in Chapter 6, “Storage pools” on page 191.

7.8.23 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 data of a volume. This information is not directly available to the host; however, it might be obtained by using a sequence of queries.

The first step is to unequivocally map a logical device seen by the host to a volume that is presented by the storage system. The best volume characteristics for this purpose is the volume ID. This ID is available to the operating system 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 operating system. After the volume ID is known, it can be used to identify physical location of data.

> **Note:** For sequential and image mode volumes, a volume copy is mapped to exactly one MDisk. This configuration often is not the case for stripped volumes, unless volume size is not greater than an extent size. Therefore, a single stripped volume uses multiple mDisks in a typical case.

On hosts that are running IBM System Storage Multipath Subsystem Device Driver, you can obtain the volume ID from the output of the **datapath query device** command. You see a long disk serial number for each vpath device, as shown in Example 7-37.

Example 7-37 The datapath query device command

```
DEV#:   0  DEVICE NAME: Disk1 Part0  TYPE: 2145       POLICY: OPTIMIZED
SERIAL: 60050768018301BF2800000000000005
============================================================================
Path#            Adapter/Hard Disk        State  Mode       Select     Errors
    0     Scsi Port2 Bus0/Disk1 Part0     OPEN   NORMAL         20          0
    1     Scsi Port3 Bus0/Disk1 Part0     OPEN   NORMAL       2343          0

DEV#:   1  DEVICE NAME: Disk2 Part0  TYPE: 2145       POLICY: OPTIMIZED
SERIAL: 60050768018301BF2800000000000004
============================================================================
Path#            Adapter/Hard Disk        State  Mode       Select     Errors
    0     Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL       2335          0
    1     Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL          0          0

DEV#:   2  DEVICE NAME: Disk3 Part0  TYPE: 2145       POLICY: OPTIMIZED
SERIAL: 60050768018301BF2800000000000006
============================================================================
Path#            Adapter/Hard Disk        State  Mode       Select     Errors
    0     Scsi Port2 Bus0/Disk3 Part0     OPEN   NORMAL       2331          0
    1     Scsi Port3 Bus0/Disk3 Part0     OPEN   NORMAL          0          0
```

> **State:** In Example 7-38, the state of each path is `OPEN`. Sometimes, the state is `CLOSED`. This state does not necessarily indicate a problem because it might be a result of the path's processing stage.

On a Linux host running native multipath driver, you can use the output of command **multipath -ll**, as shown in Example 7-38.

Example 7-38 Volume ID as returned by multipath -ll

```
mpath1 (360050768018301BF2800000000000004)  IBM,2145
[size=2.0G][features=0][hwhandler=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 sdal 66:80  [active][ready]
```

Note: Volume ID shown in the output of `multipath -ll` is generated by Linux `scsi_id`. For systems that provide the VPD via page 0x83 (such as Spectrum Virtualize devices), the ID that is obtained from the VPD page is prefixed with number 3, which is the Network Address Authority (NAA) type identifier. Therefore, the volume NAA identifier (that is, volume ID obtained via the SCSI INQUIRY command) starts at the second displayed digit. In Example 7-38, the volume ID starts with digit 6.

After you know the volume ID, complete the following steps:

1. Run the **lshostvdiskmap** command to list volumes mapped to the host. Example 7-39 lists volumes mapped to host `Almaden`.

   Example 7-39 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,60050768018301BF2800000000000004
   2,Almaden,2,28,volume_C,60050768018301BF2800000000000006
   ```

   Look for the vDisk UID that matches volume UID that was identified and note the volume name (or ID) for a volume with this UID.

2. Run the **lsvdiskmember *vdiskname*** command to list of the MDisks that contain extents allocated to the specified volume, as shown in Example 7-40.

   Example 7-40 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 MDisk IDs that were obtained in the previous step, run the **lsmdisk mdiskID** command to discover MDisk controller and LUN information. Example 7-41 shows output for MDisk 0. 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 7-41 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_WWPN 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.

8

Hosts

This chapter describes the host configuration procedures that are required to attach supported hosts to the IBM Spectrum Virtualize system. It also introduces new concepts about Host Clusters, and N-Port Virtualization ID (NPIV) support from a host's perspective.

This chapter includes the following topics:

8.1 Host attachment overview

The IBM Storwize V7000 system supports a wide range of host types (both IBM and non-IBM). This feature makes it possible to consolidate storage in an open systems environment into a common pool of storage. Then, you can use and manage the storage pool more efficiently as a single entity from a central point on 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 offered across storage systems from separate vendors
- Only one multipath driver is required for attached hosts

Hosts can be connected to Storwize V7000 system using any of the following protocols:

- Fibre Channel (FC)
- Fibre Channel over Ethernet (FCoE)
- Internet Small Computer System Interface (iSCSI)
- iSCSI Extensions over RDMA (iSER)
- Non-Volatile Memory Express (NVMe)

Hosts that connect to the Storwize V7000 system by using fabric switches that use FC or FCoE protocol must be zoned correctly, as described in 3.6, “SAN configuration planning” on page 50.

Hosts that connect to the Storwize V7000 system with iSCSI protocol must be configured correctly, as described in Chapter 3, “Planning” on page 43.

Note: Certain host operating systems can be directly connected to the Storwize V7000 system without the need for FC fabric switches. For more information, see this page of the IBM System Storage Interoperation Center (SSIC).

For load balancing and access redundancy on the host side, the use of a host multipathing driver is required in the following situations:

- Protection from fabric link failures, including port failures on the IBM Spectrum Virtualize system nodes
- Protection from a host HBA failure (if two HBAs are in use)
- Protection from fabric failures if the host is connected through two HBAs to two separate fabrics
- Provide load balancing across the host HBAs

For more information about various host operating systems and versions that are supported by IBM Storwize V7000, see this page of the IBM System Storage Interoperation Center (SSIC).

For more information about how to attach various supported host operating systems to IBM Storwize V7000, see IBM Knowledge Center.

If your host operating system is not in SSIC, you can ask an IBM representative to submit a special request for support by using the Storage Customer Opportunity REquest (SCORE) tool for evaluation (log in required).

8.2 Host clusters

IBM Storwize V7000 software supports host clusters starting with V7.7.1 and later. The host cluster allows a user to create a group of hosts to form a cluster. A cluster is treated as a single entity, which allows multiple hosts to access the same volumes.

Volumes that are mapped to a host cluster are assigned to all members of the host cluster with the same SCSI ID.

A typical user case is to define a host cluster that contains all of the WWPNs belonging to the hosts participating in a host operating system-based cluster, such as IBM PowerHA®, and Microsoft Cluster Server (MSCS).

The following new commands were added to deal with host clusters:

- `lshostcluster`
- `lshostclustermember`
- `lshostclustervolumemap`
- `mkhost` (modified to put host in a host cluster on creation)
- `rmhostclustermember`
- `rmhostcluster`
- `rmvolumehostclustermap`

8.3 N-Port Virtualization ID support

The usage model for the Storwize V7000 is based on a two-way active/active node model. This is a pair of distinct control modules that share active/active access for any specific volume. These nodes each have their own Fibre Channel worldwide node name (WWNN). Therefore, ports that are presented from each node have a set of worldwide port names (WWPNs) that are presented to the fabric.

Traditionally, if one node fails or is removed for some reason, the paths that are presented for volumes from that node go offline. In this case, it is up to the native O/S multipathing software to fail over from using both sets of WWPN to only those that remain online. Although this process is what multipathing software is designed to do, occasionally it can be problematic, particularly if paths are not seen as coming back online for some reason.

Starting with Storwize V7000 V7.7, the system can be enabled into N_Port ID Virtualization (NPIV) mode. When NPIV mode is enabled on the Storwize V7000 system, ports do not come online until they are ready to service I/O, which improves host behavior around node unpends. In addition, path failures because of an offline node are masked from hosts and their multipathing driver do not need to perform any path recovery.

When NPIV is enabled on Storwize V7000 nodes, each physical WWPN reports up to four virtual WWPNs, as listed in Table 8-1.

Table 8-1 IBM Spectrum Virtualize NPIV Ports

NPIV port	Port description
Primary Port	This is the WWPN that communicates with backend storage and can be used for node to node traffic (local or remote).
Primary scsi Host Attach Port	This is the WWPN that communicates with hosts. It is a target port only. This is the primary port, so it is based on this local node's WWNN.

NPIV port	Port description
Failover scsi Host Attach Port	This is a standby WWPN that communicates with hosts and is only brought online if the partner node within the I/O Group goes offline. This is the same as the Primary Host Attach WWPN of the partner node.
Primary nvme Host Attach Port	This is the WWPN that communicates with hosts. It is a target port only. This is the primary port, so it is based on this local node's WWNN.
Failover nvme Host Attach Port	This is a standby WWPN that communicates with hosts and is only brought online if the partner node within the I/O Group goes offline. This is the same as the Primary Host Attach WWPN of the partner node.

Figure 8-1 shows the WWPNs that are associated with a Storwize V7000 port when NPIV is enabled.

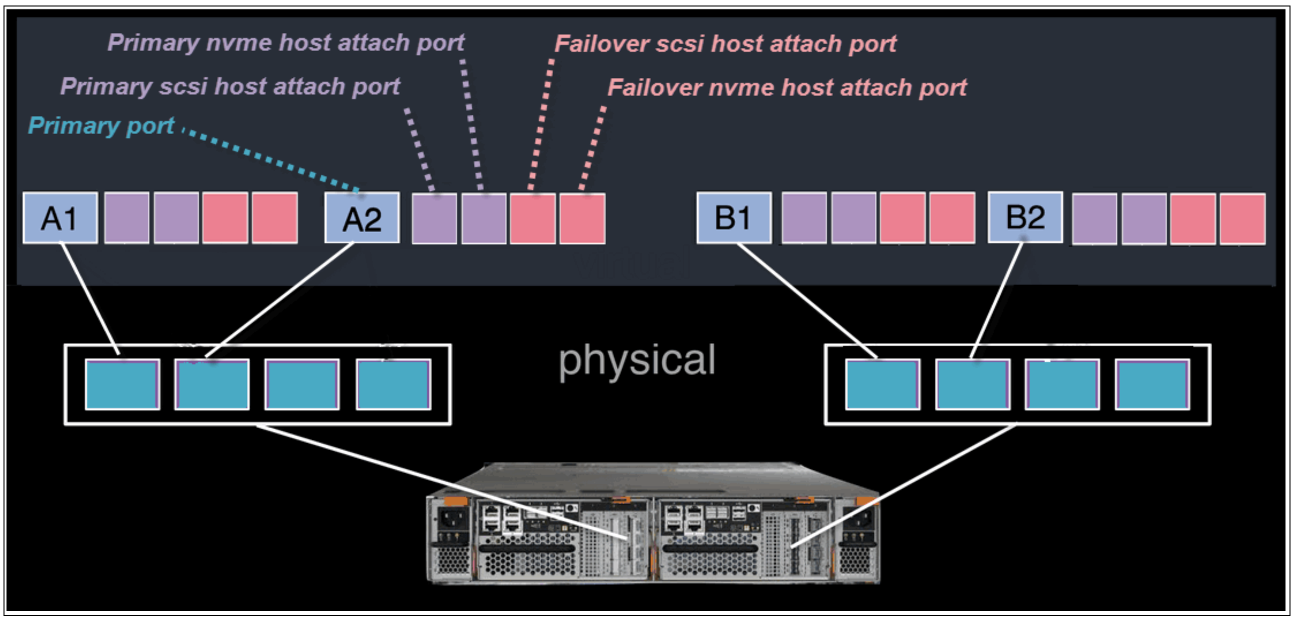


Figure 8-1 Allocation of NPIV virtual WWPN ports per physical port

The *failover host attach port* is not currently active. Figure 8-2 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 8-2 shows only two ports per node in detail, but the same 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.

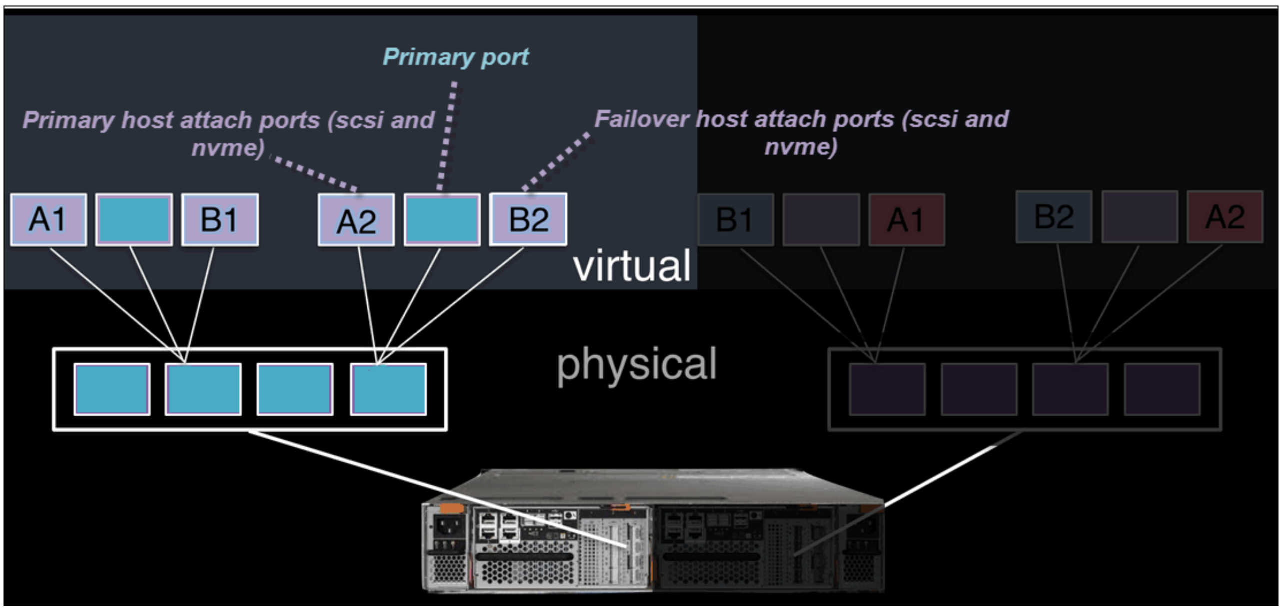


Figure 8-2 Allocation of NPIV virtual WWPN ports per physical port after a node failure

With V7.7 onwards, this process happens automatically when NPIV is enabled at a system level in Storwize V7000. This failover happens only between the two nodes in the same I/O Group.

A transitional mode allows migration of hosts from previous non-NPIV enabled systems to enabled NPIV system, allowing for a transition period as hosts are rezoned to the *primary* host attach WWPNs.

The process for enabling NPIV on a new IBM Storwize V7000 system is slightly different than on an existing system. For more information, see IBM Knowledge Center.

Note: NPIV is supported for Fibre Channel protocol only. It is *not* supported for FCoE or iSCSI protocols.

8.3.1 NPIV prerequisites

Consider the following key points for NPIV enablement:

- The IBM Storwize V7000 system must be running V7.7 or later.
- A V7.7 or later system with NPIV enabled as backend storage for a system that is earlier than V7.7 is not supported.

- Both nodes in an I/O group should have identical hardware to allow failover to work as expected.
- The Fibre Channel switches to which the Storwize V7000 ports are attached must support NPIV and have this feature enabled.

8.3.2 Enabling NPIV on a new system

New Storwize V7000 systems that are shipped with V7.7 or later should have NPIV enabled by default. If your new Storwize V7000 does not have NPIV enabled, it can be enabled by completing the following steps:

1. Run the **lsiogrp** command to list the I/O groups present in a system, as shown in Example 8-1.

Example 8-1 Listing the I/O groups in a system

```
IBM_Storwize:ITSO-V7000:superuser>lsiogrp
id name            node_count vdisk_count host_count site_id site_name
0  io_grp0         2          2           1
1  io_grp1         0          0           1
2  io_grp2         0          0           1
3  io_grp3         0          0           1
4  recovery_io_grp 0          0           0
```

2. Run the **lsiogrp <id> | grep fctargetportmode** command for the specific I/O group ID to display the `fctargetportmode` setting. If this is enabled, as shown in Example 8-2, NPIV *host target port mode* is enabled.

Example 8-2 Checking NPIV mode with the fctargetportmode field

```
IBM_Storwize:ITSO-V7000:superuser>lsiogrp 0 | grep fctargetportmode
fctargetportmode enabled
IBM_Storwize:ITSO-V7000:superuser>
```

3. The virtual WWPNs can be listed using the **lstargetportfc** command, as shown in Example 8-3. Look for the `host_io_permitted` and `virtualized` columns to be `yes`, meaning the WWPN in those lines is a primary host attach port and should be used when zoning the hosts to the Storwize V7000.

Example 8-3 Listing the virtual WWPNs

IBM_Storwize:ITSO-V7000: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	no	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	no	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	no	no	scsi
8	**500507680112A288**	**500507680100A288**	**3**	**1**	**1**	**010501**	**yes**	**yes**	**scsi**
9	**500507680114A288**	**500507680100A288**	**3**	**1**	**1**	**010502**	**yes**	**yes**	**nvme**
10	500507680120A288	500507680100A288	4	1	1	010A00	no	no	scsi
11	**500507680122A288**	**500507680100A288**	**4**	**1**	**1**	**010A02**	**yes**	**yes**	**scsi**
12	**500507680124A288**	**500507680100A288**	**4**	**1**	**1**	**010A01**	**yes**	**yes**	**nvme**
49	500507680C110009	500507680C000009	1	2	2	010500	no	no	scsi
50	**500507680C150009**	**500507680C000009**	**1**	**2**	**2**	**010502**	**yes**	**yes**	**scsi**
51	**500507680C190009**	**500507680C000009**	**1**	**2**	**2**	**010501**	**yes**	**yes**	**nvme**

```
52 500507680C120009 500507680C000009 2 2 2 010400 no  no  scsi
53 500507680C160009 500507680C000009 2 2 2 010401 yes yes scsi
54 500507680C1A0009 500507680C000009 2 2 2 010402 yes yes nvme
55 500507680C130009 500507680C000009 3 2 2 010900 no  no  scsi
56 500507680C170009 500507680C000009 3 2 2 010902 yes yes scsi
57 500507680C1B0009 500507680C000009 3 2 2 010901 yes yes nvme
58 500507680C140009 500507680C000009 4 2 2 010900 no  no  scsi
59 500507680C180009 500507680C000009 4 2 2 010901 yes yes scsi
60 500507680C1C0009 500507680C000009 4 2 2 010902 yes yes nvme
IBM_Storwize:ITSO-V7000:superuser>
```

4. NPIV enablement can be verified by checking the `fctargetportmode` field, as shown in Example 8-4.

Example 8-4 NPIV enablement verification

```
IBM_Storwize:ITSO-V7000:superuser>lsiogrp 0
id 0
name io_grp0
node_count 2
vdisk_count 2
host_count 2
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 40.0MB
raid_free_memory 38.8MB
maintenance no
compression_active yes
accessible_vdisk_count 2
compression_supported yes
max_enclosures 10
encryption_supported no
flash_copy_maximum_memory 552.0MB
site_id
site_name
fctargetportmode enabled
compression_total_memory 2047.9MB
```

You can now configure your zones for hosts by using the `primary host attach ports` (virtual WWPNs) of the Storwize V7000 ports, as shown in **bold** in the output of Example 8-3 on page 322.

8.3.3 Enabling NPIV on an existing system

When IBM Spectrum Virtualize systems that are running code earlier than V7.7.1 are upgraded to V7.7.1 or later, the NPIV feature is not turned on by default because it might require changes to host-side zoning.

Enabling NPIV on a system requires that you complete the following steps after you meet the prerequisites:

1. Audit your SAN fabric layout and zoning rules because NPIV has stricter requirements. Ensure that equivalent ports are on the same fabric and in the same zone.

2. 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.

   For more information, see IBM Knowledge Center.

3. Run the **`lstargetportfc`** command to discover the `primary host attach` WWPNs (virtual WWPNs), as shown in **bold** in Example 8-5.

Example 8-5 Using the lstargetportfc command to get primary host WWPNs (virtual WWPNs)

IBM_Storwize:ITSO-V7000: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**		**000000**	**no**	**yes**	**scsi**
3	**500507680144A288**	**500507680100A288**	**1**	**1**		**000000**	**no**	**yes**	**nvme**
4	500507680130A288	500507680100A288	2	1	1	010400	yes	no	scsi
5	**500507680132A288**	**500507680100A288**	**2**	**1**		**000000**	**no**	**yes**	**scsi**
6	**500507680134A288**	**500507680100A288**	**2**	**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**
49	500507680C110009	500507680C000009	1	2	2	010500	yes	no	scsi
50	**500507680C150009**	**500507680C000009**	**1**	**2**		**000000**	**no**	**yes**	**scsi**
51	**500507680C190009**	**500507680C000009**	**1**	**2**		**000000**	**no**	**yes**	**nvme**
52	500507680C120009	500507680C000009	2	2	2	010400	yes	no	scsi
53	**500507680C160009**	**500507680C000009**	**2**	**2**		**000000**	**no**	**yes**	**scsi**
54	**500507680C1A0009**	**500507680C000009**	**2**	**2**		**000000**	**no**	**yes**	**nvme**
55	500507680C130009	500507680C000009	3	2	2	010900	yes	no	scsi
56	**500507680C170009**	**500507680C000009**	**3**	**2**		**000000**	**no**	**yes**	**scsi**
57	**500507680C1B0009**	**500507680C000009**	**3**	**2**		**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**

4. Enable transitional mode for NPIV on the Storwize V7000 system (see Example 8-6).

Example 8-6 NPIV in transitional mode

```
IBM_Storwize:ITSO-V7000:superuser>chiogrp -fctargetportmode transitional 0
IBM_Storwize:ITSO-V7000:superuser>lsiogrp 0 |grep fctargetportmode
fctargetportmode transitional
```

5. Ensure that the primary host attach WWPNs (virtual WWPNs) now allow host traffic, as shown in **bold** in Example 8-7.

Example 8-7 Host attach WWPNs (virtual WWPNs) permitting host traffic

```
IBM_2145:ITSO-V7000: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
10 500507680120A288 500507680100A288 4       1              1               010A00  yes               no          scsi
11 500507680122A288 500507680100A288 4       1              1               010A02  yes               yes         scsi
12 500507680124A288 500507680100A288 4       1              1               010A01  yes               yes         nvme
49 500507680C110009 500507680C000009 1       2              2               010500  yes               no          scsi
50 500507680C150009 500507680C000009 1       2              2               010502  yes               yes         scsi
51 500507680C190009 500507680C000009 1       2              2               010501  yes               yes         nvme
52 500507680C120009 500507680C000009 2       2              2               010400  yes               no          scsi
53 500507680C160009 500507680C000009 2       2              2               010401  yes               yes         scsi
54 500507680C1A0009 500507680C000009 2       2              2               010402  yes               yes         nvme
55 500507680C130009 500507680C000009 3       2              2               010900  yes               no          scsi
56 500507680C170009 500507680C000009 3       2              2               010902  yes               yes         scsi
57 500507680C1B0009 500507680C000009 3       2              2               010901  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
IBM_2145:ITSO-V7000:superuser>
```

6. Add the `primary host attach ports` (virtual WWPNs) to your host zones but do not remove the Storwize V7000 WWPNs that is in the zones. Example 8-8 shows a host zone to the `Primary Port WWPNs` of the Storwize V7000 nodes.

Example 8-8 Legacy host zone

```
zone: WINDOWS_HOST_01_IBM_ITSOV7000
                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 8-9 shows that we added the `primary host attach ports` (virtual WWPNs) to our example host zone to allow us to change the host without disrupting its availability.

Example 8-9 Transitional host zone

```
zone: WINDOWS_HOST_01_IBM_ITSOV7000
                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
```

7. With the transitional zoning active in your fabrics, ensure that the host is using the new NPIV ports for host I/O. Example 8-10 shows the before and after pathing for our host. Notice that the select count now increases on the new paths and has stopped on the old paths.

Example 8-10 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: 6005076400868008380000000000000000       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: 6005076400868008380000000000000000       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 entering 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 has occurred 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 operating system, rescanning of for 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 validated, change the system NPIV to `enabled` mode by entering the command that is shown in Example 8-11.

Example 8-11 Enabling the NPIV

```
IBM_2145:ITSO:superuser>chiogrp -fctargetportmode enabled 0
```

Now NPIV is enabled on the Storwize V7000 system, and you confirmed the hosts are using the virtualized WWPNs for I/O. To complete the NPIV implementation, the host zones can be amended to remove the old primary attach port WWPNs.

Example 8-12 shows our final zone with the host HBA and the Storwize V7000 virtual WWPNs.

Example 8-12 Final host zone

```
zone: WINDOWS_HOST_01_IBM_ITSOV7000
                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 V7000 system, the system prevents you from changing the `fctargetportmode` from transitional to enabled and displays the following error:
>
> `CMMVC8019E Task could interrupt IO and force flag not set.`

8.4 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

8.4.1 Creating hosts

This section describes how to create Fibre Channel and iSCSI hosts by using the Storwize V7000 GUI. It is assumed that hosts are prepared for attachment, as described in IBM Knowledge Center, and that the host WWPNs or their iSCSI initiator names are known.

For more information, see IBM Knowledge Center.

To create a host, complete the following steps:

1. Open the host configuration window by clicking **Hosts** (see Figure 8-3).

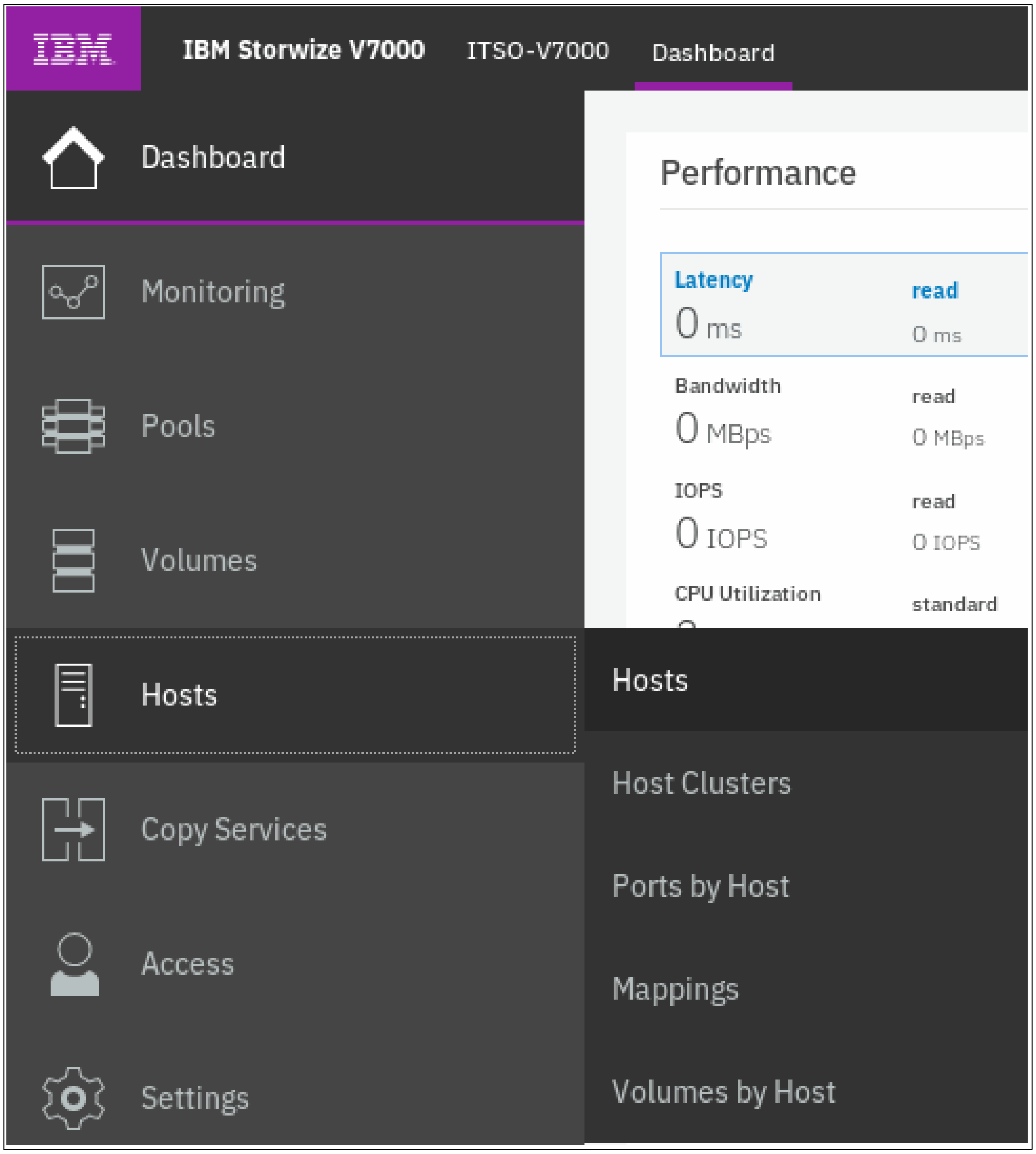


Figure 8-3 Open the host window

2. To create a host, click **Add Host**. If you want to create a Fibre Channel host, continue with "Creating Fibre Channel hosts" on page 329. To create an iSCSI host, go to "Creating iSCSI hosts" on page 331.

Creating Fibre Channel hosts

To create Fibre Channel hosts, complete the following steps:

1. Select **Fibre Channel**. The Fibre Channel configuration window opens (see Figure 8-4).

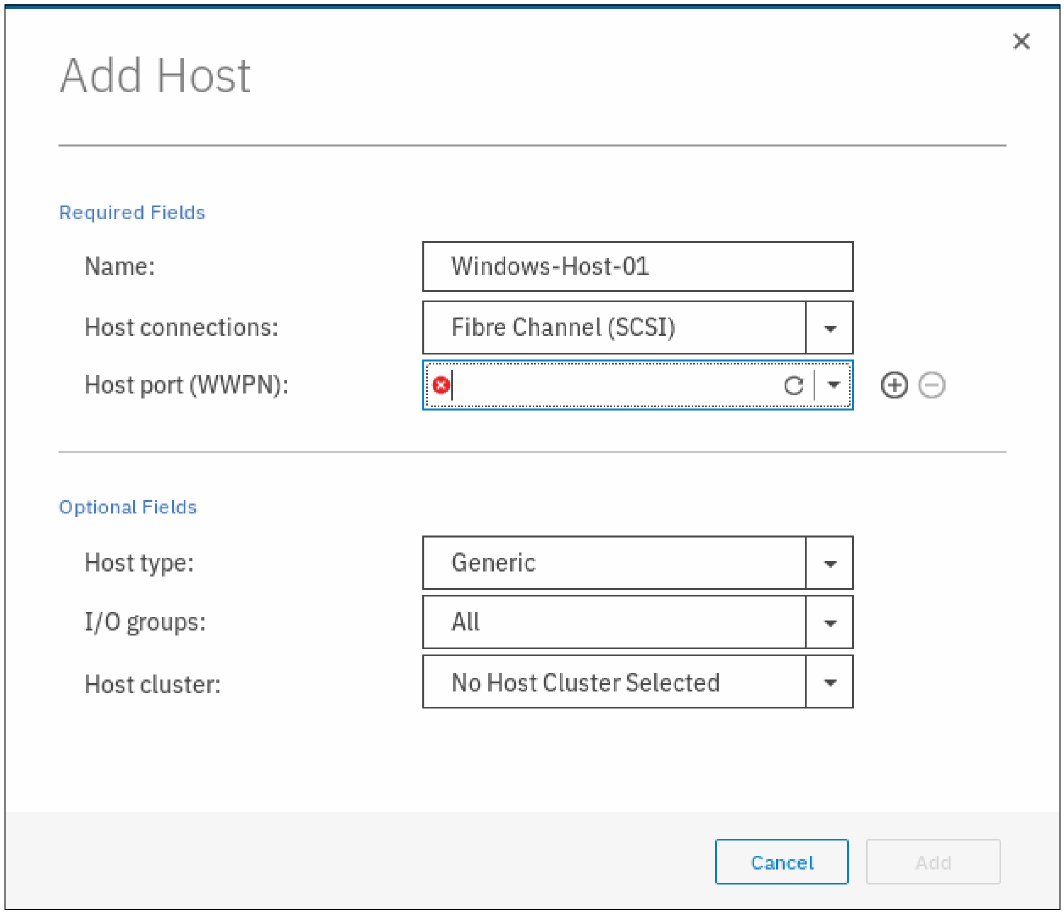


Figure 8-4 Fibre Channel host configuration

2. Enter a host name and click the **Host Port** menu to get a list of all discovered WWPNs (see Figure 8-5).

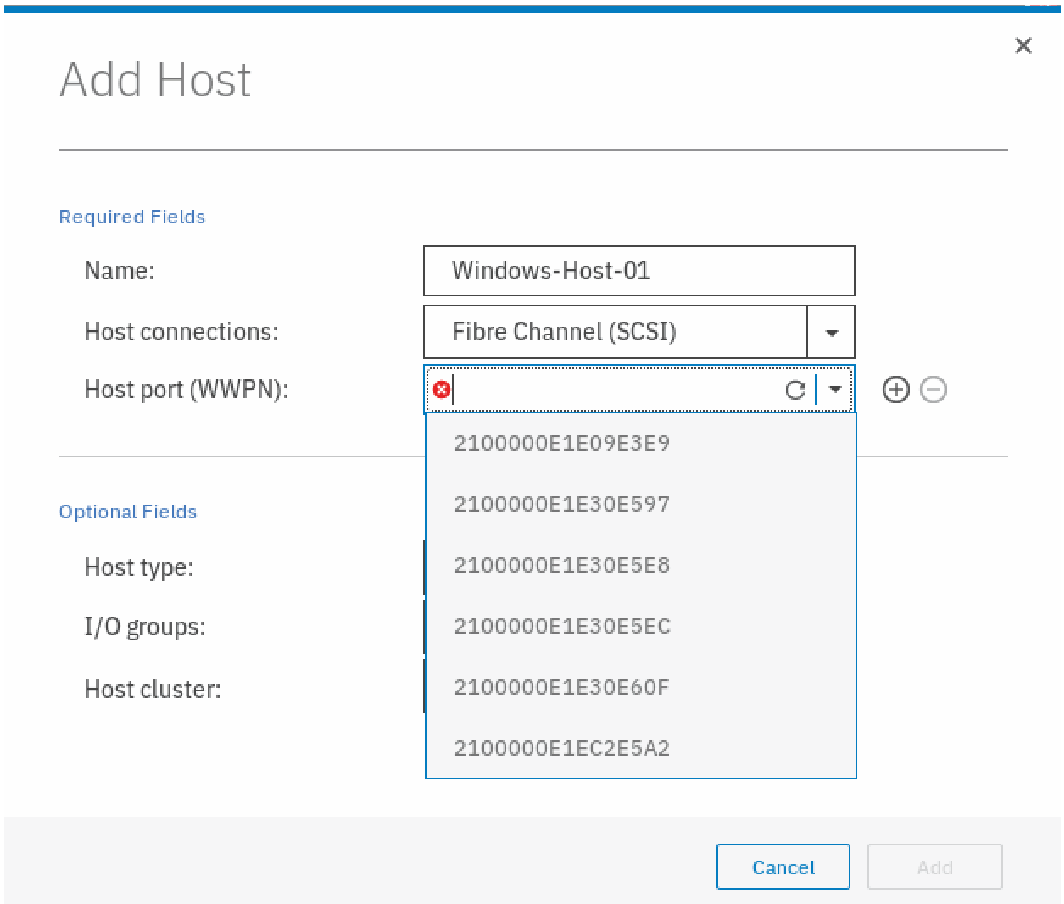


Figure 8-5 Available WWPNs

3. Select one or more WWPNs for your host. The Storwize V7000 should have the host port WWPNs available if the host is prepared, as described in IBM Knowledge Center for host attachment. If they do not appear in the list, scan for new disks as required on the respective operating system and click the **Rescan** icon in the WWPN box. If they still do not appear, check the SAN zoning and 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, click the Plus sign (**+**) 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, select the **Host type** (see Figure 8-6 on page 331). Select your host type. If your specific host type is not listed, select `generic`.

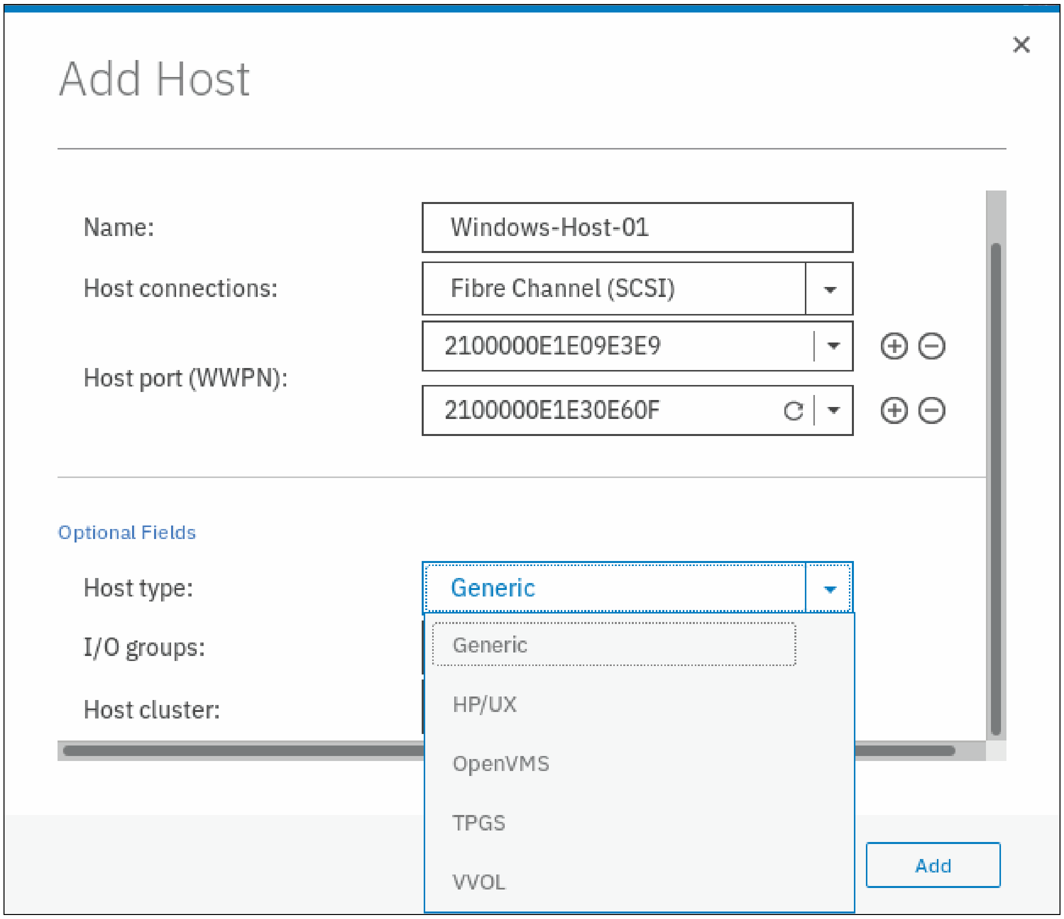


Figure 8-6 Host type selection

6. Click **Add** to create the host object.
7. Click **Close** to return to the host window. Repeat these steps for all of your Fibre Channel hosts. Figure 8-7 shows the **All Hosts** window after creating a second host.

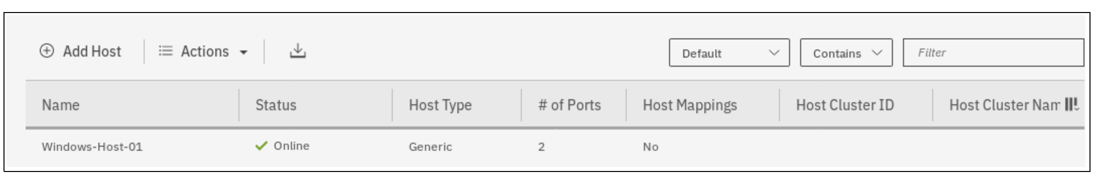


Figure 8-7 Hosts view after creating a host

After you complete the adding of Fibre Channel hosts, see Chapter 7, "Volumes" on page 241 to create volumes and map them to the created hosts.

Creating iSCSI hosts

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 a Storwize V7000 host object in the same way that you add FC WWPNs.
- Host objects can have WWPNs and IQNs.

- Standard iSCSI host connection procedures can be used to discover and configure the Storwize V7000 as an iSCSI target.
- The Storwize V7000 supports the Challenge Handshake Authentication Protocol (CHAP) authentication methods for iSCSI.
- The name iqn.1986-03.com.ibm:2076.<*cluster_name*>.<*node_name*> is the iSCSI qualified name (IQN) for a Storwize V7000 node. Because the IQN contains the clustered system name and the node name, it is important not to change these names after iSCSI is deployed.
- Each node can be given an iSCSI alias as an alternative to the IQN.

To create iSCSI hosts, complete the following steps:

1. Click **iSCSI** and the iSCSI configuration options are displayed (see Figure 8-8).

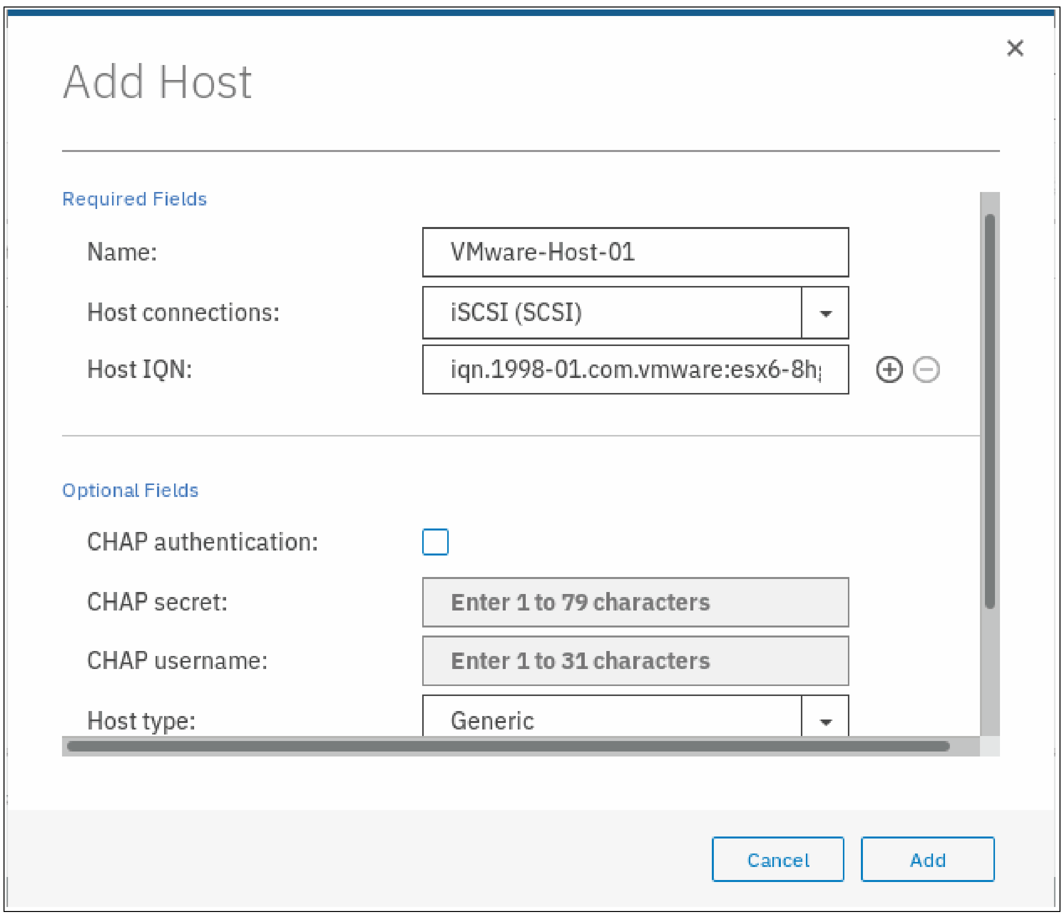


Figure 8-8 Add an iSCSI host

2. Enter a host name and the iSCSI initiator name into the **iSCSI host IQN** field. Click the plus sign (**+**) if you want to add initiator names to one host.
3. If you are connecting an HP-UX or TPGS host, click the **Host type** menu and then select the correct host type. For our ESX host, we selected **VVOL**. However, generic can be selected if you are not using VVOLs.
4. Click **Add** and then, click **Close** to complete the host object definition.

5. Repeat these steps for every iSCSI host that you want to create. Figure 8-9 shows the All Hosts window after creating two Fibre Channel hosts and one iSCSI host.

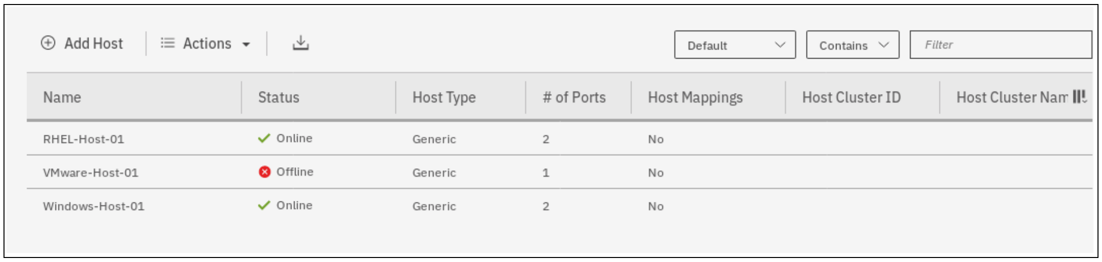


Figure 8-9 Defined Hosts list

Although the iSCSI host is now configured, the iSCSI Ethernet ports must also be configured to provide connectivity.

Complete the following steps to enable iSCSI connectivity:

1. Select **Settings** → **Network** and select the iSCSI tab (see Figure 8-10).

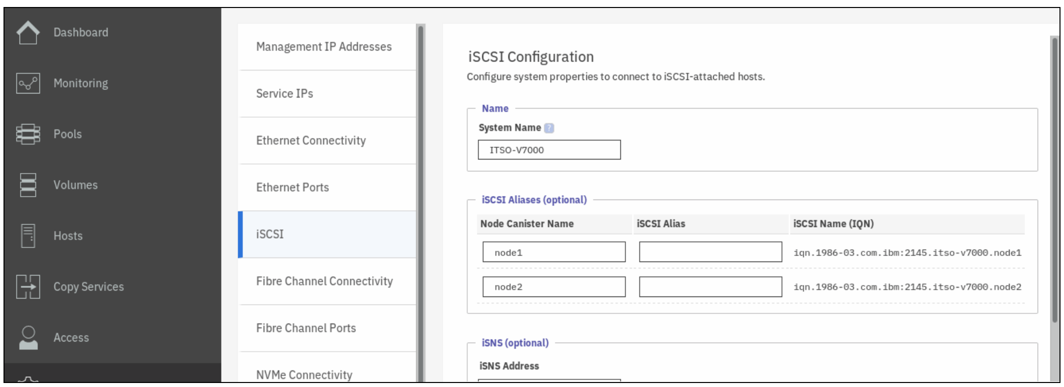


Figure 8-10 Network: iSCSI settings

2. In the iSCSI Configuration window, we can modify the system name, node names, and provide an optional iSCSI Alias for each node, if wanted (see Figure 8-11).

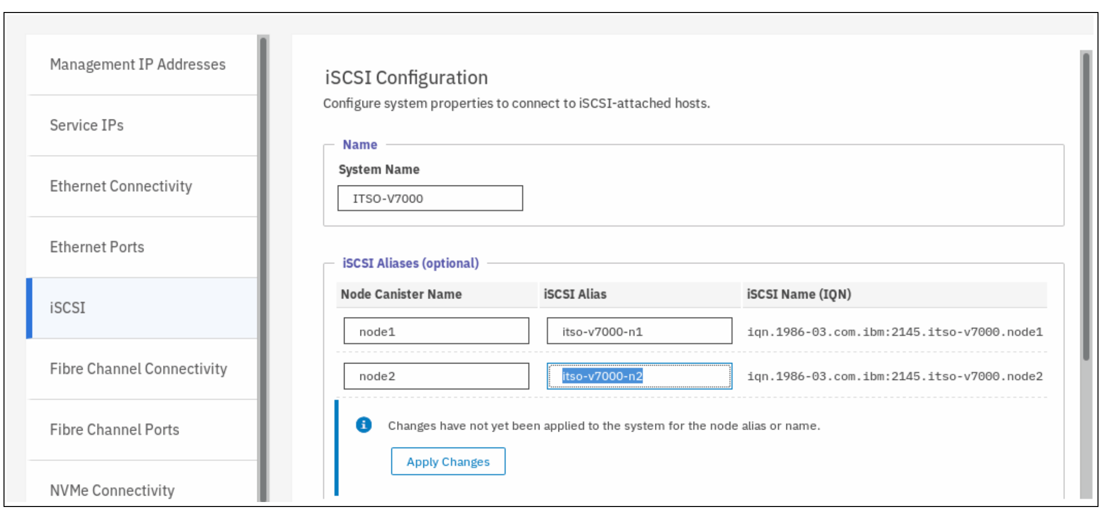


Figure 8-11 iSCSI Configuration window

3. The window displays an **Apply Changes** prompt to apply any changes you made before continuing,
4. Lower in the configuration window, you can also configure internet Storage Name Service (iSNS) addresses and CHAP if you need these 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 Storwize V7000 does not allow it to perform I/O to volumes. Also, you can assign a CHAP secret to the cluster.

5. Click the **Ethernet Ports** tab to set the iSCSI IP address for each node (see Figure 8-12).

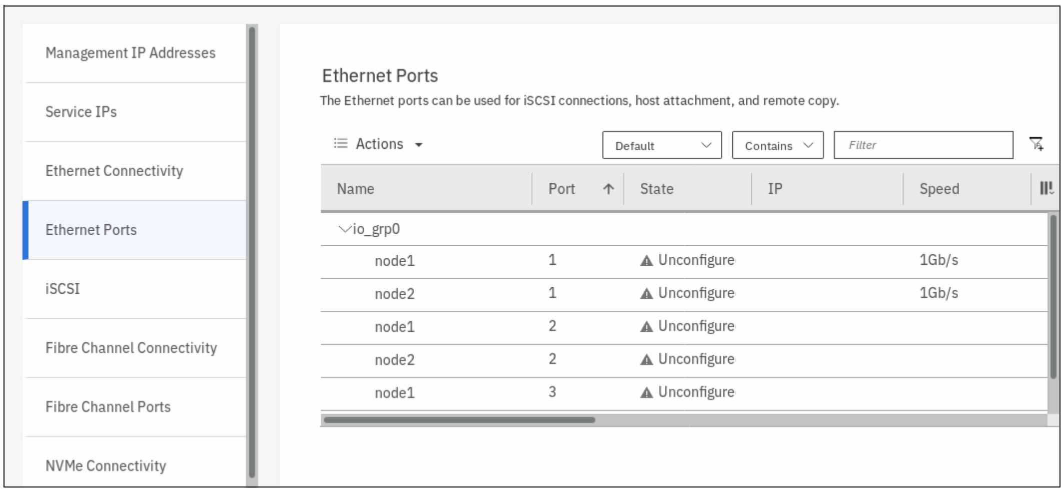


Figure 8-12 Enter an iSCSI IP address

6. Select the port that is wanted to set the iSCSI IP information. Click **Actions**, and then, **Modify IP Settings** (see Figure 8-13 on page 335).

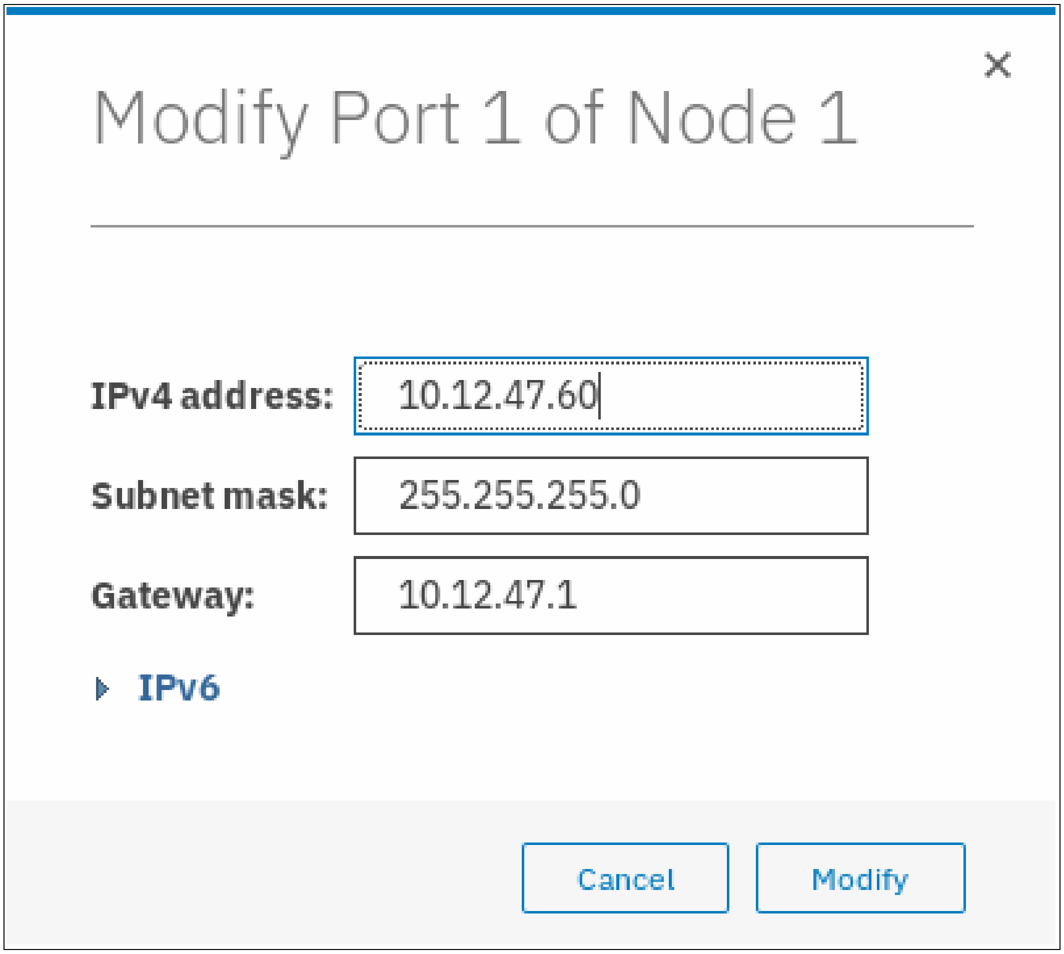


Figure 8-13 Modifying IP settings

7. After entering the IP address for a port, click **Modify** to enable the configuration. After the changes are successfully applied, click **Close.**

   You can see that iSCSI is enabled for host I/O on the required interfaces with the yes under the Host Attach column (see Figure 8-14).

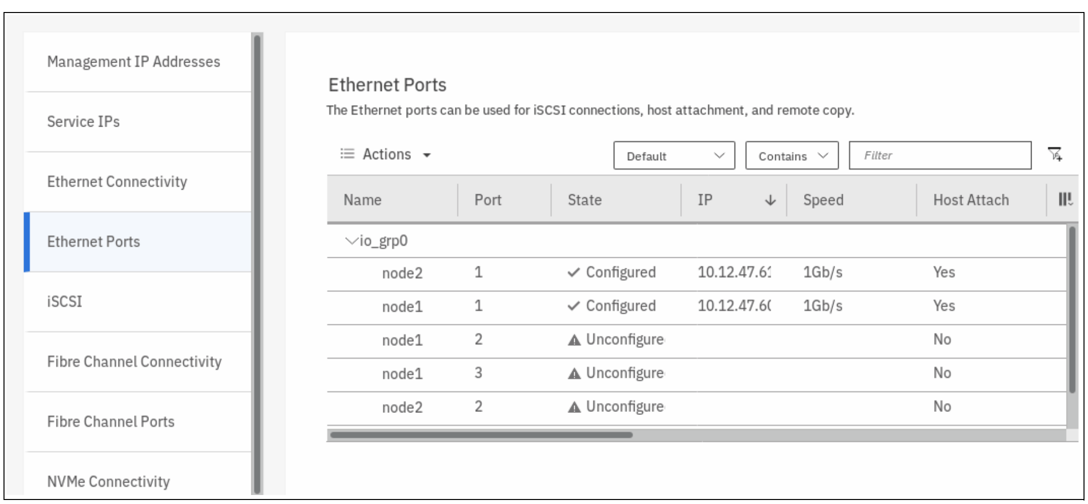


Figure 8-14 Actions menu on iSCSI ports

8. By default, iSCSI host connection is enabled after setting the IP address. To disable any interfaces you do not want to be used for host connections, click **Actions** and select **Modify iSCSI Hosts** (see Figure 8-15).

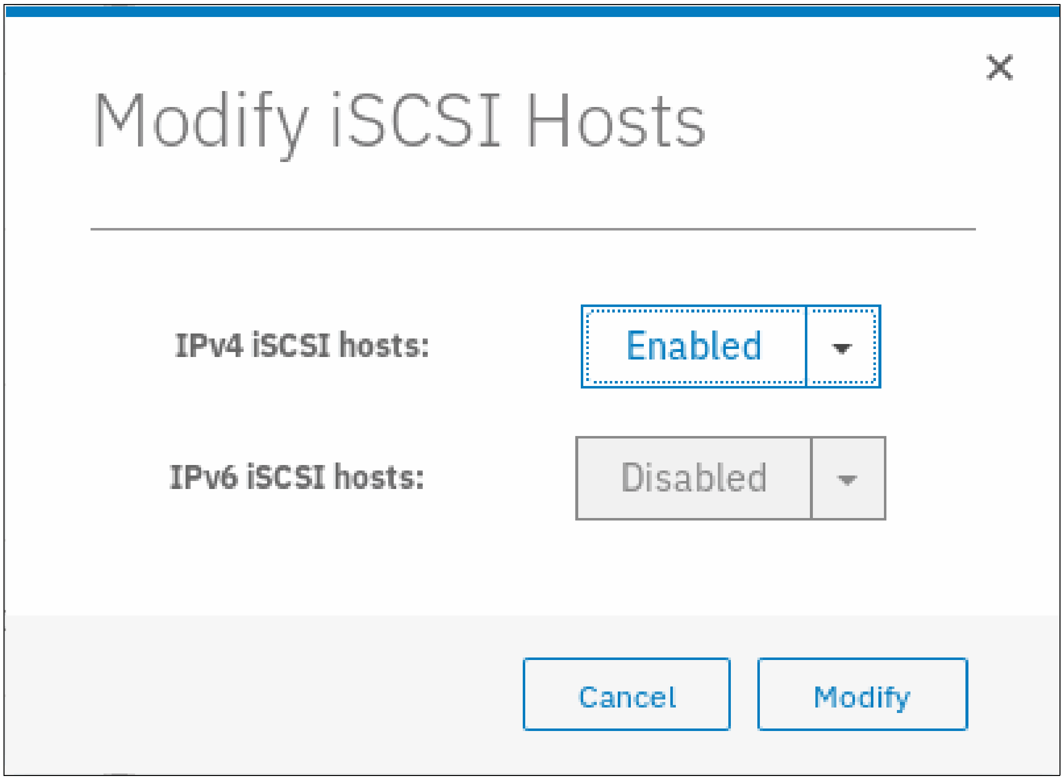


Figure 8-15 Disabling iSCSI host connections

9. It is recommended to use the iSCSI network in a separate subnet. It is also possible to set a VLAN for the iSCSI traffic. To enable the VLAN, click **Actions**, then **Modify VLAN**, as shown in Figure 8-16 on page 337.

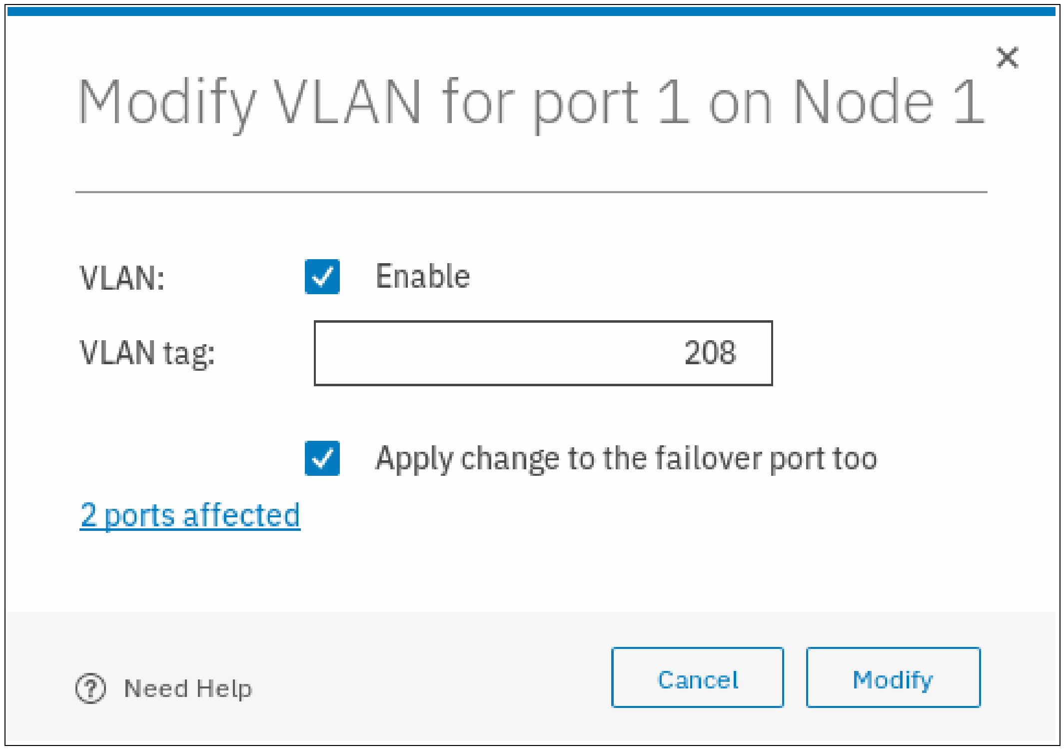


Figure 8-16 Modifying VLAN

The IBM Storwize V7000 is now configured and ready for iSCSI host use. Note the initiator IQN names of your Storwize V7000 node canisters (see Figure 8-11 on page 333) because you need them when you add storage on your host. For more information about creating volumes and mapping them to a host, see Chapter 7, “Volumes” on page 241.

Creating NVMe hosts

Running V8.2, Storwize V7000 can use NVMe hosts by using FC-NVMe. FC-NVMe uses Fibre Channel Protocol (FCP) as its underlying transport, which puts the data transfer in control of the target and transfers data direct from host memory, similar to RDMA. In addition, FC-NVMe allows for a host to send commands and data together (first burst), which eliminates the first data “read” by the target and provides better performance at distances.

For more information about NVMe, see *IBM Storage and the NVM Express Revolution*, REDP-5437.

Note: See SSIC to check whether your hosts and Storwize system are compatible.

To configure an NVMe host, complete the following steps:

1. Go to the host window, click **Add Host**. Then, on Host connections, select **NVMe**, as shown in Figure 8-17.

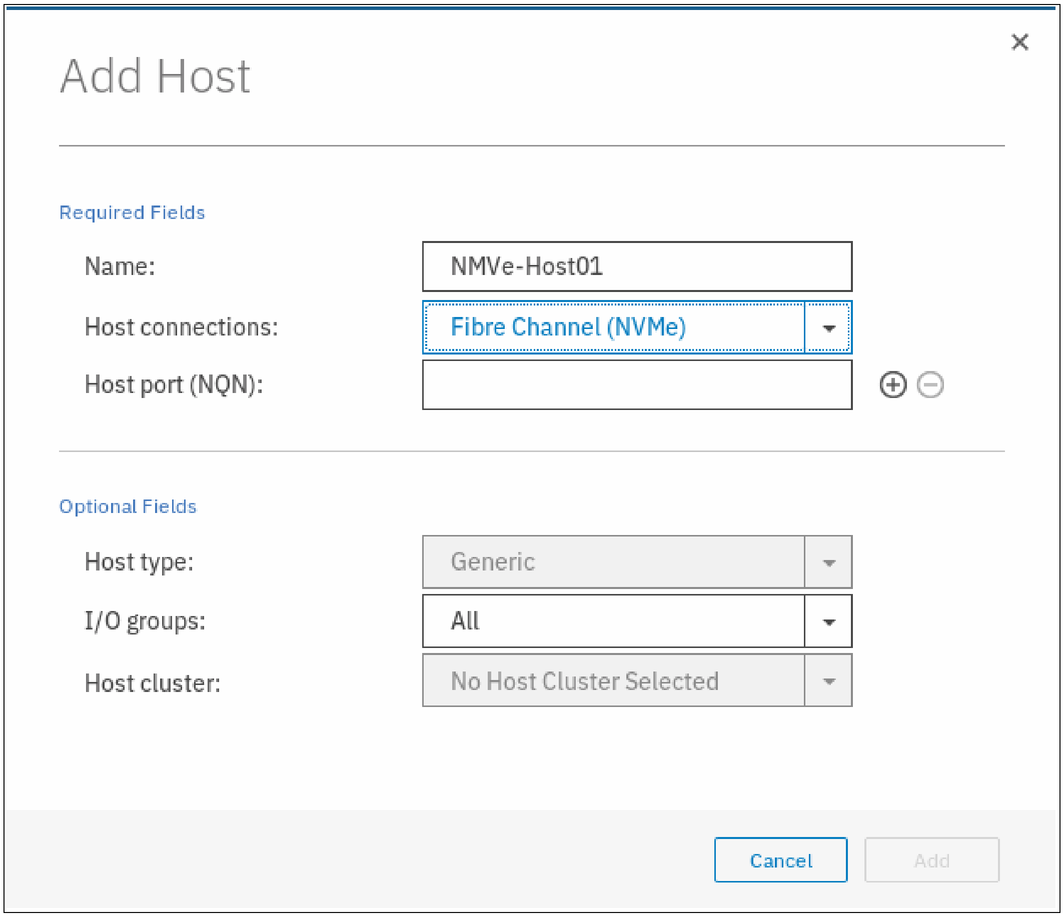


Figure 8-17 Creating NVMe host

2. Enter the host name, NVMe Qualified Name (NQN) of the host, as shown in Figure 8-18.

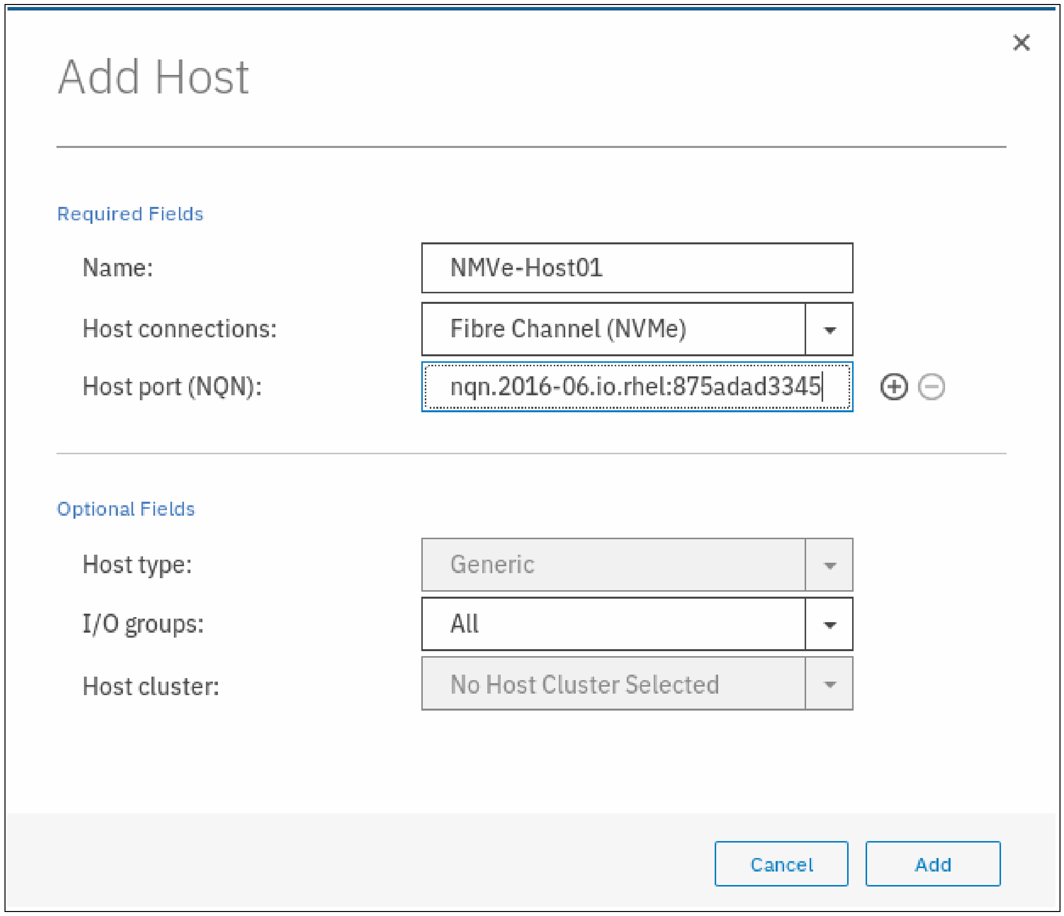


Figure 8-18 Defining NQN

3. Click **Add**. Your host appears in the defined host list.
4. The I/O Group NQN must be configured on the host. You can get the I/O Group NQN by using the **lsiogrp** command, as shown in Example 8-13.

Example 8-13 lsiogrp command

```
BM_Storwize:ITSO-V7000:superuser>lsiogrp 0
id 0
name io_grp0
node_count 2
vdisk_count 11
host_count 0
flash_copy_total_memory 20.0MB
flash_copy_free_memory 20.0MB
remote_copy_total_memory 20.0MB
remote_copy_free_memory 19.9MB
mirroring_total_memory 20.0MB
mirroring_free_memory 19.9MB
raid_total_memory 40.0MB
raid_free_memory 38.7MB
maintenance no
compression_active yes
accessible_vdisk_count 11
```

```
compression_supported yes
max_enclosures 21
encryption_supported yes
flash_copy_maximum_memory 2048.0MB
site_id
site_name
fctargetportmode enabled
compression_total_memory 11756.9MB
deduplication_supported yes
deduplication_active no
nqn nqn.1986-03.com.ibm:nvme:2145.0000010029C056C2.iogroup0
IBM_Storwize:ITSO-V7000:superuser>
```

5. You can now configure your NVMe host to use the IBM SAN Volume Controller as a target.

> **Note:** For more information about a compatibility matrix and supported hardware, see IBM Knowledge Center and the SSIC.

8.4.2 Host clusters

Storwize V7000 V7.7 introduced the concept of a *host cluster*. A host cluster allows a user to group individual hosts to form a cluster, which is treated as one entity instead of dealing with all of the hosts individually in the cluster.

The host cluster is useful for hosts that are participating in a cluster at host operating system levels. Examples are Microsoft Clustering Server, IBM PowerHA, Red Hat Cluster Suite, and VMware ESX. By defining a host cluster, a user can map one or more volumes to the host cluster object.

As a result, the volume or set of volumes gets mapped to each individual host object that is part of the host cluster. Importantly, each of the volumes gets mapped with the same SCSI ID to each host that is part of the host cluster with single command.

Although a host is part of a host cluster, 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, and another set of SCSI IDs to be used for individual 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, wanted volumes are not shared, and other volumes can be shared. For example, the boot volume of each host can be kept private, while data and application volumes can be shared.

This section describes how to create a host cluster. It is assumed that individual hosts were created as described in 8.4.3, “Advanced host administration” on page 344. Complete the following steps:

1. From the menu on the left side, select **Hosts** → **Host Clusters** (see Figure 8-19).

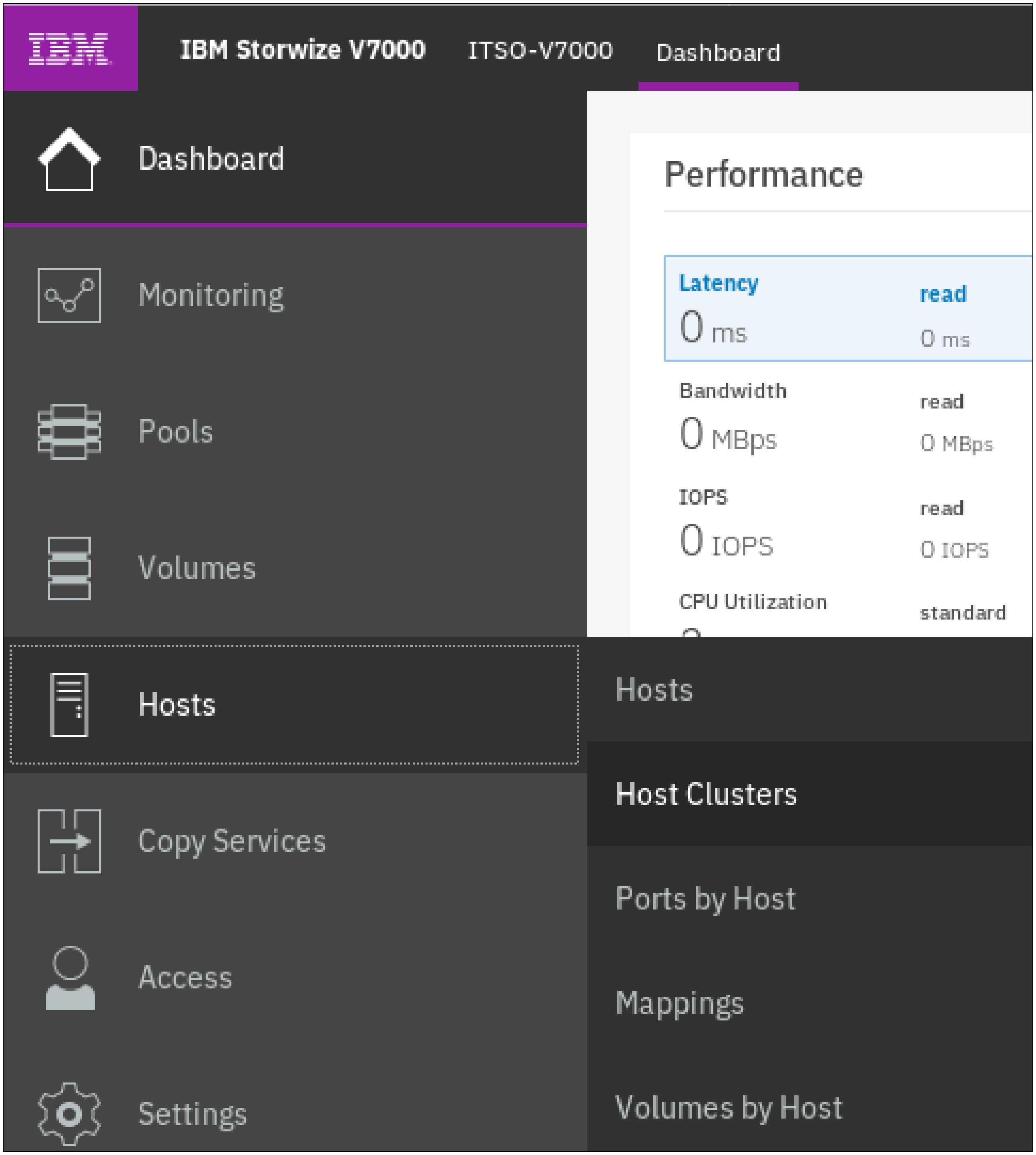


Figure 8-19 Host Clusters

2. Click **Create Host Cluster** to open the wizard that is shown in Figure 8-20.

Figure 8-20 Create host cluster

3. Enter a cluster name and select the individual nodes that you want in the cluster object, as shown in Figure 8-21.

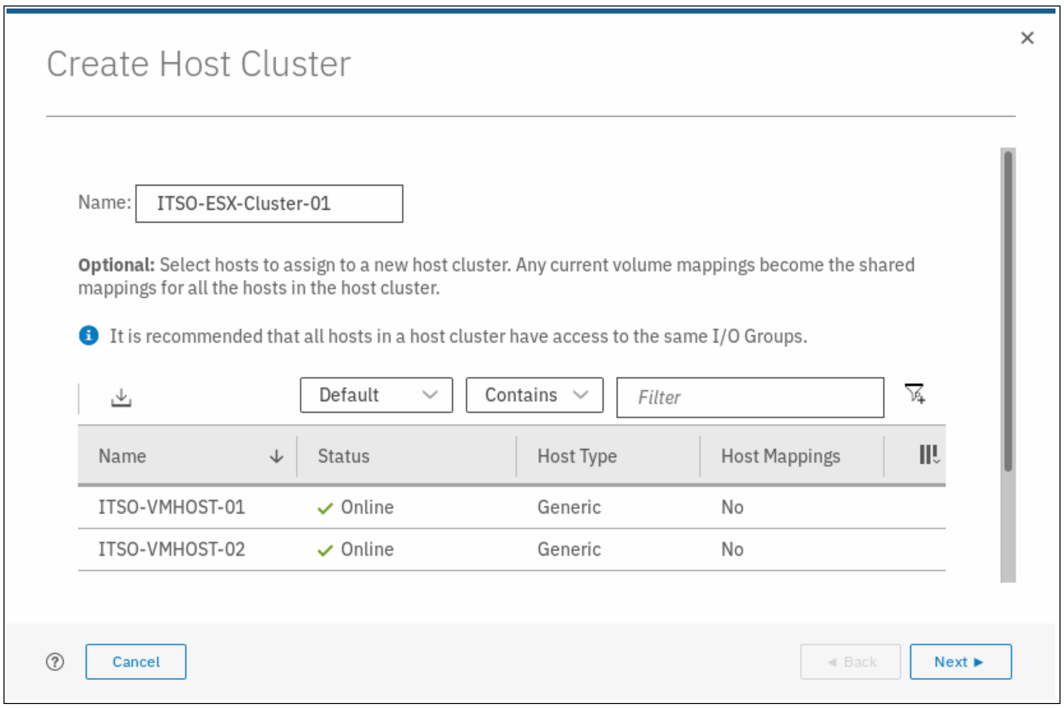


Figure 8-21 Creating a host cluster: details

4. A summary is displayed in which you can confirm that you selected the correct hosts. Click **Make Host Cluster** (see Figure 8-22).

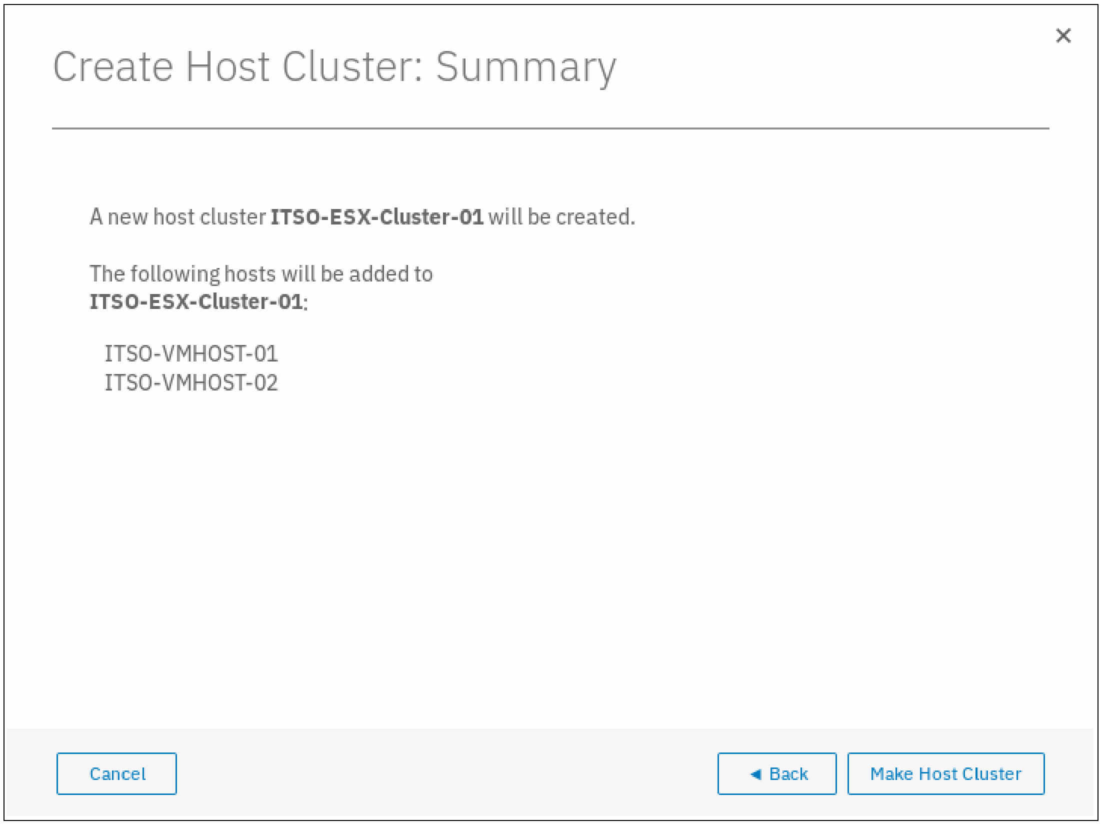


Figure 8-22 Create host cluster summary

5. After the task completes, click **Close** to return to the Host Cluster view, where you can see the cluster that you created (see Figure 8-23).

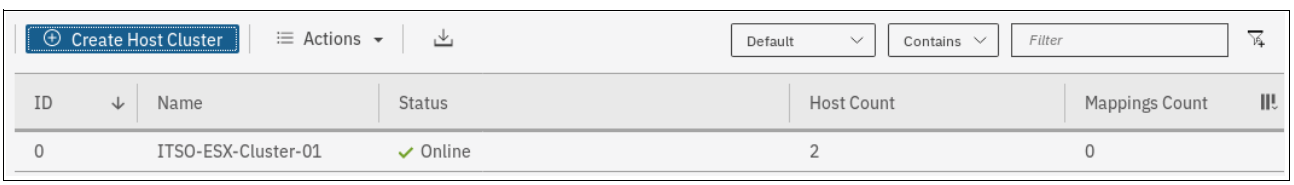


Figure 8-23 Host Cluster view

> **Note:** The host cluster status depends on its member hosts. One offline or degraded host sets the host cluster status as degraded.

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 8-24).

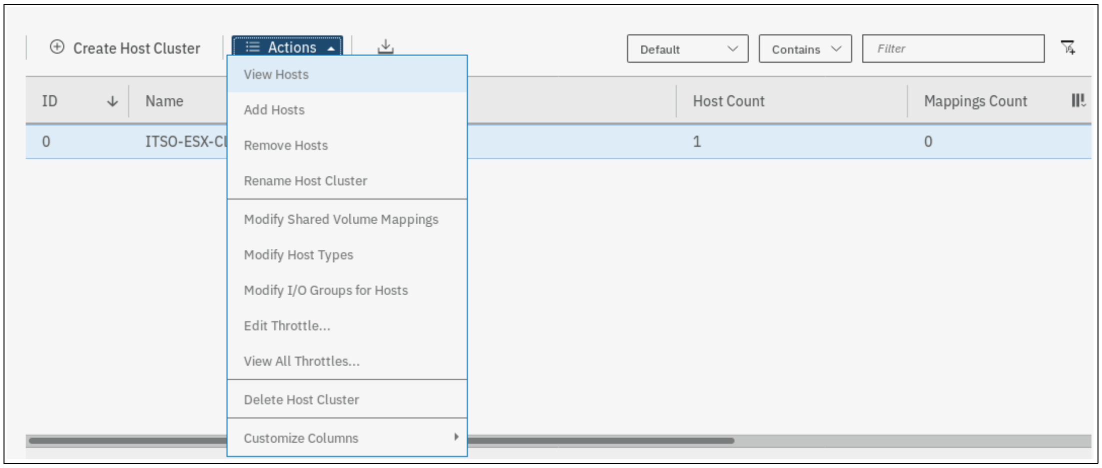


Figure 8-24 Host Clusters Actions menu

From the **Actions** menu, we can perform the following tasks:

- View Hosts status within the cluster.
- Add or Remove Hosts from the cluster.
- Rename the host cluster.
- Modify Shared Volume mappings allows you to add or remove volumes mapped to all hosts in the cluster while maintaining the same SCSI ID for all hosts.
- Modify Host Type can be used to change from generic to VVOLs, for example.
- Modify I/O Groups for Hosts is used to assign or restrict volume access to specific I/O groups.
- `Edit Throttle` is used restrict MBps or IOPS bandwidth for the Host Cluster.
- View All Throttles displays any throttling settings and allows for changing, deleting, or refining Throttle settings.

8.4.3 Advanced host administration

This section covers host administration, including topics, such as host modification, host mappings, and deleting hosts. Basic host creation by using FC and iSCSI connectivity is described in 8.4.1, “Creating hosts” on page 327.

It is assumed that hosts were created in your IBM Spectrum Virtualize GUI, and that some volumes are mapped to them. This section describes the following functions that are available in the Hosts section of the IBM Spectrum Virtualize GUI (see Figure 8-25 on page 345):

- Hosts (“Modifying Volume Mappings” on page 346)
- Ports by Host (8.4.4, “Adding and deleting host ports” on page 361)
- Host Mappings (8.4.5, “Host mappings overview” on page 371)

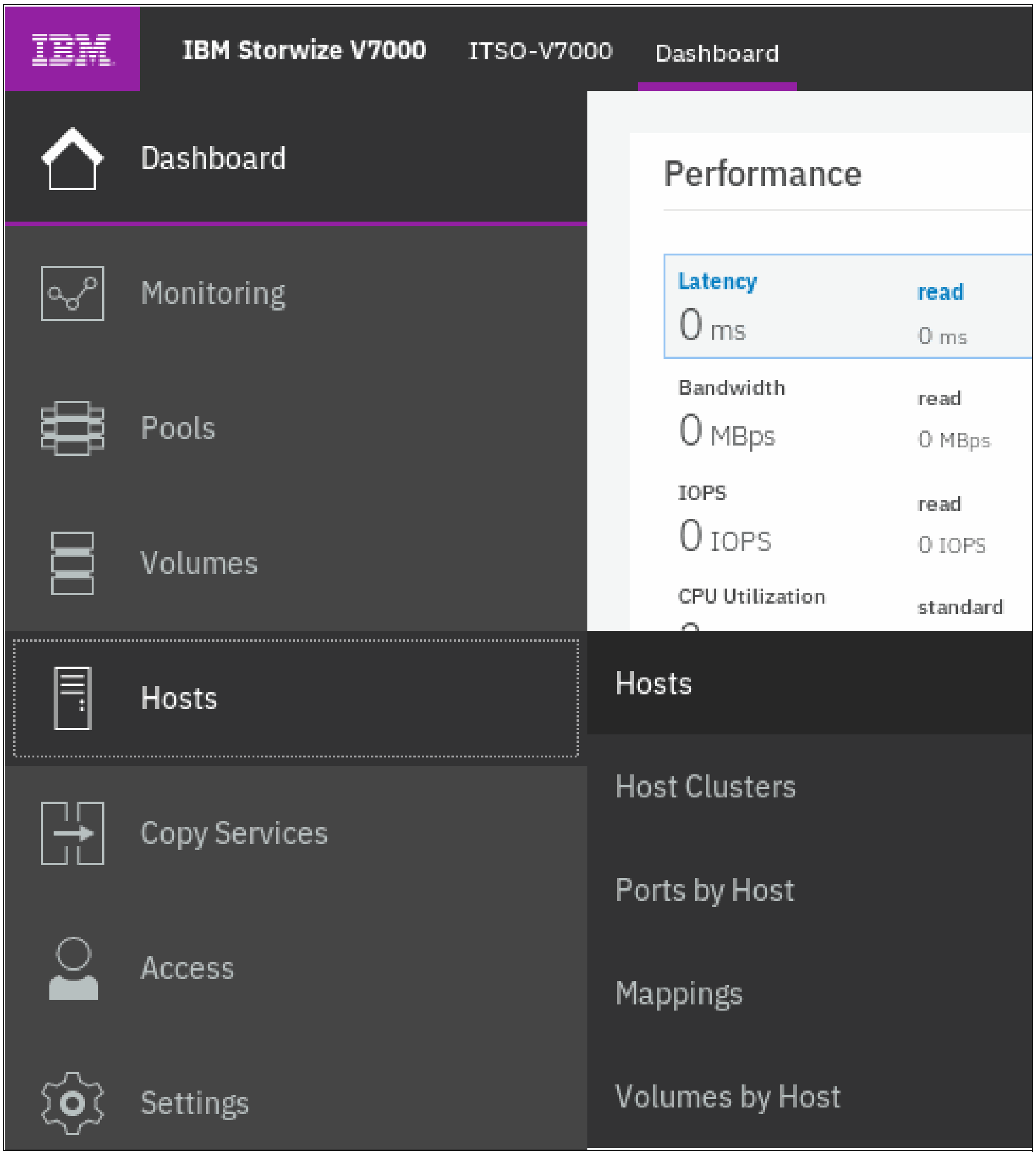


Figure 8-25 IBM Spectrum Virtualize Hosts menu

In the **Hosts** → **Hosts** view, three hosts were created and volumes are mapped to them in our example. If needed, we can now modify these hosts by selecting a host and click **Actions**, or right-click the host to see the available tasks (see Figure 8-26 on page 346).

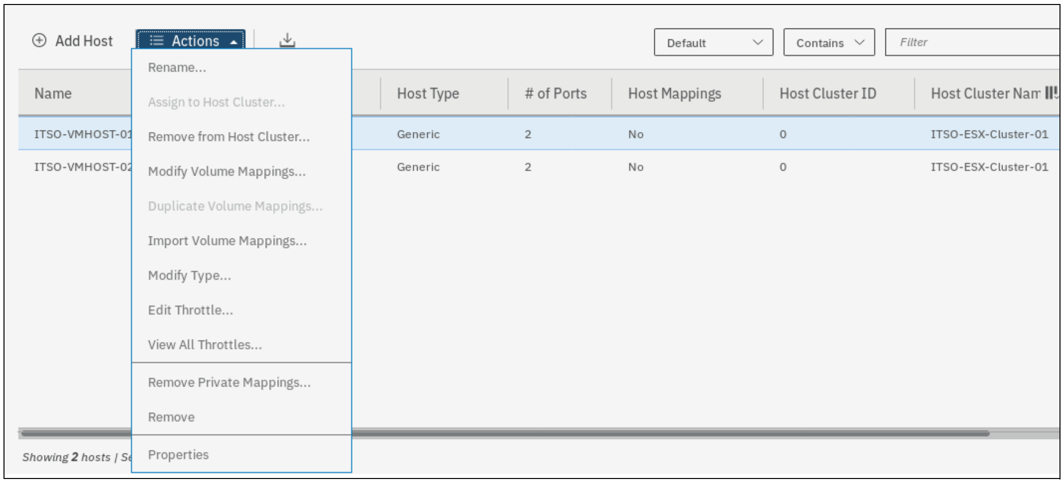


Figure 8-26 Host Actions

Modifying Volume Mappings

To modify the volumes are mapped to a specific host, complete the following steps:

1. From the **Actions** menu, select **Modify Volume Mappings** (see Figure 8-26). The window that is shown in Figure 8-27 opens. At the upper left, you can confirm that the correct host is targeted. The list shows all volumes that are mapped to the selected host. In our example, one volume with SCSI ID `0` is mapped to the host `ITSO-VMHOST-01`.

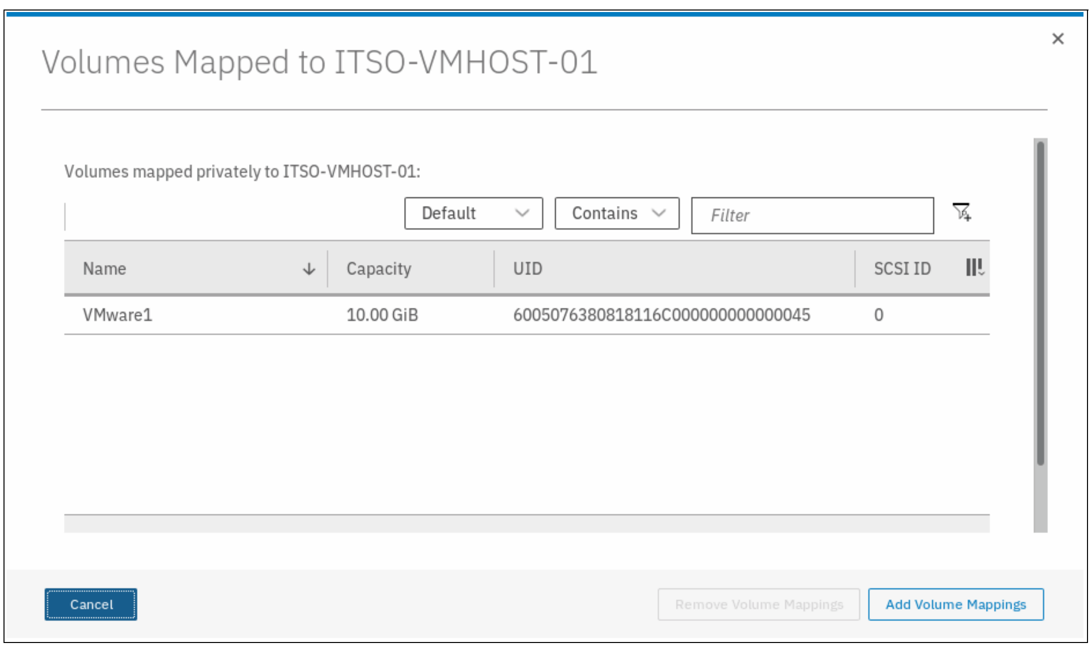


Figure 8-27 Modify Host Volume Mapping window

1. By selecting a listed volume, you can remove that Volume map from the host. However, in our case, we want to add a volume to our host. Continue by clicking **Add Volume Mapping** (see Figure 8-28 on page 347).

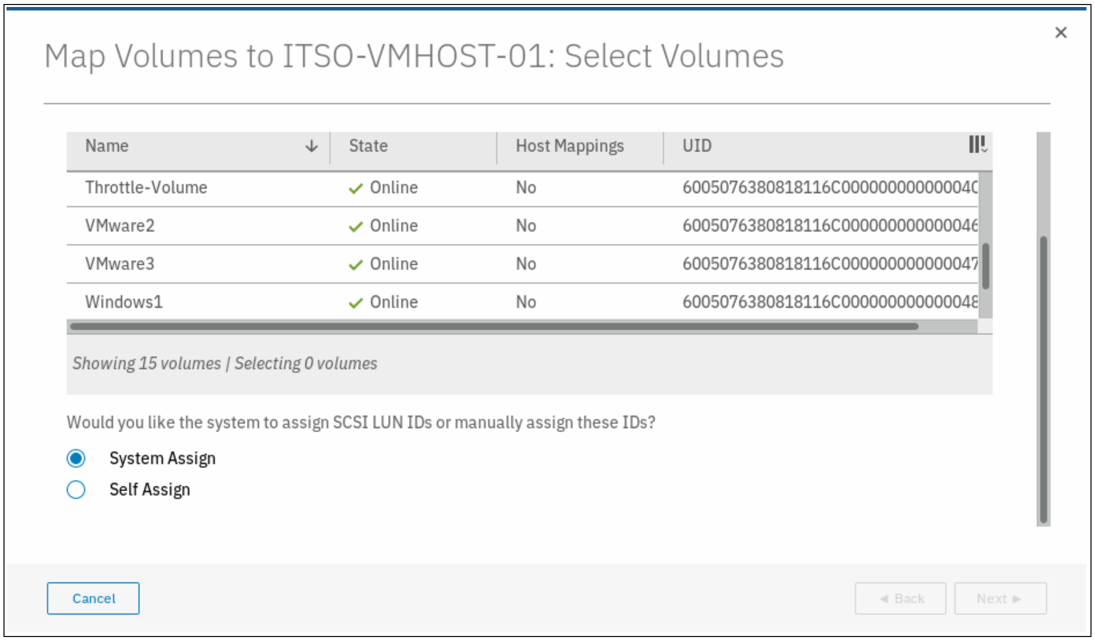


Figure 8-28 Volumes selection list

A new list displays that shows all volumes. You can easily identify if a volume you want to map is mapped to another host.

A new list displays that shows all volumes. You can easily identify if a volume you want to map is mapped to another host, as shown in Figure 8-29.

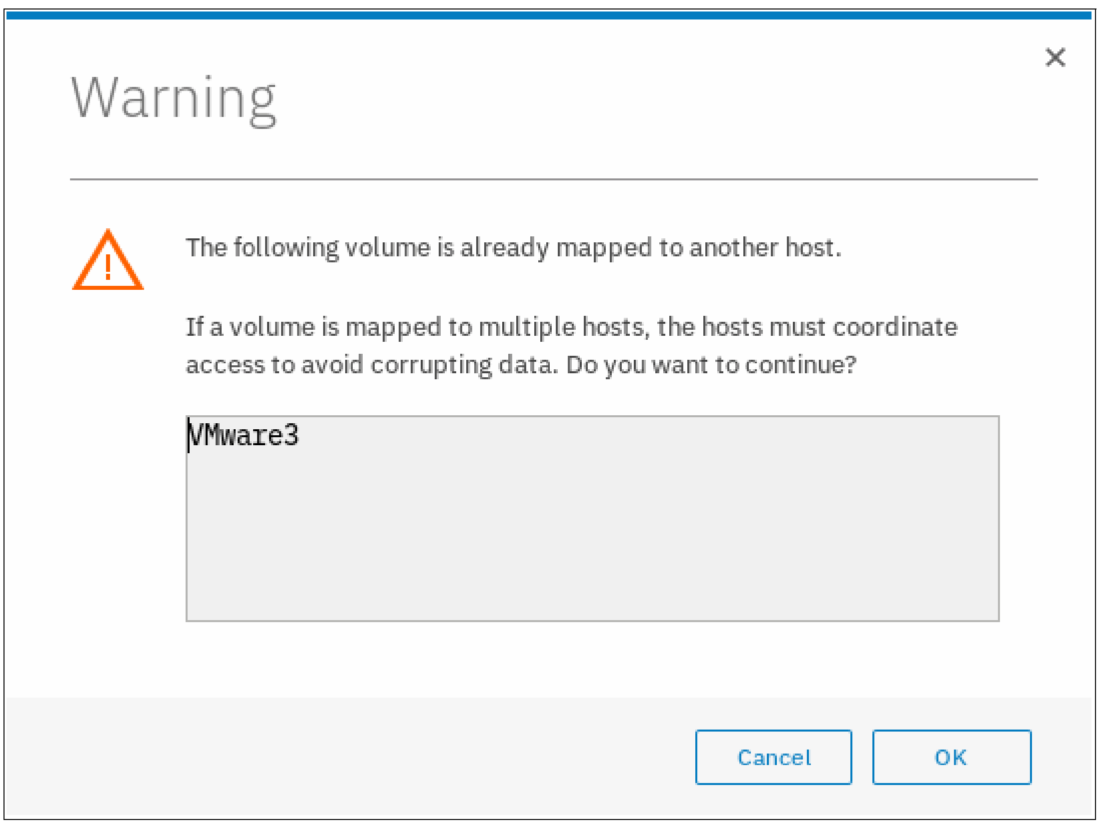


Figure 8-29 Volume mapped to another host warning

2. To map a volume, select it and click **Next** to map it to the host. The volume is assigned the next available SCSI ID if you leave **System Assign** selected. However, by selecting **Self Assign**, you can manually set SCSI IDs, as shown in Figure 8-30.

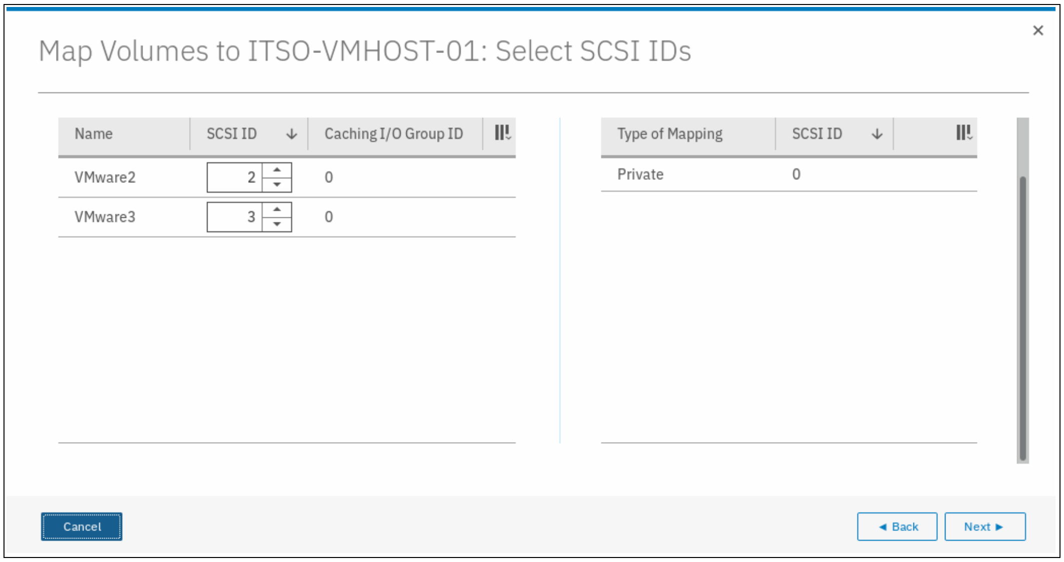


Figure 8-30 Modify Host Volume Mappings: Assign SCSI ID

If you select a SCSI ID that is in use for the host, you cannot proceed. As shown in Figure 8-29 on page 347, we selected SCSI ID 0. However, you can see in the right column SCSI ID 0 is allocated. By changing to SCSI ID 1, we can click **Next**.

3. A summary window opens showing the new mapping details (see Figure 8-31). After confirming that this result is what you planned, click **Map Volumes** and then, click **Close**.

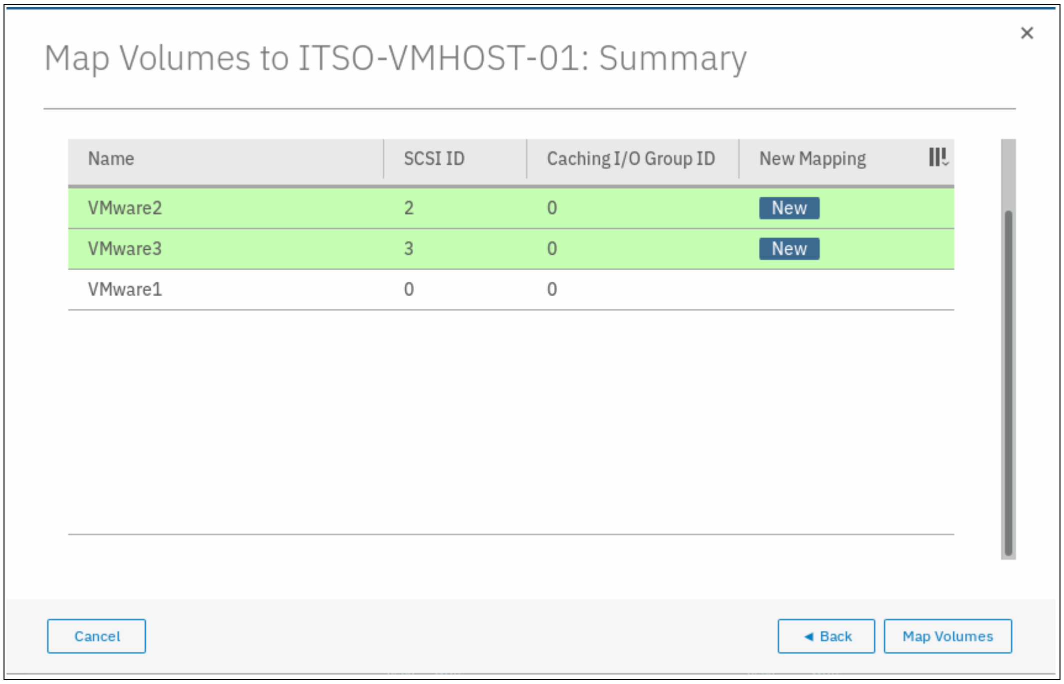


Figure 8-31 Confirm Modified Mappings

Note: The SCSI ID of the volume can be changed only before it is mapped to a host. Changing it after is not possible unless the volume is unmapped from it again.

Removing private mappings from a host

A host can access only those volumes on your Storwize V7000 system that are mapped to it. If we want to remove access to all volumes for one host, regardless of how many volumes are mapped to it, complete the following steps:

1. From the Hosts pane, select the host, click **Actions** → **Remove Private Mappings** to remove all access that the selected host has to its volumes (see Figure 8-32).

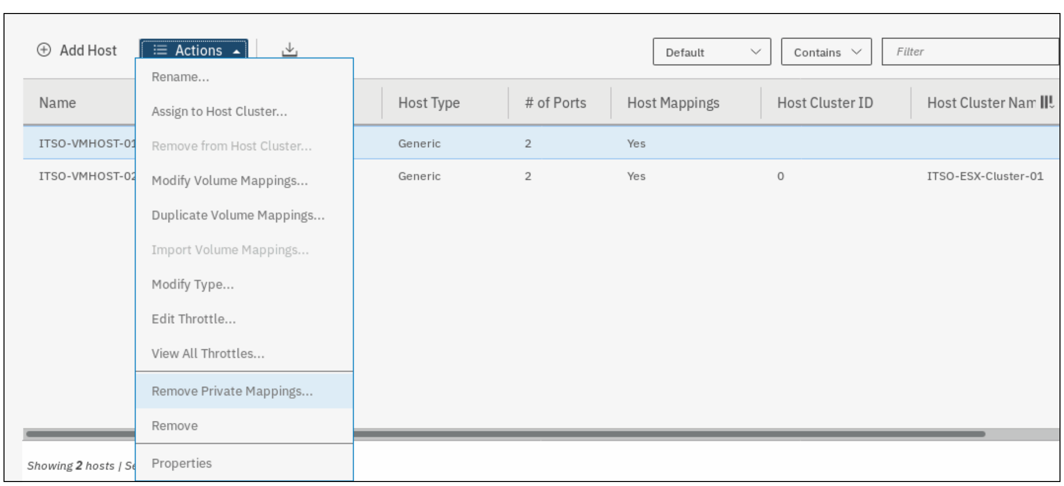


Figure 8-32 Unmap All Volumes action

2. You are prompted to confirm the number of mappings to be removed. To confirm your action, enter the number of volumes to be removed and click **Unmap** (see Figure 8-33). In this example, we removed three volume mappings.

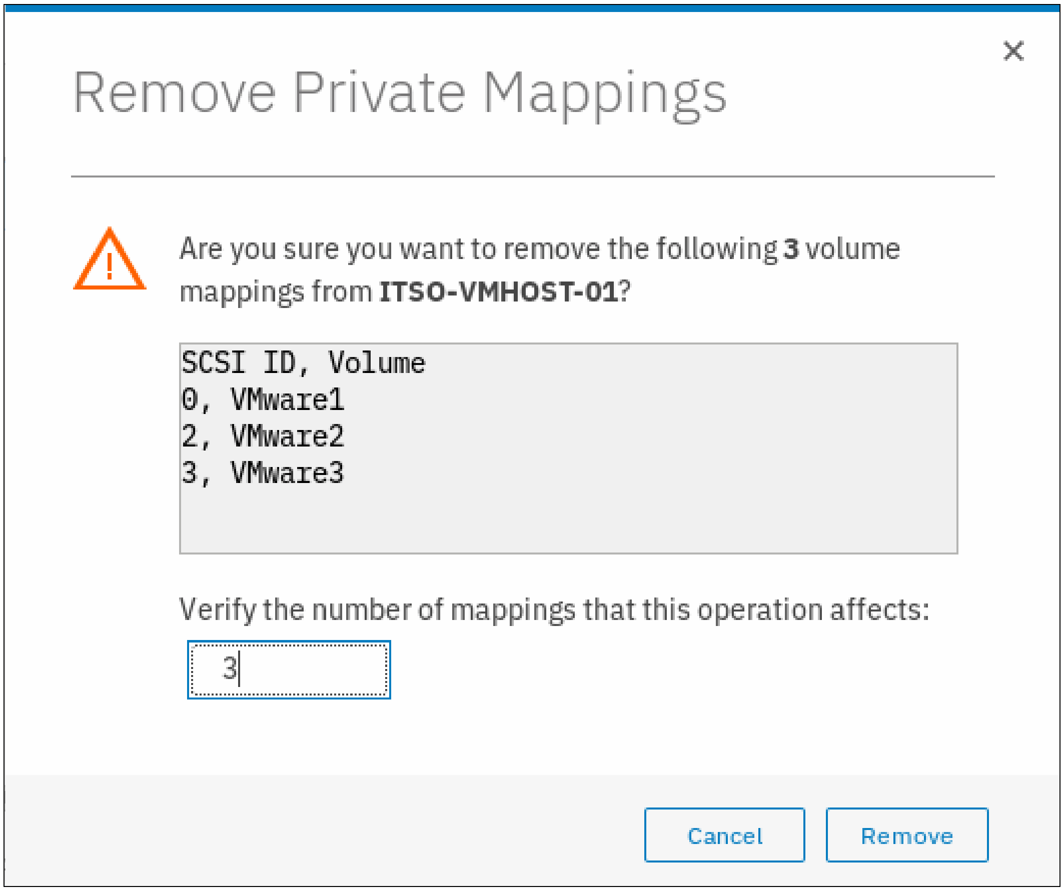


Figure 8-33 Confirm the number of mappings to be removed

> **Unmapping:** If you click **Unmap**, all access for this host to volumes that are controlled by the Storwize V7000 system is removed. Ensure that you run the required procedures on your host operating system, 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 Storwize V7000 GUI.

3. The changes are applied to the system. Click **Close**. Figure 8-34 shows that the selected host no longer has any host mappings.

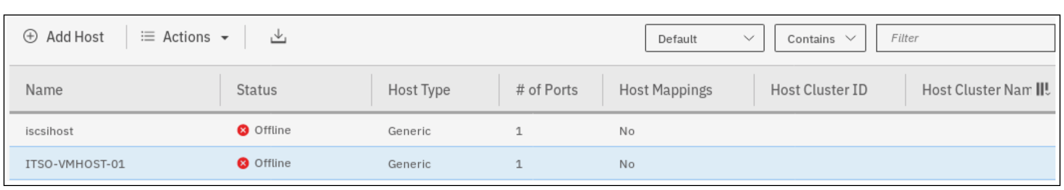


Figure 8-34 All mappings for host ITSO-VMHOST-01have been removed

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: by duplicating the mappings on the existing host object to the new host object, or by importing the host mappings to the new host. To duplicate the mappings, complete the following steps:

1. To duplicate an existing host mapping, select the host that you want to duplicate and click **Actions** → **Duplicate Volume Mappings** (see Figure 8-35). In our example, we duplicate the volumes that are mapped to host `ITSO-VMHOST-01` to the new host `ITSO-VMHOST-02`.

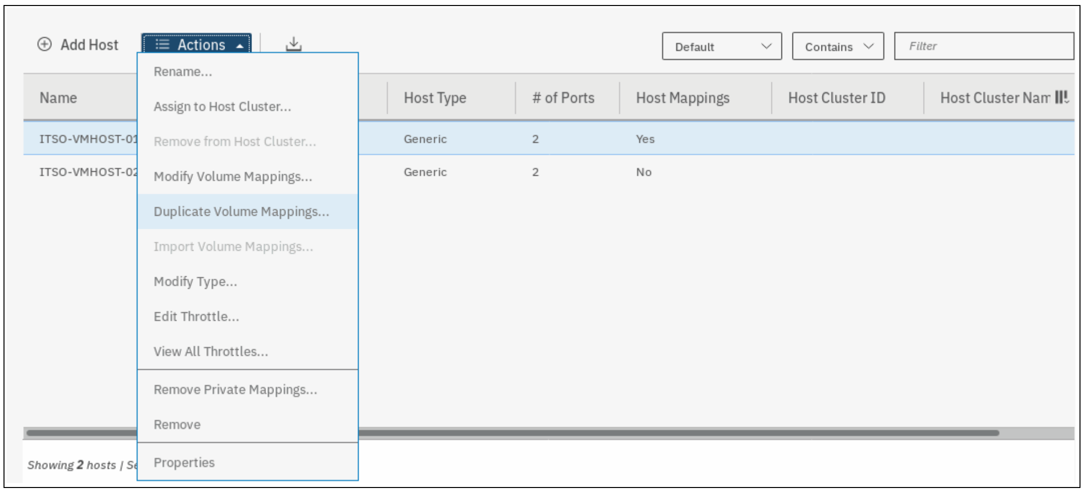


Figure 8-35 Duplicate Host Volume Mappings menu

2. The Duplicate Mappings window opens. Select a listed target host object to which you want to map all the existing source host volumes and click **Duplicate** (see Figure 8-36).

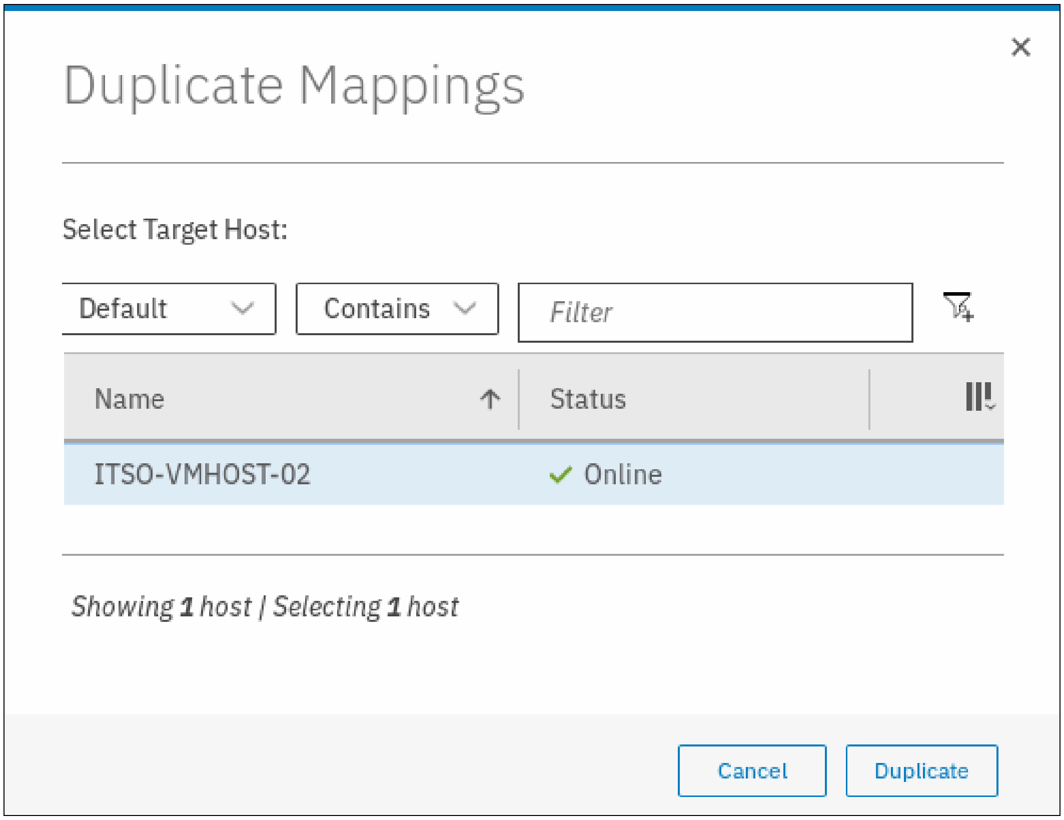


Figure 8-36 Duplicate mappings window

Note: You can duplicate mappings only to a host that has no volumes mapped.

3. After the task completion is displayed, we can verify the new mappings on the new host object. From the **Hosts** menu (see Figure 8-35 on page 351), right-click the target host and select **Properties**.

4. Click the **Mapped Volumes** tab and verify that the required volumes were mapped to the new host (see Figure 8-37).

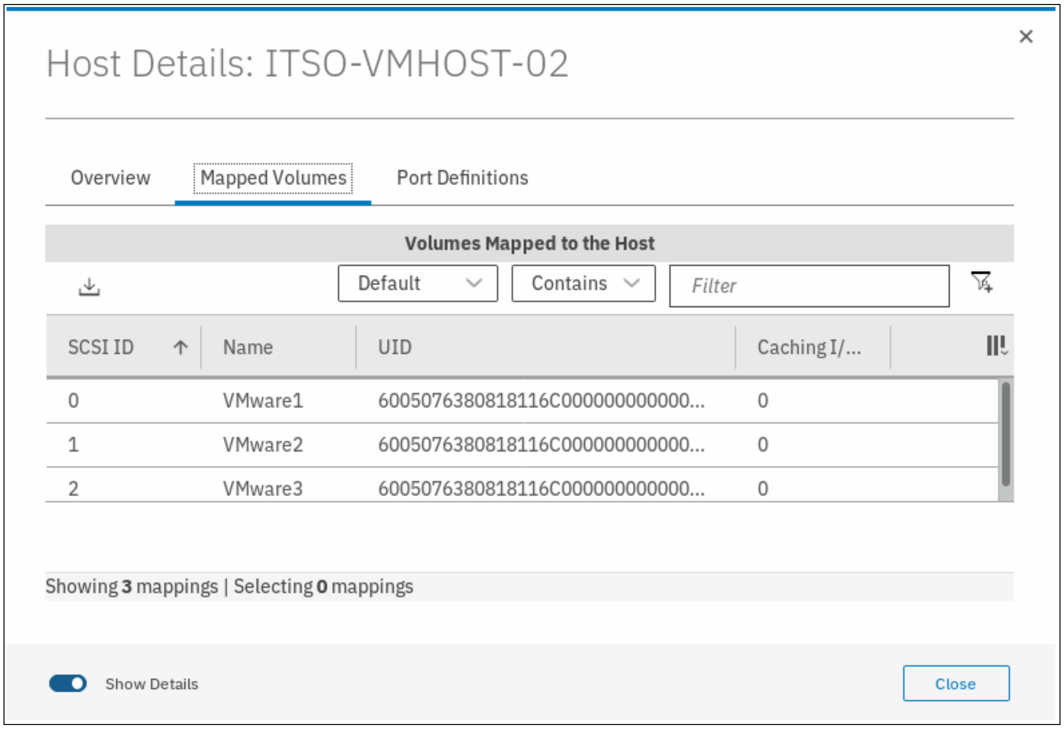


Figure 8-37 Host Details, new mappings on target host

The same volume mappings can be accomplished by importing existing hosts mappings to the new host:

1. Select the new host without any mapped volumes and click **Actions** → **Import Volume Mappings** (see Figure 8-38).

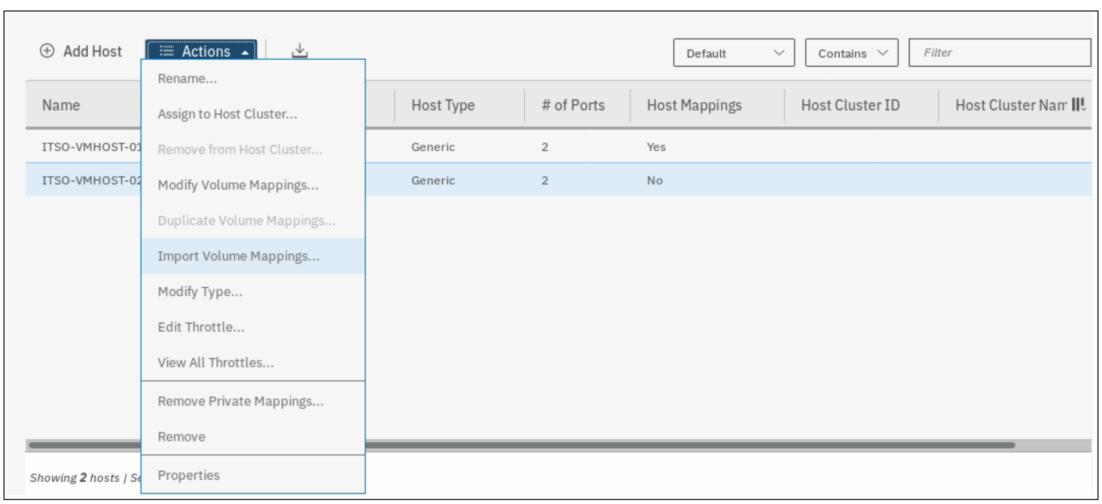


Figure 8-38 Hosts Actions: Import Volume Mappings

2. The import mappings window opens. Select the source host from which you want to import the volume mappings. As shown in Figure 8-39, we select the host `ITSO-VMHOST-01` and click **Import**.

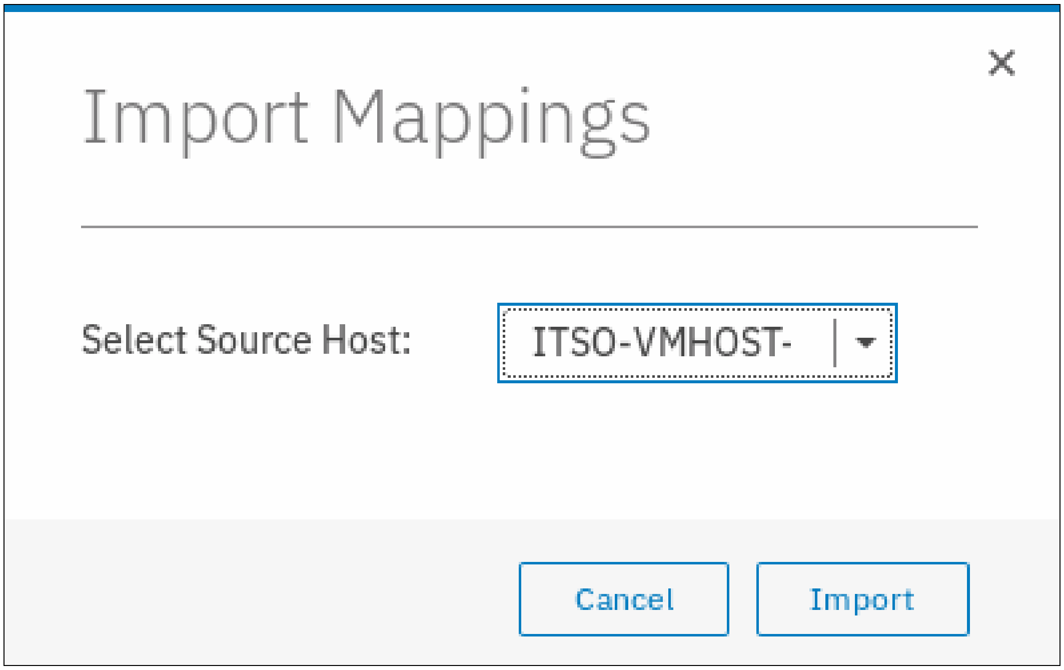


Figure 8-39 import volume mappings source host selection

3. After the task completes, verify that the mappings are as expected from the **Hosts** menu (see Figure 8-26 on page 346), right-click the target host, and select **Properties**. Then, click the **Mapped Volumes** tab and verify that the required volumes were mapped to the new host (see Figure 8-37 on page 353).

Renaming a host

To rename a host, complete the following steps:

1. Select the host, and then right-click and select **Rename** (see Figure 8-40).

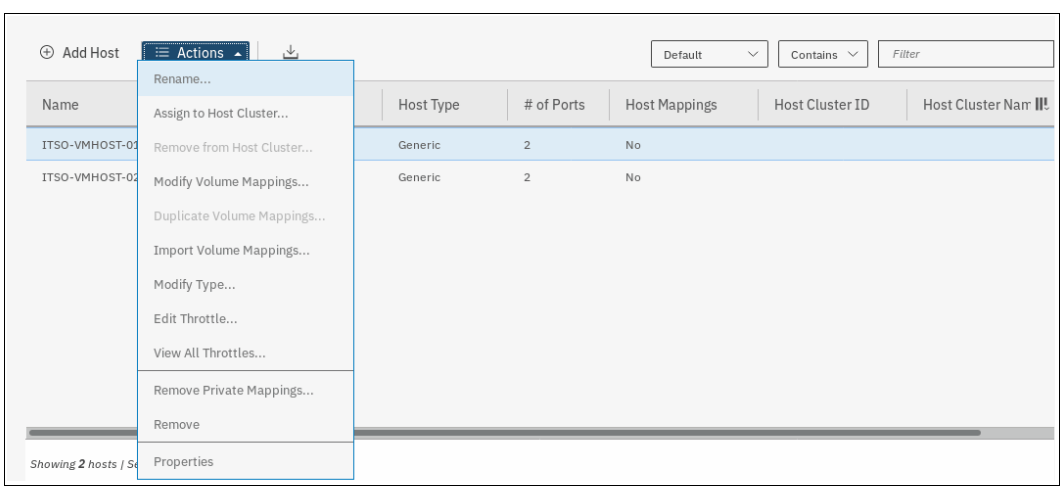


Figure 8-40 Rename a host

2. Enter a new name and click **Rename** (see Figure 8-41). If you click **Reset**, the changes are reset to the original host name.

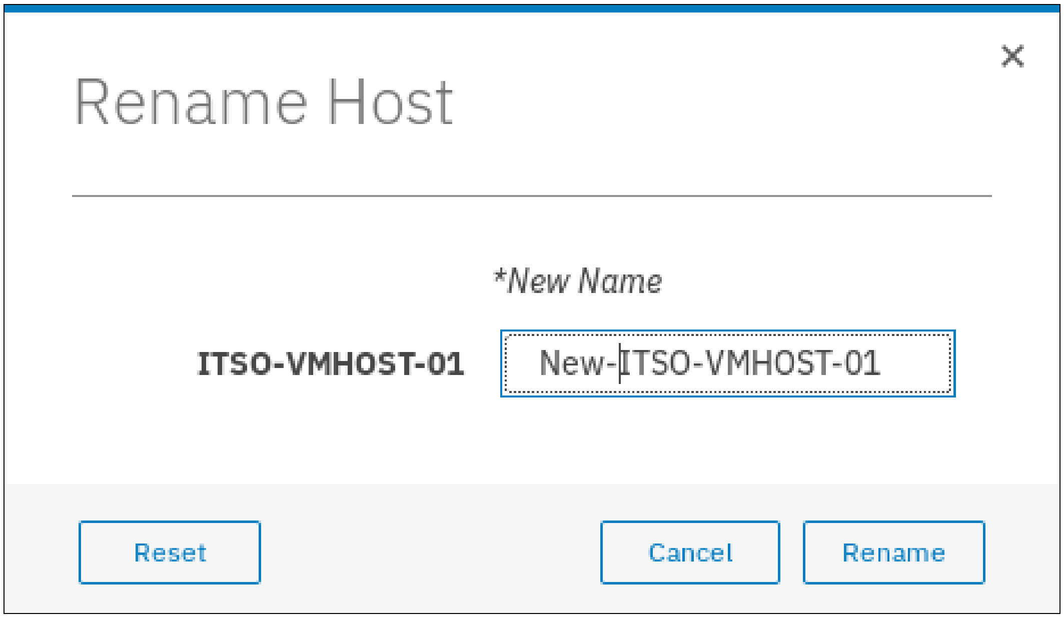


Figure 8-41 Rename Host window

3. After the changes are applied to the system, click **Close**.

Removing a host

To remove a host object definition, complete the following steps:

1. From the Hosts pane, select the host and right-click it or click **Actions** → **Remove** (see Figure 8-42).

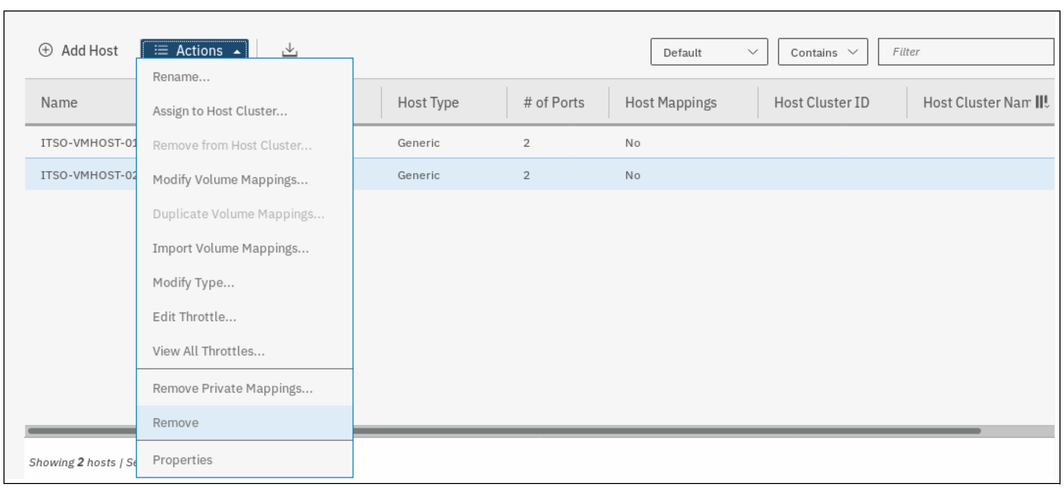


Figure 8-42 Remove a host

2. Confirm that the window displays the correct list of hosts that you want to remove by entering the number of hosts to remove, and clicking **Delete** (see Figure 8-43).

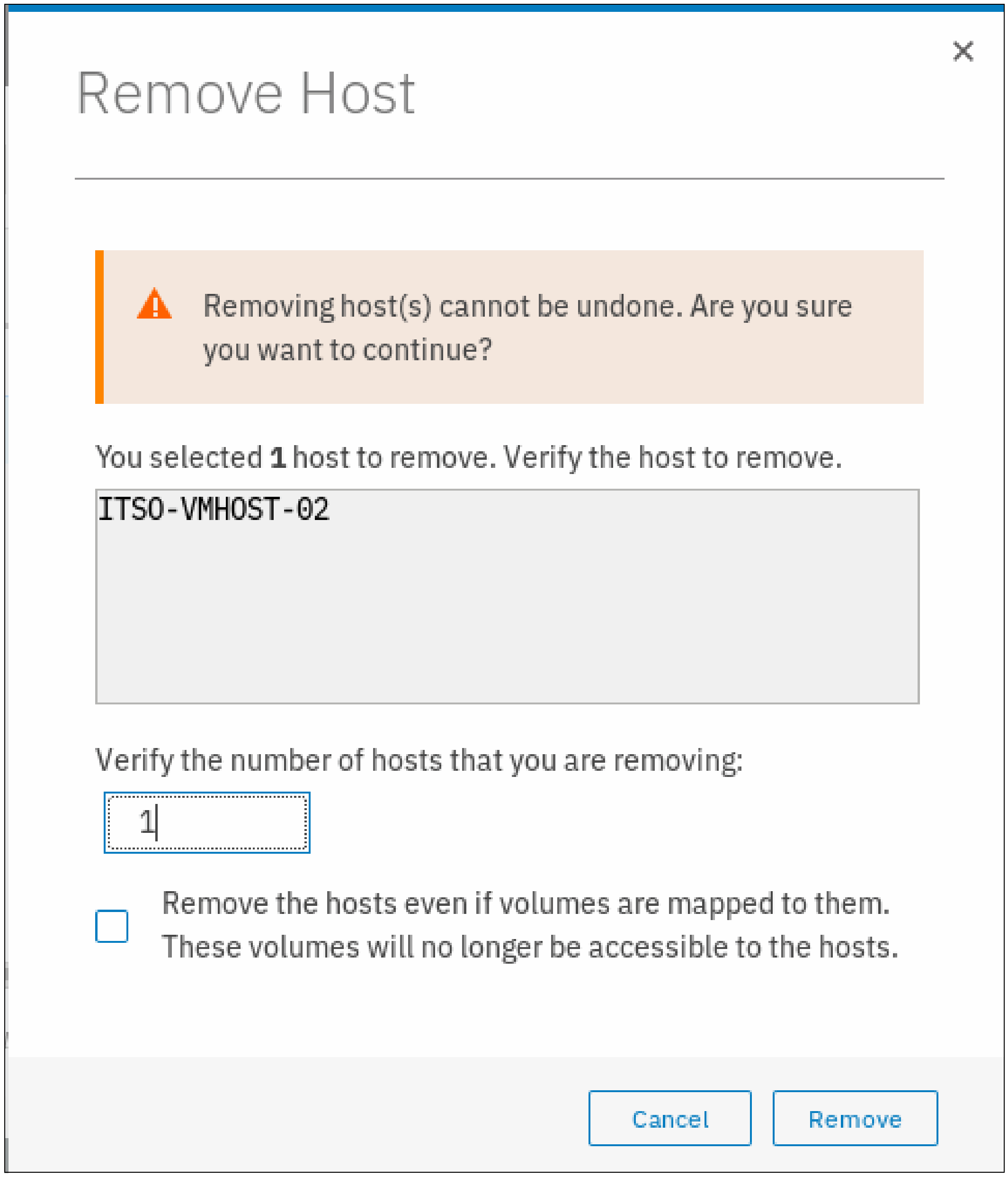


Figure 8-43 Confirm the removal of the host

3. If the host that you are removing has volumes mapped to it, force the removal by selecting the **Remove the hosts even if volumes are mapped to them** option in the lower part of the window. By selecting this option, the host is removed and it no longer can access any volumes on this system.
4. After the task is completed, click **Close**.

Host properties

To view a host object's properties, complete the following steps:

1. From the IBM Spectrum Virtualize GUI Hosts pane, select a host and right-click it or click **Actions** → **Properties** (see Figure 8-44).

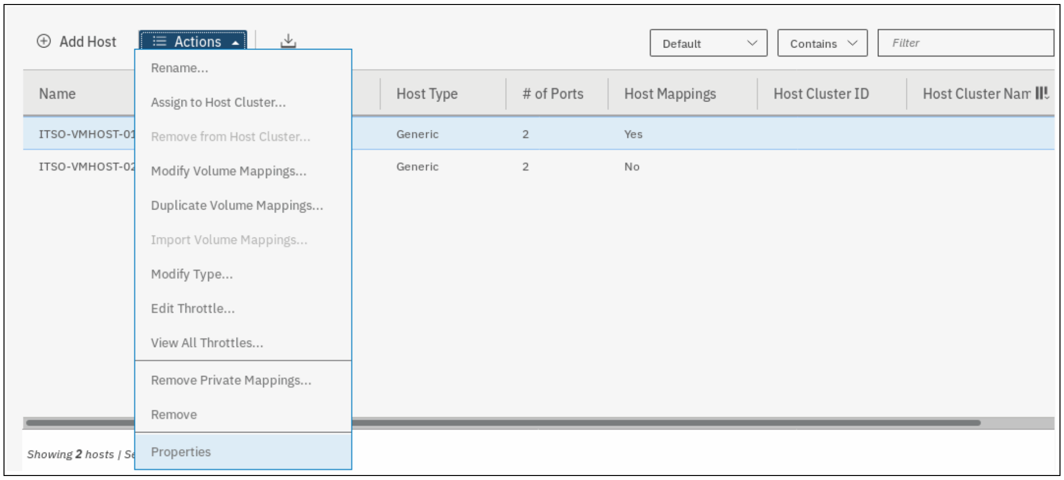


Figure 8-44 Host properties

The Host Details window opens (see Figure 8-45).

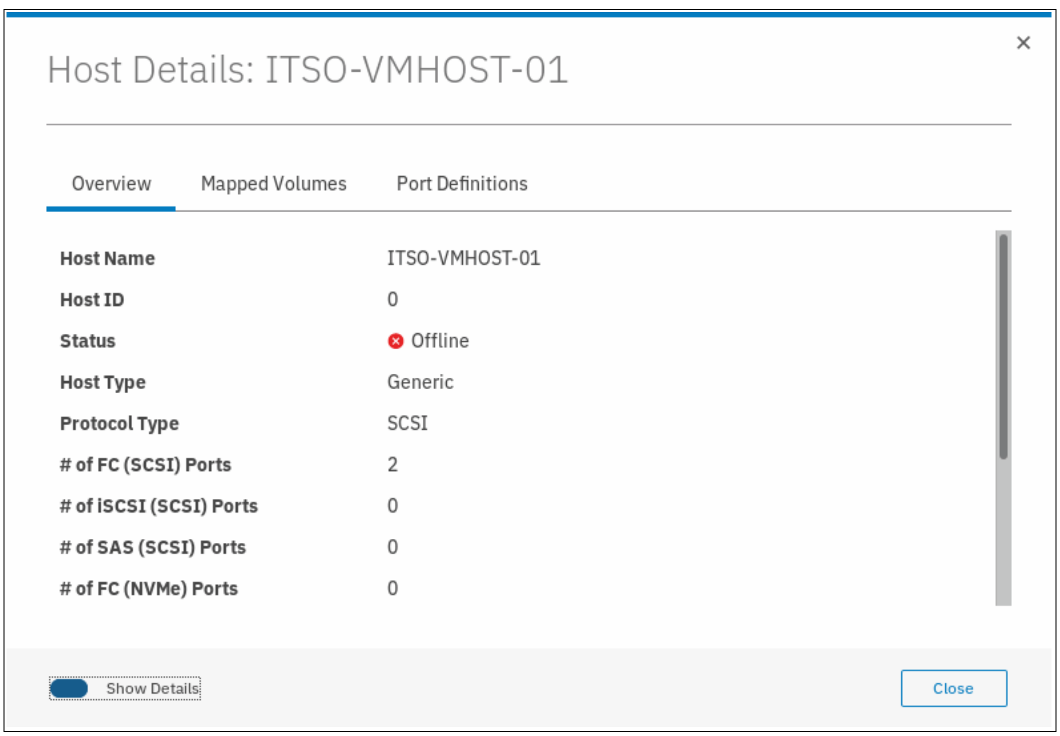


Figure 8-45 Host properties overview

The Host Details window shows an overview of the selected host properties. It includes three tabs: Overview, Mapped Volumes, and Port Definitions. The Overview tab is shown in Figure 8-45.

2. Select the **Show Details** slider to see more information about the host (see Figure 8-46).

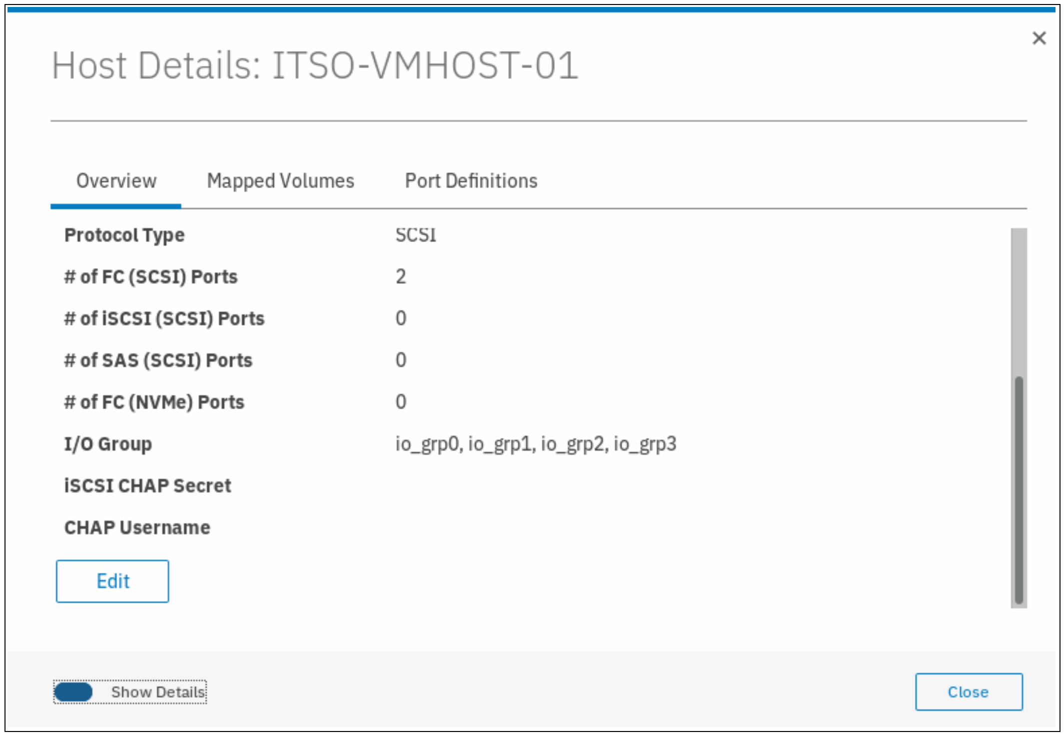


Figure 8-46   Host Properties: Show details

3. Click **Edit** to change the host properties (see Figure 8-47).

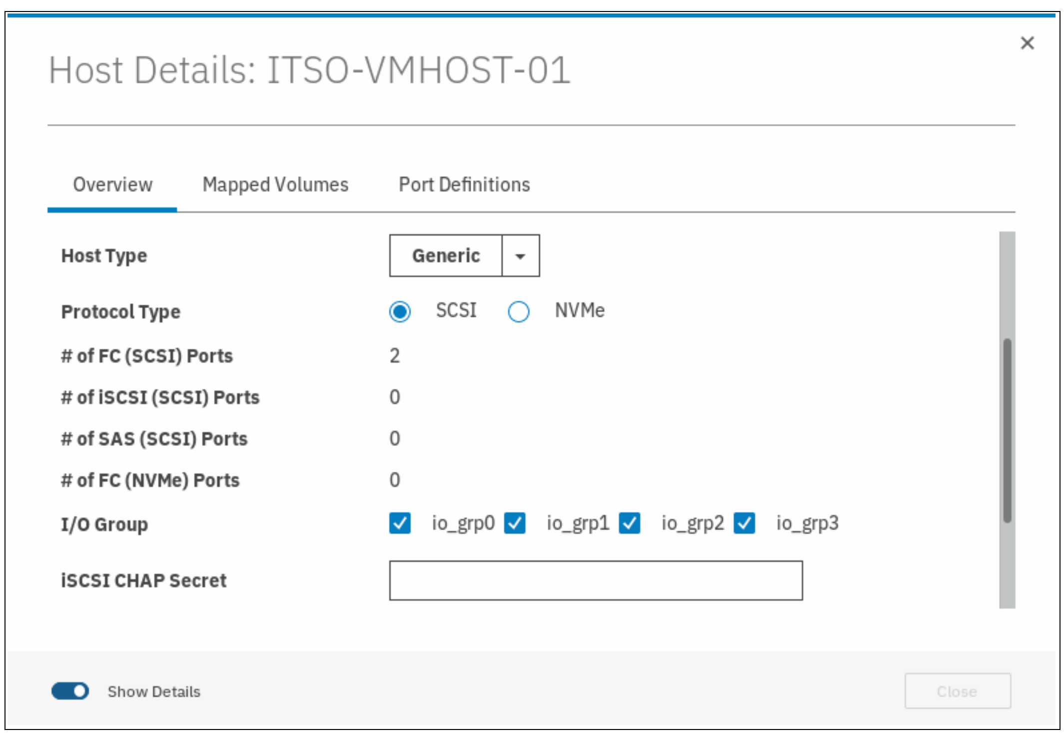


Figure 8-47   Edit host properties

In the window that is shown in Figure 8-47 on page 358, you can modify the following properties:

– Host Name: Change the host name.

– Host Type: Change this setting if you are going to attach HP/UX, OpenVMS, or TPGS hosts.

– I/O Group: Host can access volumes that are mapped from selected I/O Groups.

– iSCSI CHAP Secret: Enter or change the iSCSI CHAP secret if this host is using iSCSI.

4. When finished making changes, click **Save** to apply them. The editing window closes.

   The Mapped Volumes tab shows a summary of which volumes are currently mapped with which SCSI ID and UID to this host (see Figure 8-48). The Show Details slider does not show any more information for this list.

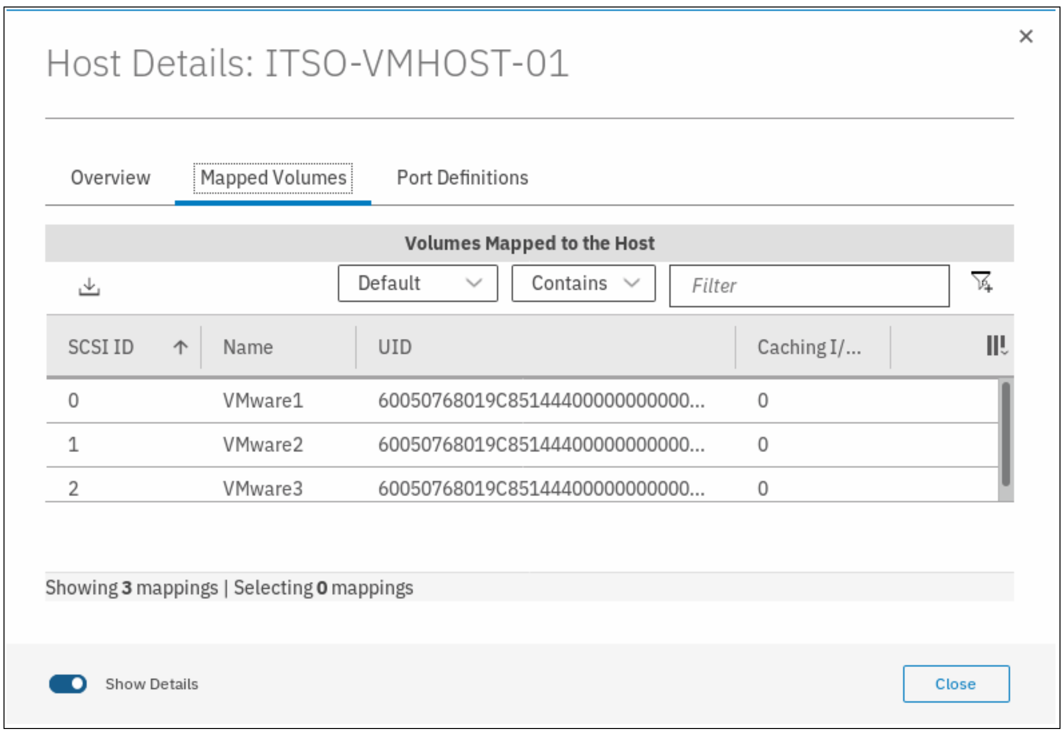


Figure 8-48 Mapped volumes tab

The Port Definitions tab shows the configured host ports of a host and provides status information about them (see Figure 8-49).

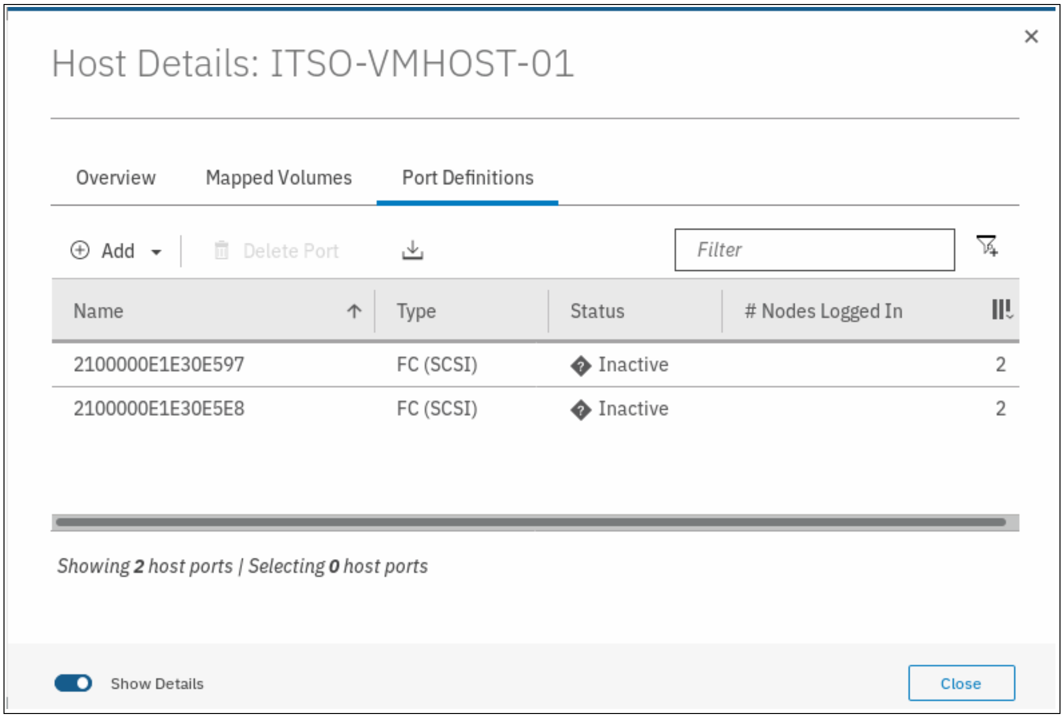


Figure 8-49 Port definitions

This window offers the option to **Add** or **Delete Port** on the host, as described in 8.4.4, "Adding and deleting host ports" on page 361.

5. Click **Close** to close the Host Details window.

8.4.4 Adding and deleting host ports

To configure host ports, complete the following steps:

1. From the left menu, select **Hosts** → **Ports by Host** to open the associated pane (see Figure 8-50).

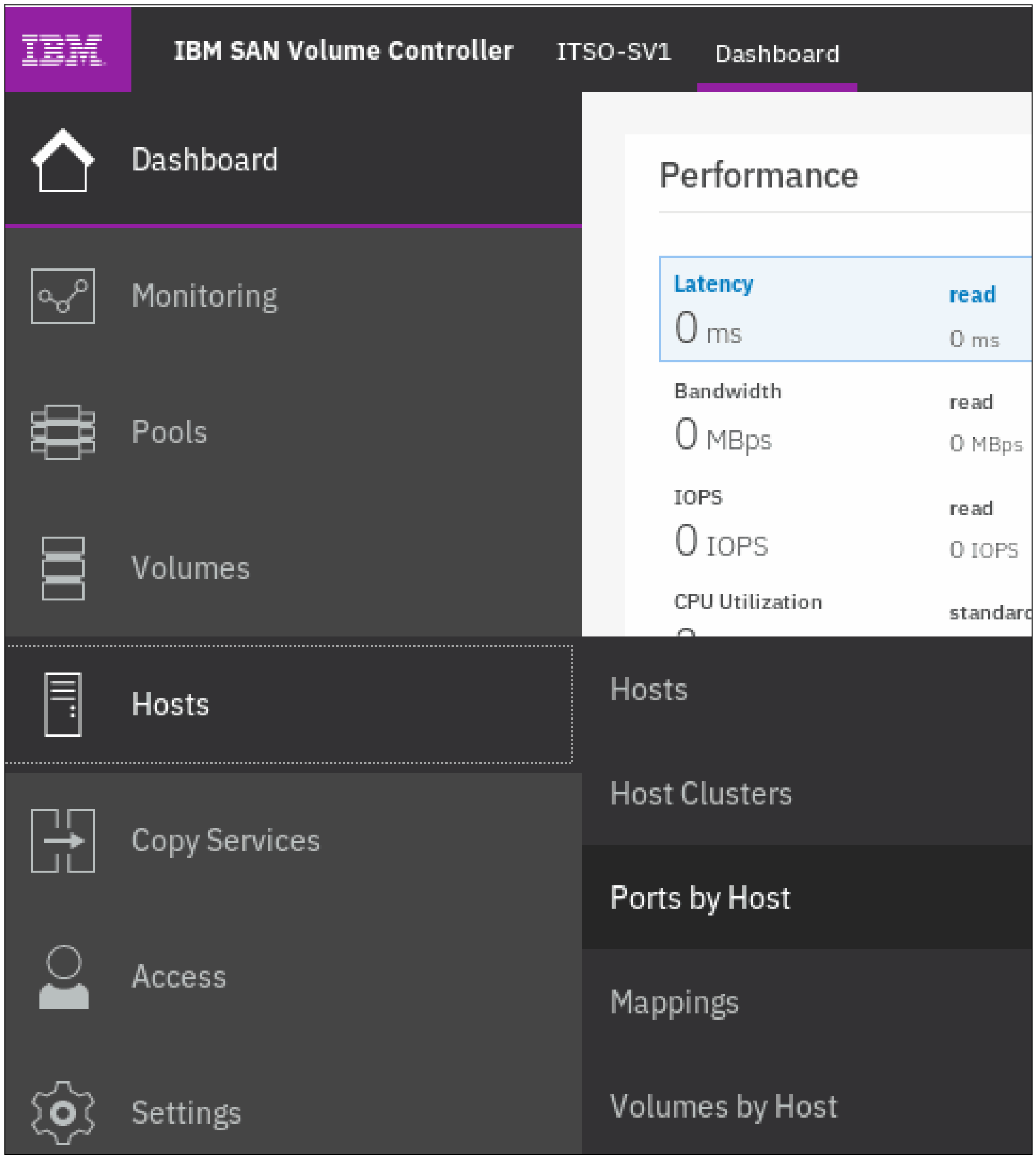


Figure 8-50 Ports by Host

2. A list of all the hosts is displayed. The function icons indicate whether the host is Fibre Channel, iSCSI, or SAS attached. The port details of the selected host are shown to the right. You can add a new host object by clicking **Add Host**. If you click **Actions** (see Figure 8-51), the tasks that are described in “Modifying Volume Mappings” on page 346 can be selected.

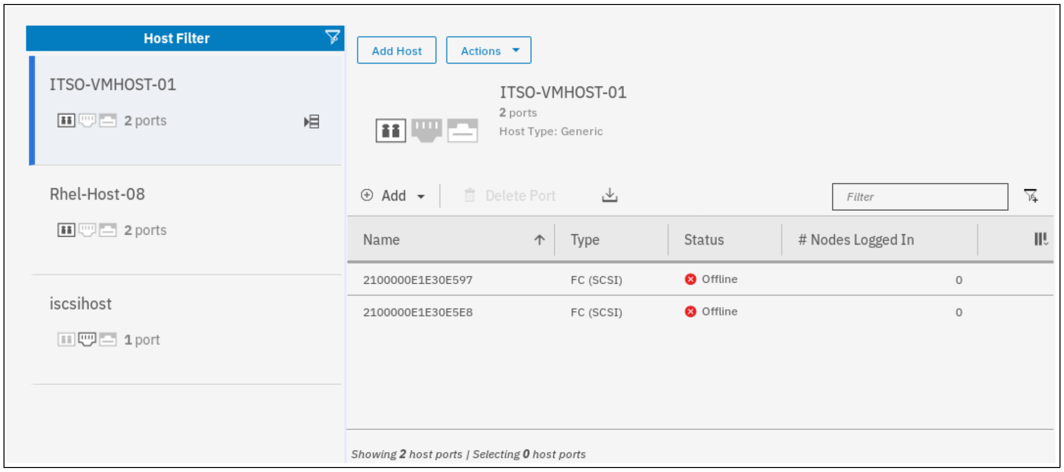


Figure 8-51 Ports by Host actions

Adding a Fibre Channel or iSCSI host port

To add a host port, complete the following steps:

1. Select the host.
2. Click **Add** (see Figure 8-52) and select one of the following options:
   a. **Fibre Channel Port** (see “Adding a Fibre Channel port” on page 363)
   b. **iSCSI Port** (see “Adding an iSCSI host port” on page 366)

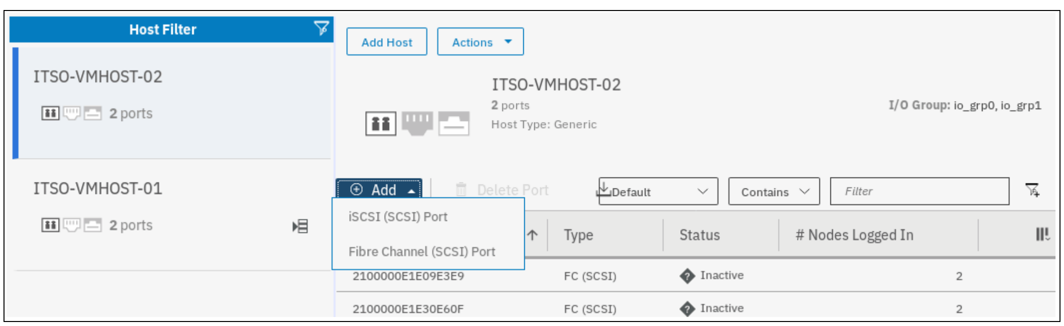


Figure 8-52 Add host ports

Adding a Fibre Channel port

To add a Fibre Channel port, complete the following steps:

1. Click **Fibre Channel Port** (see Figure 8-52 on page 362). The Add Fibre Channel Ports window opens (see Figure 8-53).

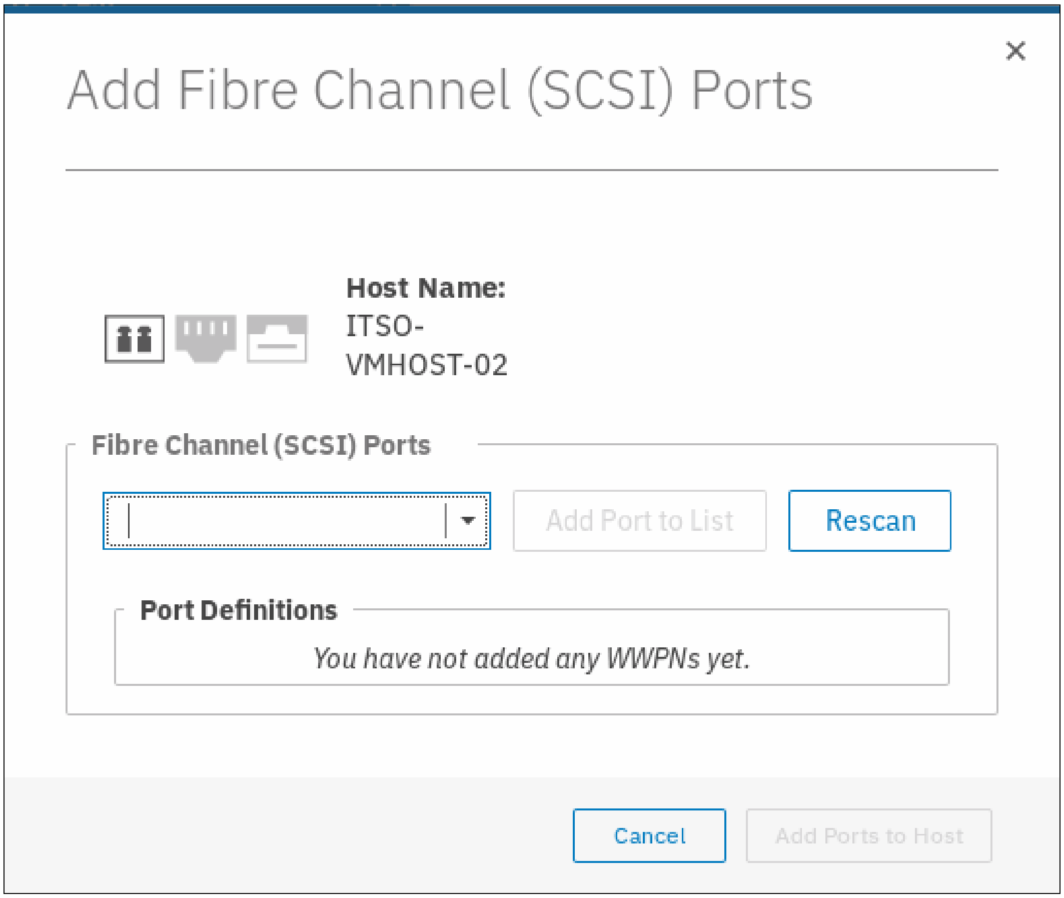


Figure 8-53 Add Fibre Channel Ports window

2. Click the drop-down menu to display a list of all discovered Fibre Channel WWPNs. If the WWPN of your host is not available in the menu, enter it manually or check the SAN zoning to ensure that connectivity is configured. Then, rescan storage from the host.

3. Select the WWPN that you want to add and click **Add Port to List** (see Figure 8-54).

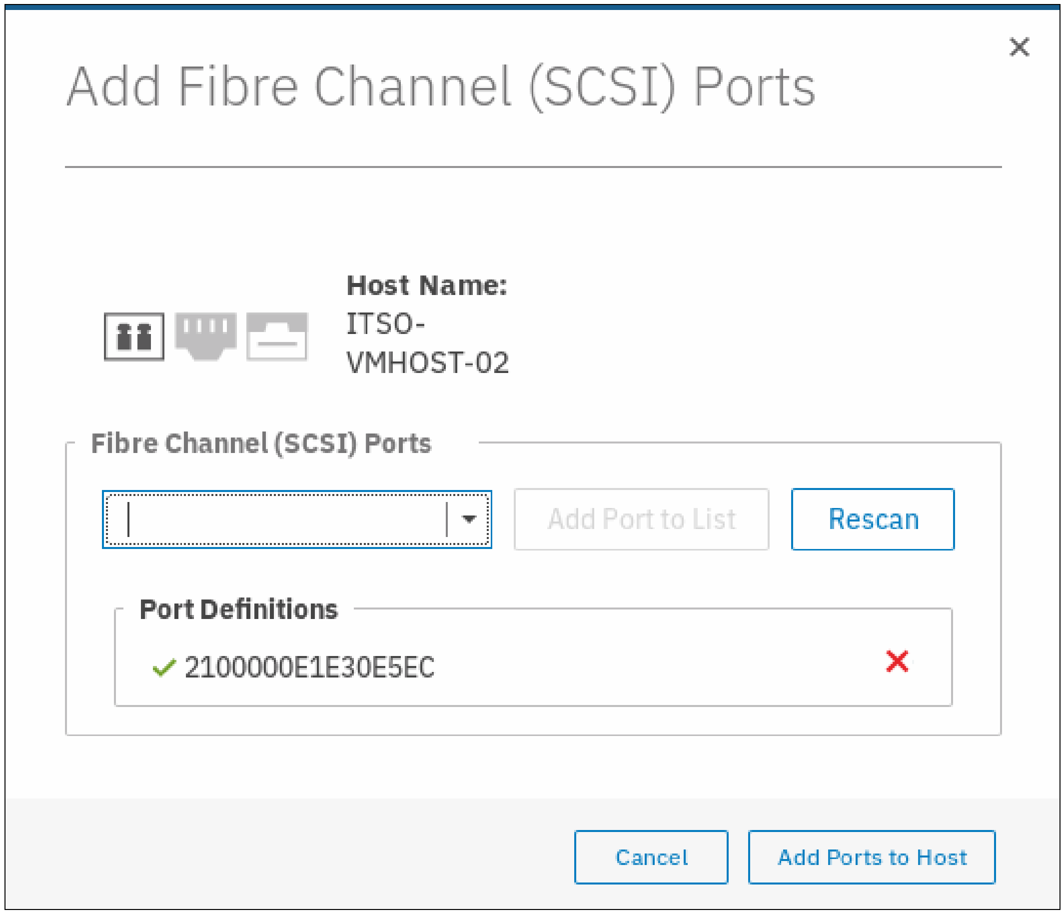


Figure 8-54 Add a port to list

This step can be repeated to add other ports to the host.

4. To add an offline port (if the WWPN of your host is not available in the drop-down menu), manually enter the WWPN of the port into the **Fibre Channel Ports** field and click **Add Port to List**.

The port is unverified (see Figure 8-55) because it is not logged on to the Storwize V7000 system. The first time that it logs on, its state is automatically changed to `online`, and the mapping is applied to this port.

Figure 8-55 Unverified port

5. To remove a port from the adding list, click the red X next to the port. In this example, we delete the manually added FC port so only the detected port remains.

6. Click **Add Ports to Host** to apply the changes and click **Close**.

Adding an iSCSI host port

To add an iSCSI host port, complete the following steps:

1. Click **iSCSI Port** (see Figure 8-52 on page 362). The Add iSCSI Ports window opens (see Figure 8-56).

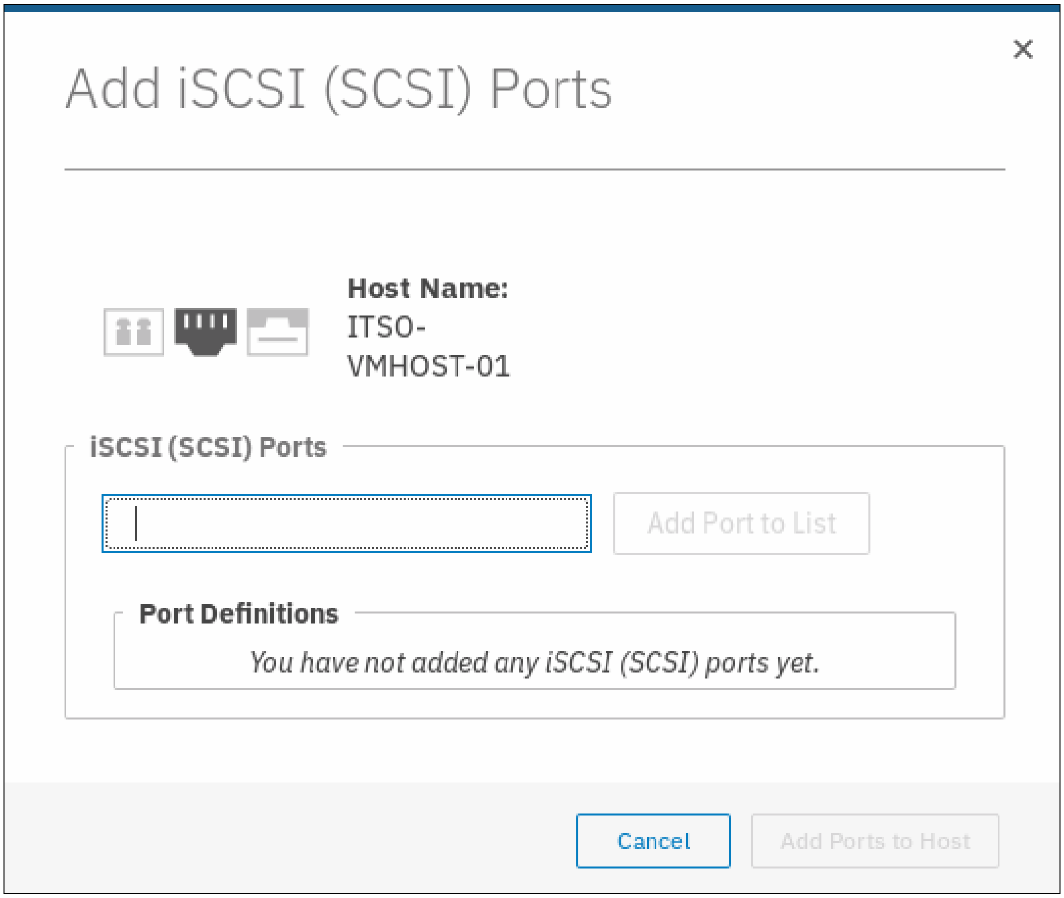


Figure 8-56 Add iSCSI host ports

2. Enter the initiator name of your host (see Figure 8-57) and click **Add Port to List**.

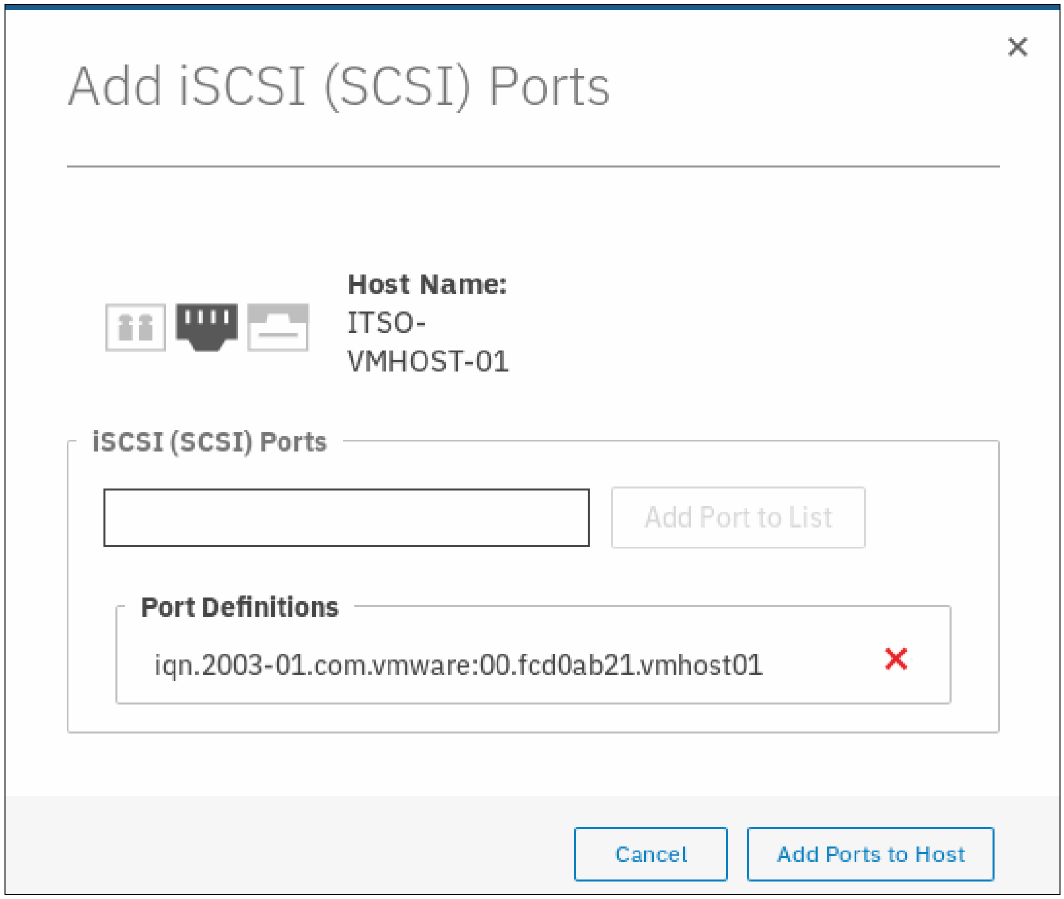


Figure 8-57 Enter the initiator name

3. Click **Add Ports to Host** to apply the changes to the system and then, click **Close**.

Adding a NVMe host port

To add an NVMe host port, complete the following steps:

1. Click **Add**, then **Fibre Channel (NVMe) Port** and the Add Fibre Channel (NVMe) Ports window opens, as shown in Figure 8-58.

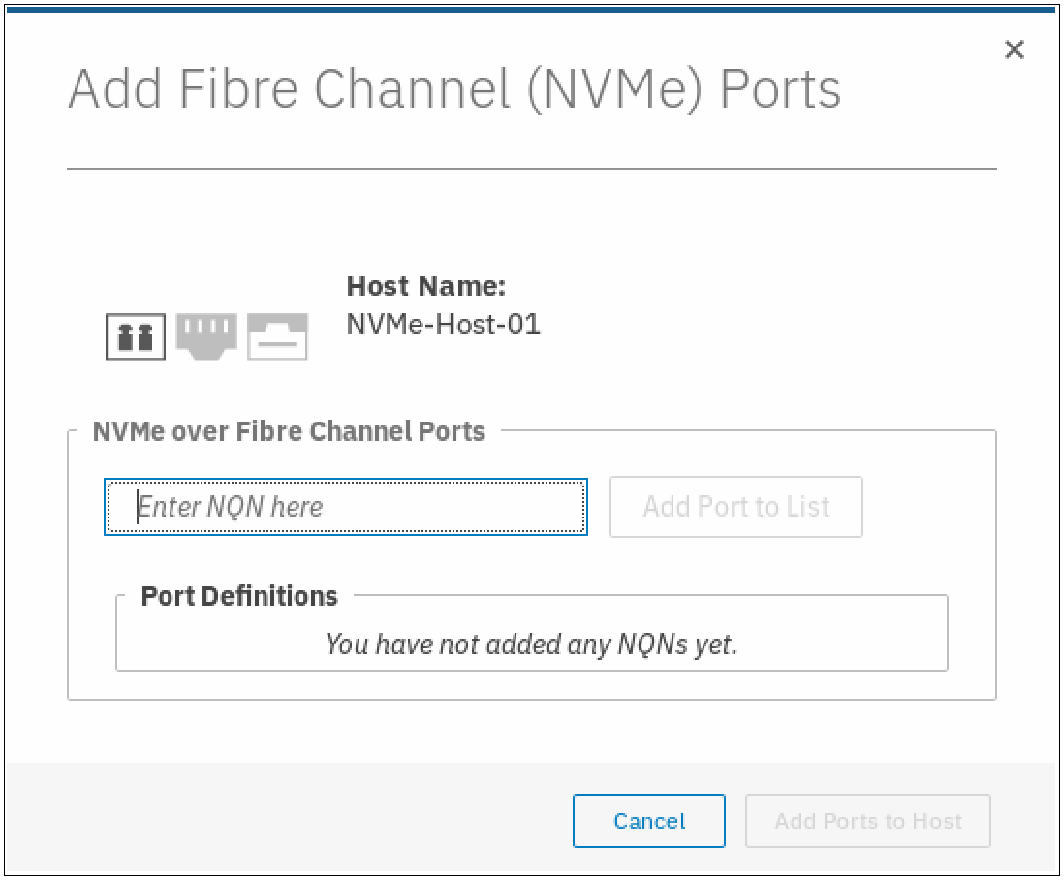


Figure 8-58 Add Fibre Channel (NVMe) Ports

2. Enter the NVMe Qualified Name (NQN). Then, click **Add Ports to Host**, as shown in Figure 8-59.

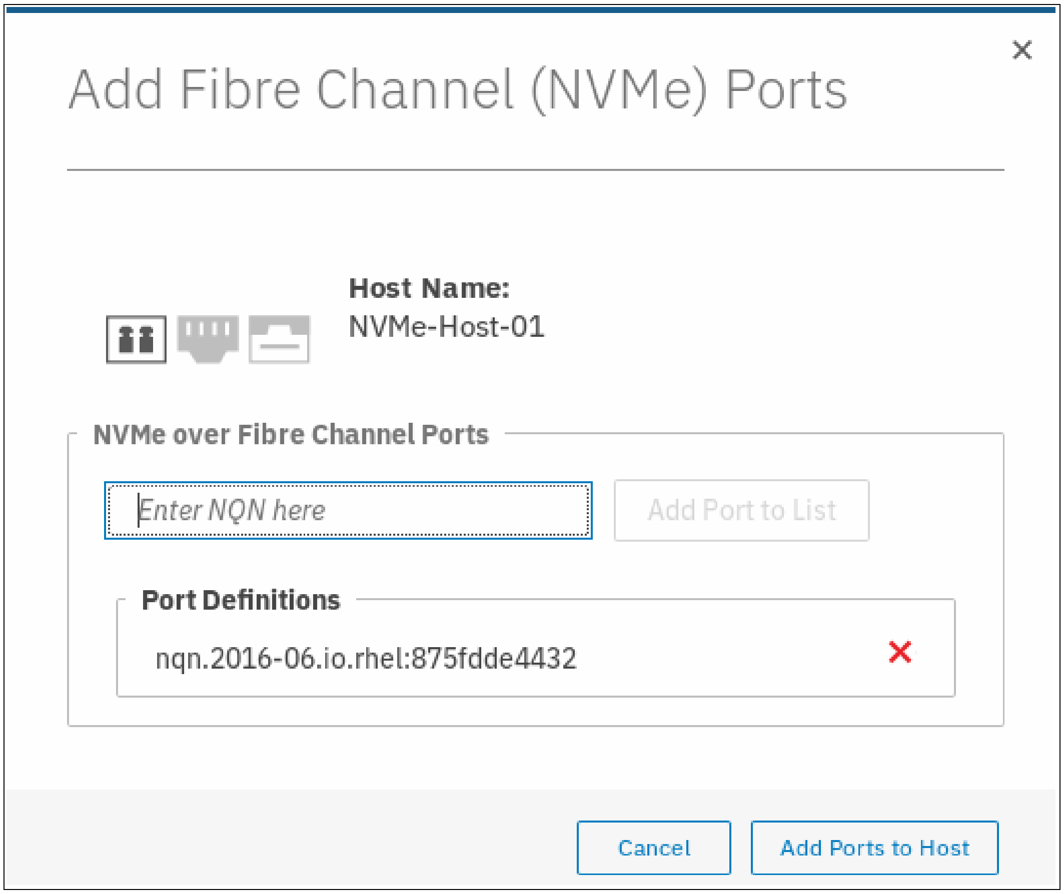


Figure 8-59 Enter the NVMe Qualified Name (NQN)

Deleting a host port

To delete a host port, complete the following steps:

1. Highlight the host port and right-click it or click **Delete Port** (see Figure 8-60).

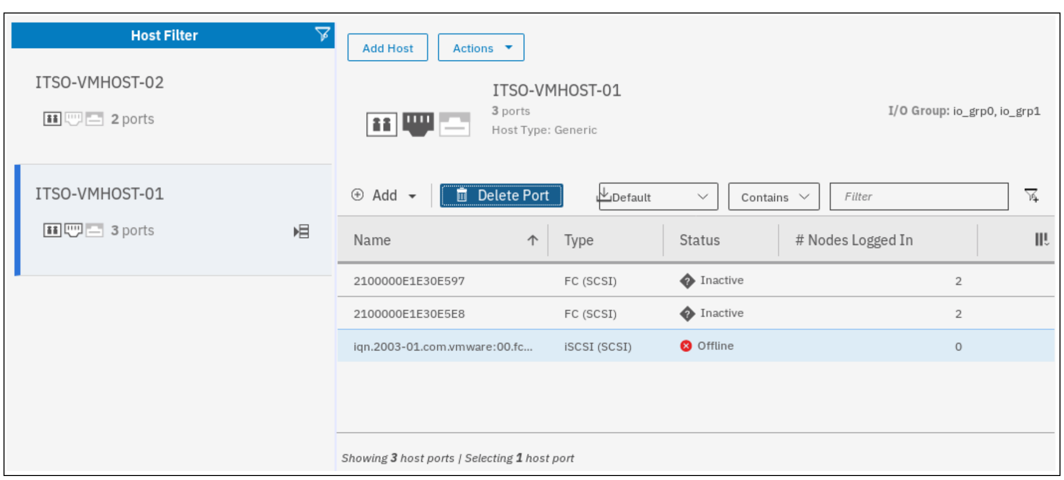


Figure 8-60 Delete host port

You can also press the **Ctrl** key while you select several host ports to delete (see Figure 8-61).

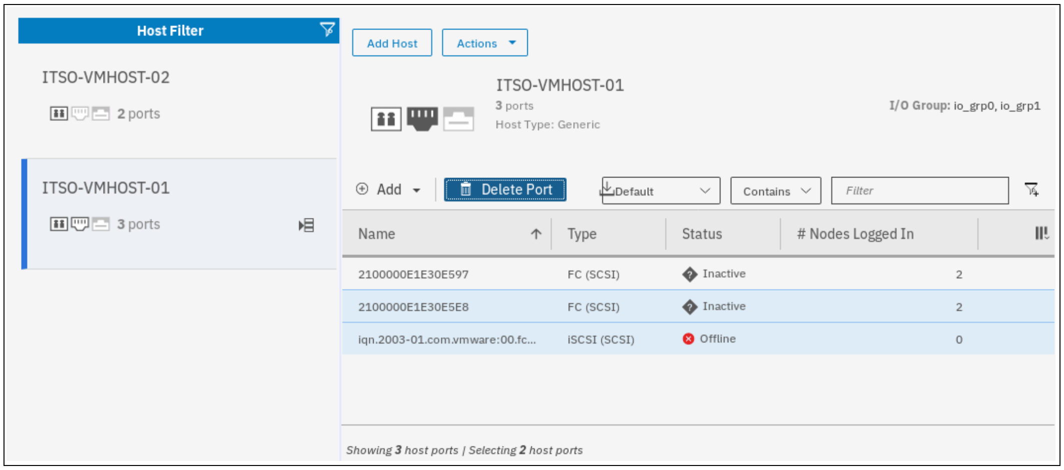


Figure 8-61 Delete several host ports

2. Click **Delete** and confirm the number of host ports that you want to remove by entering that number in the **Verify** field (see Figure 8-62).

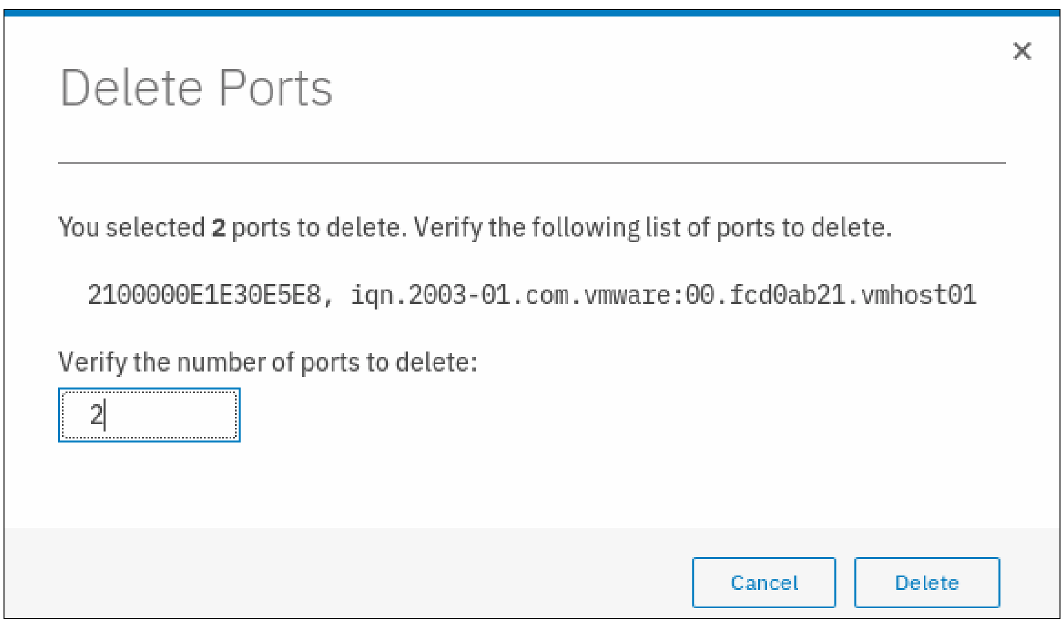


Figure 8-62 Enter the number of host ports to delete

3. Click **Delete** to apply the changes and then, click **Close**.

Note: Deleting FC (including NVMe) and iSCSI ports is done the same way.

8.4.5 Host mappings overview

To see an overview of the host mappings, click **Hosts** → **Mappings**, as shown in Figure 8-63.

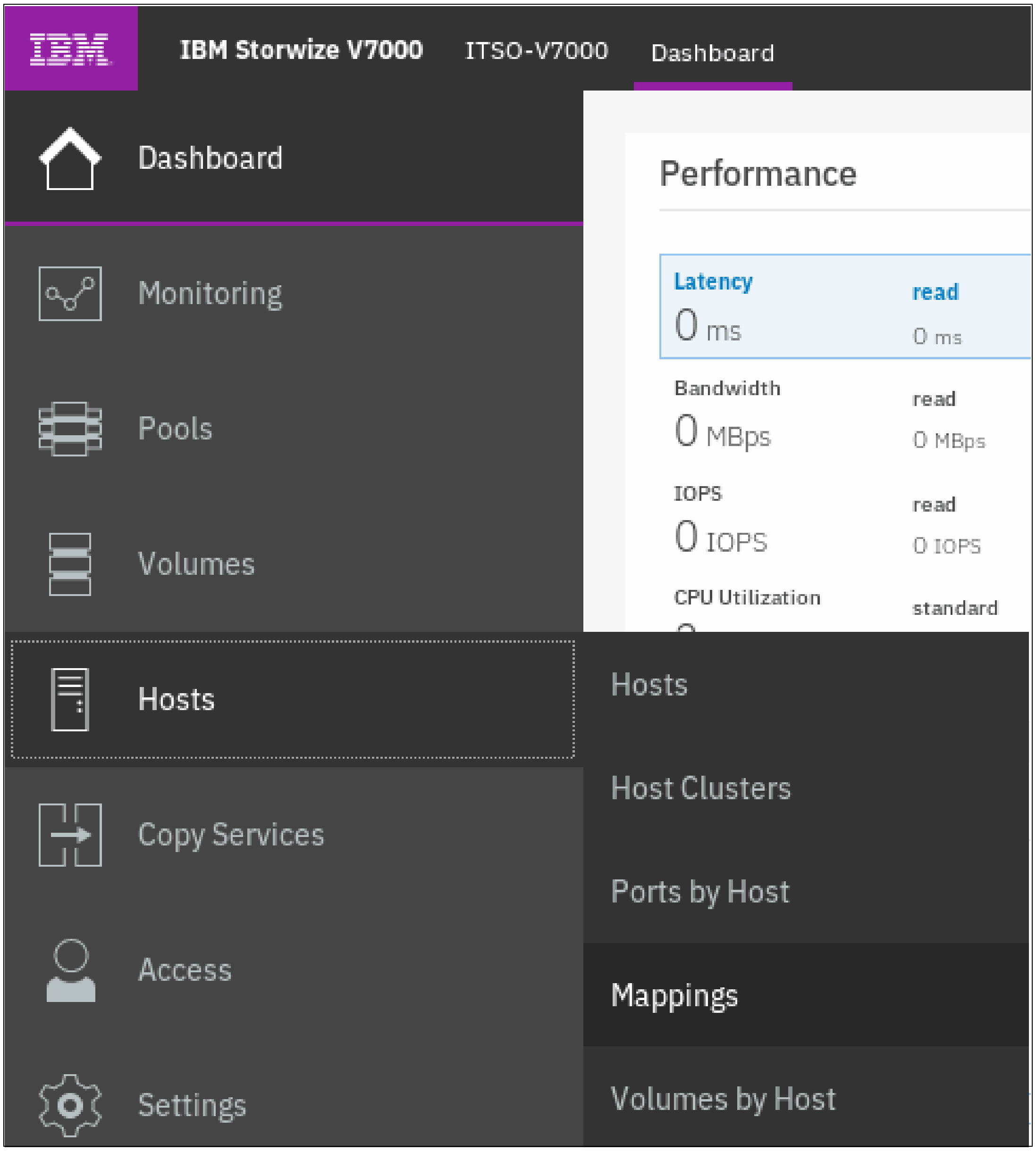


Figure 8-63 Host mappings

A list of all volume mappings is displayed, as shown in Figure 8-64.

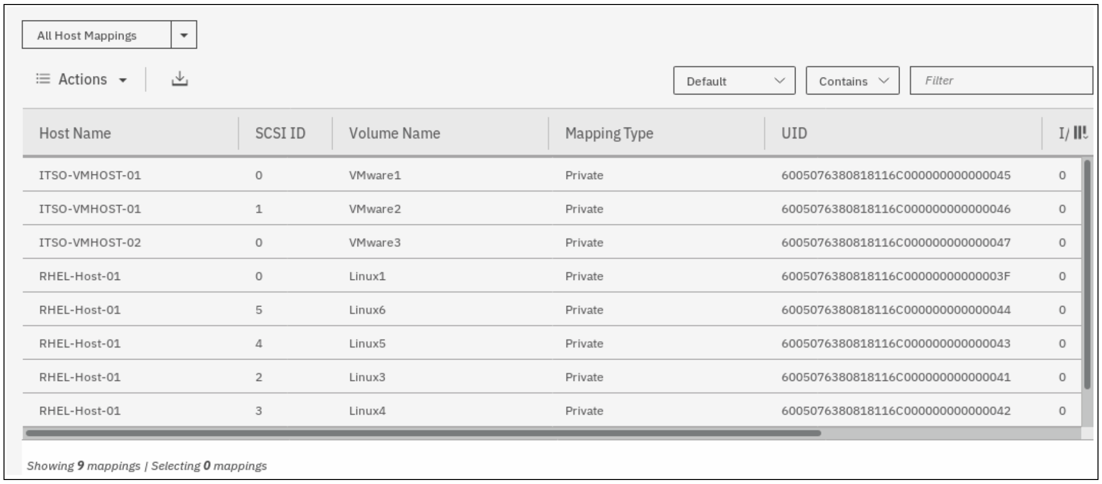


Figure 8-64 Mappings list

This window lists all hosts and volumes. This example shows that the host `ITSO-VMHOST-01` has two mapped volumes, and their associated SCSI ID, Volume Name, and Volume Unique Identifier (UID). If you have more than one caching I/O group, you also see which volume is handled by which I/O group.

If you select one line and click **Actions** (see Figure 8-65), the following tasks are available:

- Unmap Volumes
- Properties (Host)
- Properties (Volume)

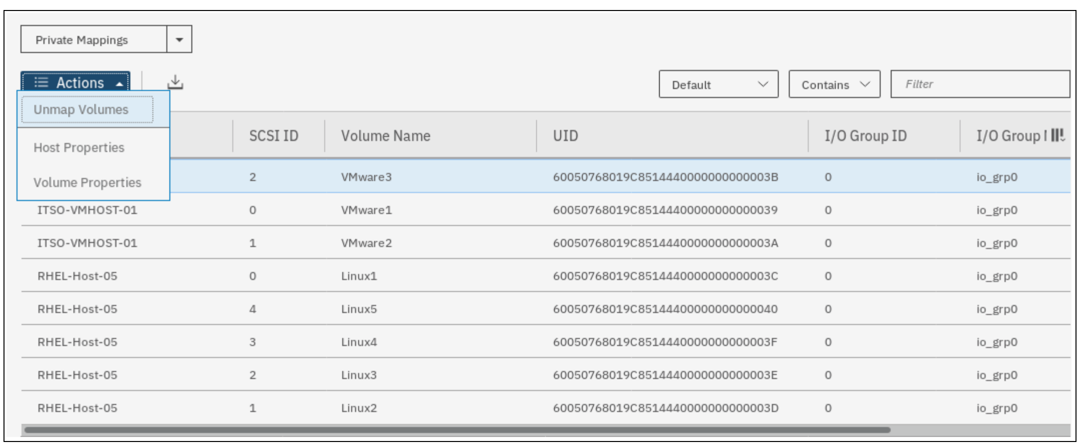


Figure 8-65 Host Mappings Actions menu

Unmapping a volume

This action removes the mappings for all selected entries. From the **Actions** menu that is shown in Figure 8-65 on page 372, select one or more lines (while holding the **Ctrl** key). Click **Unmap Volumes**. Confirm how many volumes are to be unmapped by entering that number in the **Verify** field (see Figure 8-66), and then, click **Unmap**.

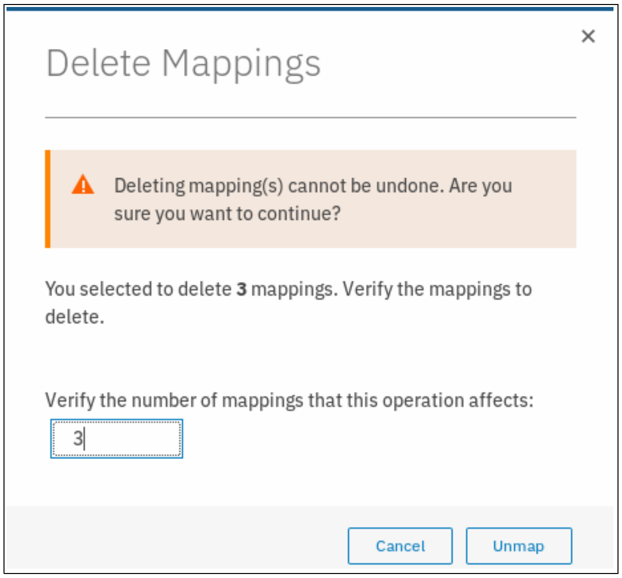


Figure 8-66 Unmap selected volumes

Properties (Host)

Selecting an entry and clicking **Properties (Host)**, as shown in Figure 8-65 on page 372, opens the host properties window. The contents of this window are described in "Host properties" on page 357.

Properties (Volume)

Selecting an entry and clicking **Properties (Volume)**, as shown in Figure 8-65 on page 372, opens the volume properties view. The contents of this window are described in Chapter 7, "Volumes" on page 241.

8.5 Hosts operations by using command-line interface

This section describes some of the host-related actions that can be taken within the Storwize V7000 system by using the command-line interface (CLI).

8.5.1 Create a host by using CLI

This section describes how to create Fibre Channel and iSCSI hosts by using the IBM Spectrum Virtualize CLI. It is assumed that hosts are prepared for attachment, as described in the Storwize V7000 host attachment section of IBM Knowledge Center.

Creating Fibre Channel hosts

To create a Fibre Channel host, complete the following steps:

1. Rescan the SAN on Storwize V7000 by using the **detectmdisk** command (see Example 8-14).

   Example 8-14 Rescanning the SAN

   ```
   IBM_Storwize:ITSO-V7000:superuser>detectmdisk
   ```

   Note: The **detectmdisk** command does not return any response.

   If the zoning was implemented correctly, any new WWPNs are discovered by the Storwize V7000 system after running the **detectmdisk** command.

2. List the candidate WWPNs and identify the WWPNs belonging to the new host, as shown in Example 8-15.

   Example 8-15 Available WWPNs

   ```
   IBM_Storwize:ITSO-V7000:superuser>lsfcportcandidate
   fc_WWPN
   2100000E1E09E3E9
   2100000E1E30E5E8
   2100000E1E30E60F
   2100000E1EC2E5A2
   2100000E1E30E597
   2100000E1E30E5EC
   ```

3. Run the **mkhost** command with the required parameters, as shown in Example 8-16.

   Example 8-16 Host creation

   ```
   IBM_Storwize:ITSO-V7000:superuser>mkhost -name ITSO-VMHOST-03 -fcwwpn
   2100000E1E30E597:2100000E1E30E5EC
   Host, id [3], successfully created
   IBM_Storwize:ITSO-V7000:superuser>
   ```

Creating iSCSI hosts

Before you create an iSCSI host in Storwize V7000, the iSCSI qualified name (IQN) address of the host must be known. See your host operating system-specific documentation to find the IQN of the host.

Create a host by completing the following steps:

1. Create the iSCSI host by using the **mkhost** command (see Example 8-17).

   Example 8-17 Creating an iSCSI host by using the mkhost command

   ```
   IBM_Storwize:ITSO-V7000:superuser>mkhost -iscsiname
   iqn.1994-05.com.redhat:e6ff477b58 -name RHEL-Host-06
   Host, id [4], successfully created
   IBM_Storwize:ITSO-V7000:superuser>
   ```

2. The iSCSI host can be verified by using the **lshost** command, as shown in Example 8-18.

   Example 8-18 Verifying the iSCSI host by using the lshost command

   ```
   IBM_Storwize:ITSO-V7000:superuser>lshost 4
   ```

```
id 4
name RHEL-Host-06
port_count 1
type generic
mask 1111111111111111111111111111111111111111111111111111111111111111
iogrp_count 4
status offline
site_id
site_name
host_cluster_id
host_cluster_name
protocol scsi
iscsi_name iqn.1994-05.com.redhat:e6ff477b58
node_logged_in_count 0
state offline
IBM_Storwize:ITSO-V7000:superuser>
```

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 Storwize V7000 system, use the **chhost** command with the **chapsecret** parameter. If you must display a CHAP secret for a defined server, use the **lsiscsiauth** command. The **lsiscsiauth** command lists the CHAP secret that is configured for authenticating an entity to the Storwize V7000 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 need to know the NVMe qualified name (NQN) address of the host. See your host operating system-specific documentation to find the NQN of the host.

Create a host by completing the following steps:

1. Create the NVMe host by using the **mkhost** command, as shown in Example 8-19.

Example 8-19

```
IBM_Storwize:ITSO-V7000:superuser>svctask mkhost -name NVMe-Host-01 -nqn
nqn.2014-08.com.redhat:nvme:nvm-nvmehost01-edf223876 -protocol nvme -type generic
Host, id [6], successfully created
IBM_Storwize:ITSO-V7000:superuser>
```

2. The NVMe host can be verified by using the **lshost** command, as shown in Example 8-20.

Example 8-20

```
IBM_Storwize:ITSO-V7000:superuser>lshost 6
id 6
name NVMe-Host-01
port_count 1
type generic
mask 1111111111111111111111111111111111111111111111111111111111111111
iogrp_count 4
status offline
site_id
```

```
site_name
host_cluster_id
host_cluster_name
protocol nvme
nqn nqn.2014-08.com.redhat:nvme:nvm-nvmehost01-edf223876
node_logged_in_count 0
state offline
IBM_Storwize:ITSO-V7000:superuser>
```

8.5.2 Advanced 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 already mapped to a different host
- Unmapping a volume from a host
- Renaming 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, use the **mkvdiskhostmap** command, as shown in Example 8-21.

Example 8-21 Mapping a volume

```
IBM_Storwize:ITSO-V7000:superuser>mkvdiskhostmap -host RHEL_HOST -scsi 0
RHEL_VOLUME
Virtual Disk to Host map, id [0], successfully created
```

2. The volume mapping can then be checked by issuing the **lshostvdiskmap** command against that host, as shown in Example 8-22.

Example 8-22 Checking the mapped volume

```
IBM_Storwize:ITSO-V7000:superuser>lshostvdiskmap RHEL_HOST
id name      SCSI_id vdisk_id vdisk_name  vdisk_UID
IO_group_id IO_group_name mapping_type host_cluster_id host_cluster_name
7  RHEL_HOST 0     109     RHEL_VOLUME 600507680C81825B0000000000000154 0
io_grp0       private
```

Mapping a volume that is already mapped to a different host

To map a volume to another host that is mapped to one host, complete the following steps:

1. Issue the **mkvdiskhost -force** command, as shown in Example 8-23.

Example 8-23 Mapping the same volume to a second host

```
IBM_Storwize:ITSO-V7000:superuser>svctask mkvdiskhostmap -force -host
RHEL-Host-06 -scsi 0 Linux1
Virtual Disk to Host map, id [0], successfully created
IBM_Storwize:ITSO-V7000:superuser>
```

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 (MSCS), IBM PowerHA, and VMware ESX clustering.

2. The volume RHEL_VOLUME is mapped to two hosts (RHEL_HOST and iSCSI_RHEL_HOST), and can be seen by issuing the **lsvdiskhostmap** command, as shown in Example 8-24.

Example 8-24 Ensuring the same volume is mapped to multiple hosts

```
IBM_Storwize:ITSO-V7000:superuser>lsvdiskhostmap Linux1
id name   SCSI_id host_id host_name    vdisk_UID
IO_group_id IO_group_name mapping_type host_cluster_id host_cluster_name
protocol
11 Linux1 0      3      RHEL-Host-05 60050768019C8514440000000000003C 0
io_grp0       private                                    scsi
11 Linux1 0      4      RHEL-Host-06 60050768019C8514440000000000003C 0
io_grp0       private                                    scsi
IBM_Storwize:ITSO-V7000:superuser>
```

Unmapping a volume from a host

To unmap a volume from the host, the **rmvdiskhostmap** command is used, as shown in Example 8-25.

Example 8-25 Unmapping a volume from a host

```
IBM_Storwize:ITSO-V7000:superuser>rmvdiskhostmap -host RHEL-Host-06 Linux1
IBM_Storwize:ITSO-V7000:superuser>
```

Important: Before unmapping a volume from a host on Storwize V7000, ensure that the host side action is completed on that volume by using the respective host operating system 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, issue the **chhost -name** command, as shown in Example 8-26. In this example, the host RHEL_HOST is renamed to FC_RHEL_HOST.

Example 8-26 Renaming a host

```
IBM_Storwize:ITSO-V7000:superuser>chhost -name RHEL-Host-07 RHEL-Host-06
IBM_Storwize:ITSO-V7000:superuser>
```

Removing a host

To remove a host from the Storwize V7000, the **rmhost** command is used, as shown in Example 8-27.

Example 8-27 Removing a host

```
IBM_Storwize:ITSO-V7000:superuser>rmhost RHEL-Host-07
IBM_Storwize:ITSO-V7000:superuser>
```

Note: Before removing a host from Storwize V7000, ensure that all of the volumes are unmapped from that host, as shown in Example 8-25.

Host properties

To get more information about a host, the **lshost** command can be used with the `hostname` or `host id` as a parameter, as shown in Example 8-28.

Example 8-28 Host details

```
IBM_Storwize:ITSO-V7000:superuser>lshost ITSO-VMHOST-01
id 0
name ITSO-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 ITSO-ESX-Cluster-01
protocol scsi
WWPN 2100000E1E30E597
node_logged_in_count 0
state offline
WWPN 2100000E1E30E5E8
node_logged_in_count 0
state offline
IBM_Storwize:ITSO-V7000:superuser>
```

8.5.3 Adding and deleting a host port by using the CLI

This section describes adding and deleting a host port to and from Storwize V7000.

Adding ports to a defined host

If an HBA is added, or if a network interface controller (NIC) is to be added to a server that is defined within the Storwize V7000, the **addhostport** command can be used to add the new port definitions to the host configuration.

If the host is connected through SAN with FC, and if the WWPN is zoned to the Storwize V7000 system, issue the **lshbaportcandidate** command to compare it with the information that is available from the server administrator, as shown in Example 8-29.

Example 8-29 Listing the new available WWPN

```
IBM_Storwize:ITSO-V7000:superuser>lsfcportcandidate
fc_WWPN
2100000E1E09E3E9
2100000E1E30E5E8
2100000E1E30E60F
2100000E1EC2E5A2
IBM_Storwize:ITSO-V7000:superuser>
```

Use host or SAN switch utilities to verify whether the WWPN matches the information for the new WWPN. If the WWPN matches, use the **addhostport** command to add the port to the host, as shown in Example 8-30.

Example 8-30 Adding the newly discovered WWPN to the host definition

```
IBM_Storwize:ITSO-V7000:superuser>addhostport -hbawwpn 2100000E1E09E3E9:2100000E1E30E5E8
ITSO-VMHOST-01
IBM_Storwize:ITSO-V7000:superuser>
```

This command adds the WWPNs `2100000E1E09E3E9` and `2100000E1E30E5E8` to the ITSO-VMHOST-01 host.

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 8-31.

Example 8-31 Adding a WWPN to the host definition using the -force option

```
IBM_Storwize:ITSO-V7000:superuser>addhostport -hbawwpn 2100000000000001 -force
ITSO-VMHOST-01
IBM_Storwize:ITSO-V7000:superuser>
```

This command forces the addition of the WWPN 2100000000000001 to the host called ITSO-VMHOST-01.

Note: WWPNs are not case-sensitive within the CLI.

The host port count can be verified by running the **lshost** command again. The host ITSO-VMHOST-01 has an updated port count of 3, as shown in Example 8-32.

Example 8-32 Host with updated port count

```
IBM_Storwize:ITSO-V7000:superuser>lshost
id name           port_count iogrp_count status   site_id site_name host_cluster_id
host_cluster_name protocol
0  ITSO-VMHOST-02 0          2           offline
scsi
1  ITSO-VMHOST-01 3          2           degraded
scsi
2  RHEL-Host-01   1          2           offline
scsi
3  ITSO-VMHOST-03 2          2           online
scsi
IBM_Storwize:ITSO-V7000:superuser>
```

If the host uses iSCSI as a connection method, the new iSCSI IQN ID is used to add the port. Unlike FC-attached hosts, with iSCSI, available candidate ports cannot be checked.

After getting the other iSCSI IQN, issue the **addhostport** command, as shown in Example 8-33.

Example 8-33 Adding an iSCSI port to the defined host

```
IBM_Storwize:ITSO-V7000:superuser>addhostport -iscsiname
iqn.1994-05.com.redhat:e6ddffaab567 RHEL-Host-05
IBM_Storwize:ITSO-V7000:superuser>
```

If the host uses NVMe as a connection method, the new NVMe NQN ID is used to add the port. Unlike SCSI FC-attached hosts, with NVMe, available candidate ports cannot be checked.

After getting the other NVMe NQN, issue the **addhostport**, command as shown in Example 8-34.

Example 8-34

```
IBM_Storwize:ITSO-V7000:superuser>addhostport -nqn nqn.2016-06.io.rhel:875adad3345
RHEL-Host-08
IBM_Storwize:ITSO-V7000:superuser>
```

Deleting ports from a defined host

If a mistake is made while adding a port, or if an HBA is removed from a server that is defined within the Storwize V7000, use the **rmhostport** command to remove WWPN definitions from a host.

Before removing the WWPN, ensure that it is the correct WWPN by issuing the **lshost** command, as shown in Example 8-35.

Example 8-35 Using the lshost command to check the WWPNs

```
IBM_2145:ITSO-SV1:superuser>lshost ITSO-VMHOST-01
id 0
name ITSO-VMHOST-01
port_count 2
type generic
mask 1111111111111111111111111111111111111111111111111111111111111111
iogrp_count 4
status online
site_id
site_name
host_cluster_id 0
host_cluster_name ITSO-ESX-Cluster-01
protocol scsi
WWPN 2100000E1E30E597
node_logged_in_count 2
state online
WWPN 2100000E1E30E5E8
node_logged_in_count 2
state online
IBM_2145:ITSO-SV1:superuser>
```

When the WWPN or iSCSI IQN that needs to be deleted is known, use the **rmhostport** command to delete a host port, as shown in Example 8-36.

Example 8-36 Using the rmhostport command to remove a WWPN

```
IBM_Storwize:ITSO-V7000:superuser>rmhostport -fcwwpn 2100000E1E30E597 ITSO-VMHOST-01
IBM_Storwize:ITSO-V7000:superuser>
```

Use the command to remove the iSCSI IQN, as shown in Example 8-37.

Example 8-37 Removing iSCSI port from the host

```
IBM_Storwize:ITSO-V7000:superuser>rmhostport -iscsiname iqn.1994-05.com.redhat:e6ddffaab567
RHEL-Host-05
IBM_Storwize:ITSO-V7000:superuser>
```

Use the command to remove the NVMe NQN, as shown in Example 8-38.

Example 8-38 Removing NQN port from the host

```
IBM_Storwize:ITSO-V7000:superuser>rmhostport -nqn nqn.2016-06.io.rhel:875adad3345
RHEL-Host-08
IBM_Storwize:ITSO-V7000:superuser>
```

Note: Multiple ports can be removed at one time by using the separator or colon (:) between the port names, as shown in the following example:

```
rmhostport -hbawwpn 210000E08B054CAA:210000E08B892BCD Angola
```

8.5.4 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`)
- Unmapping a volume 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, the command **`mkhostcluster`** can be used, as shown in Example 8-39.

Example 8-39 Creating a host cluster by using mkhostcluster

```
IBM_Storwize:ITSO-V7000:superuser>mkhostcluster -name ITSO-ESX-Cluster-01
Host cluster, id [0], successfully created.
IBM_Storwize:ITSO-V7000:superuser>
```

Note: While creating the host cluster, if you want it to inherit the volumes mapped to a particular host, use the **`-seedfromhost`** flag option. Any volume mapping that does not need to be shared can be kept private by using **`-ignoreseedvolume`** flag option.

Adding a host to a host cluster

After creating a host cluster, a host or a list of hosts can be added by using the **`addhostclustermember`** command, as shown in Example 8-40.

Example 8-40 Adding a host or hosts to a host cluster

```
IBM_Storwize:ITSO-V7000:superuser>addhostclustermember -host ITSO-VMHOST-01:ITSO-VMHOST-02
ITSO-ESX-Cluster-01
IBM_Storwize:ITSO-V7000:superuser>
```

In Example 8-40, the hosts `ITSO-VMHOST-01` and `ITSO-VMHOST-02` were added as part of host cluster `ITSO-ESX-Cluster-01`.

Listing the host cluster member

To list the host members that are part of a particular host cluster, the **`lshostclustermember`** command can be used, as shown in Example 8-41.

Example 8-41 Listing host cluster members with the lshostclustermember command

```
IBM_Storwize:ITSO-V7000:superuser>lshostclustermember ITSO-ESX-Cluster-01
host_id host_name      status  type    site_id site_name
0       ITSO-VMHOST-01 offline generic
4       ITSO-VMHOST-02 offline generic
IBM_Storwize:ITSO-V7000: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, the **mkvolumehostclustermap** command can be used, as shown in Example 8-42.

Example 8-42 Mapping a volume to a host cluster

```
IBM_Storwize:ITSO-V7000:superuser>mkvolumehostclustermap -hostcluster ITSO-ESX-Cluster-01
VMware1
Volume to Host Cluster map, id [0], successfully created
IBM_Storwize:ITSO-V7000: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, the **lshostclustervolumemap** command can be used, as shown in Example 8-43.

Example 8-43 Listing volumes that are mapped to a host cluster by using lshostclustervolumemap

```
IBM_Storwize:ITSO-V7000:superuser>lshostclustervolumemap ITSO-ESX-Cluster-01
id name                SCSI_id volume_id volume_name volume_UID                       IO_group_id
IO_group_name protocol
0  ITSO-ESX-Cluster-01 0       8         VMware1     60050768019C8514440000000000000039 0           io_grp0
scsi
0  ITSO-ESX-Cluster-01 1       9         VMware2     60050768019C851444000000000000003A 0           io_grp0
scsi
0  ITSO-ESX-Cluster-01 2       10        VMware3     60050768019C851444000000000000003B 0           io_grp0
scsi
IBM_Storwize:ITSO-V7000:superuser>
```

> **Note:** The **lshostvdiskmap** command can be run against each host that is part of a host cluster to ensure that the mapping type for the shared volume is `shared`, and is `private` for the non-shared volume.

Removing a volume mapping from a host cluster

To remove a volume mapping to a host cluster, use the **rmvolumehostclustermap** command, as shown in Example 8-44.

Example 8-44 Removing a volume mapping

```
IBM_Storwize:ITSO-V7000:superuser>rmvolumehostclustermap -hostcluster ITSO-ESX-Cluster-01
VMware3
IBM_Storwize:ITSO-V7000:superuser>
```

In Example 8-44, volume VMware3 is unmapped from the host cluster ITSO-ESX-Cluster-01. The current volume mapping can be checked to ensure that it is unmapped, as shown in Example 8-43.

> **Note:** To specify the host or hosts that acquire private mappings from the volume that is being removed from the host cluster, specify the **-makeprivate** flag.

Removing a host cluster member

To remove a host cluster member, use the **rmhostclustermember** command, as shown in Example 8-45.

Example 8-45 Removing a host cluster member

```
IBM_2145:ITSO-SV1:superuser>rmhostclustermember -host ITSO-VMHOST-02 -removemappings
ITSO-ESX-Cluster-01
IBM_2145:ITSO-SV1:superuser>
```

In Example 8-45, the host ITSO-VMHOST-02 was removed as a member from the host cluster ITSO-ESX-Cluster-01, along with the associated volume mappings because of the **-removemappings** flag was specified.

Removing a host cluster

To remove a host cluster, use the **rmhostcluster** command, as shown in Example 8-46.

Example 8-46 Removing a host cluster

```
IBM_Storwize:ITSO-V7000:superuser>rmhostcluster -removemappings ITSO-ESX-Cluster-01
IBM_Storwize:ITSO-V7000:superuser>
```

The use of the **-removemappings** flag also causes the system to remove any host mappings to volumes that are shared. 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, specify 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`.

9

Storage migration

This chapter describes the steps that are involved in migrating data from an external storage system to the capacity of the Storwize V7000 by using the storage migration wizard. Migrating data from other storage systems to the Storwize V7000 consolidates storage. It also allows for IBM Spectrum Virtualize features, such as Easy Tier, thin provisioning, compression, encryption, storage replication, and the easy-to-use graphical user interface (GUI) to be realized across all volumes.

Storage migration uses the volume mirroring functionality to allow reads and writes during the migration, and minimizing disruption and downtime. After the migration is complete, the existing system can be retired. Storwize V7000 supports migration through Fibre Channel and internet Small Computer Systems Interface (iSCSI) connections. Storage migration can be used to migrate data from other storage systems and IBM SAN Volume Controller.

This chapter includes the following topics:

> **Note:** This chapter does not cover migration outside of the storage migration wizard. To migrate data outside of the wizard, you must use Import. For more information about the Import action, see Chapter 6, “Storage pools” on page 191.

9.1 Storage migration overview

To migrate data from a storage system to the Storwize V7000, it is necessary to use the built-in external virtualization capability. This capability places external connected Logical Units (LUs) under the control of the Storwize V7000. After volumes are virtualized, hosts continue to access them, but do so through the Storwize V7000, which acts as a proxy.

> **Attention:** The system does not require a license for its own control and expansion enclosures. However, a license is required for each enclosure of any external systems that are being virtualized. Data can be migrated from storage systems to your system by using the external virtualization function within 45 days of purchase of the system without purchase of a license. After 45 days, any ongoing use of the external virtualization function requires a license for each enclosure in each external system.
>
> 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 systems divide storage into many Small Computer System Interface LUs that are presented to hosts.
- I/O to the LUs must be stopped and changes made to the mapping of the external storage system LUs and to the fabric configuration so that the original LUs are presented directly to the Storwize V7000 and not to the hosts anymore. The Storwize V7000 discovers the external LUs as *unmanaged* MDisks.
- The unmanaged MDisks are *imported* to the Storwize V7000 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 system and the data remains unchanged. The Storwize V7000 is presenting active images of the LUs and is acting as a proxy.
- The hosts must have the storage system multipath device driver removed, and are then configured for Storwize V7000 attachment. The Storwize V7000 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 Storwize V7000 volumes through a host rescan or reboot operation.
- IBM Spectrum Virtualize volume mirroring operations are then initiated. The image-mode volumes are mirrored to generic 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.

9.1.1 Interoperability and compatibility

Interoperability is an important consideration when a new storage system is set up in an environment that contains storage infrastructure. Before attaching any external storage systems to the Storwize V7000, see the IBM System Storage Interoperation Center (SSIC).

At the SSIC site, select **IBM System Storage Midrange Disk** as the Storage Family, then **Storwize V7000 Storage Controller Support** as the Storage Model. You can then refine your search by selecting the external storage controller that you want to use in the **Storage Controller** menu.

The matrix results indicate the external storage that you want to attach to the Storwize V7000, such as minimum firmware level or support for disks greater than 2 TB.

9.1.2 Prerequisites

Before the storage migration wizard can be started, the external storage system must be visible to the Storwize V7000. You also need to confirm that the restrictions and prerequisites are met.

Administrators can migrate data from the external storage system to the system that uses iSCSI connections and Fibre Channel or Fibre Channel over Ethernet connections.

Prerequisites for Fibre Channel connections

The following prerequisites for Fibre Channel connections must be met:

- Cable this system into the SAN of the external storage that you want to migrate. Ensure that your system is cabled into the same storage area network (SAN) as the external storage system that you are migrating.

  If you are using Fibre Channel, connect the Fibre Channel cables to the Fibre Channel ports in *both* canisters of your system, and then to the Fibre Channel network. If you are using Fibre Channel over Ethernet, connect Ethernet cables to the 10 Gbps Ethernet ports.

- Change VMware ESX host settings, or do not run VMware ESX. If you have VMware ESX server hosts, you must change settings on the VMware host so copies of the volumes can be recognized by the system after the migration is completed. To enable volume copies to 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 LUN has a different unique identifier after it is imported. It resembles a mirrored volume to the VMware server.

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 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, extra redundant iSCSI connection can be established for increased throughput.

- Change VMware ESX host settings, or do not run VMware ESX. If you have VMware ESX server hosts, you must change settings on the VMware host so copies of the volumes can be recognized by the system after the migration is completed. To enable volume copies to be recognized by the system for VMware ESX hosts, 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 LUN has a different unique identifier after it is imported. It appears as a mirrored volume to the VMware server.

If the external storage system is not detected, the warning message that is shown in Figure 9-1 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.

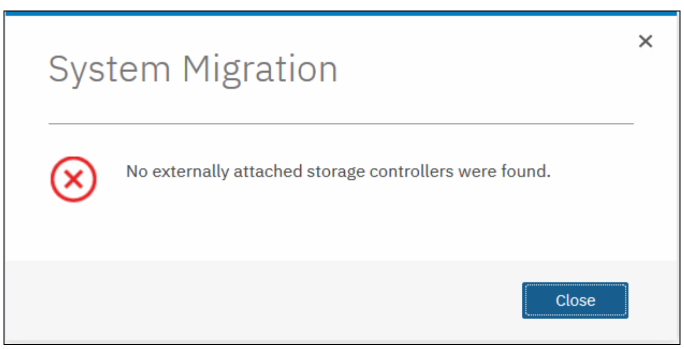


Figure 9-1 Error message if no external storage is detected

9.2 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 the storage migration wizard is used 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 systems, and the Storwize V7000 before the wizard is used.

Complete the following steps to complete the migration by using the storage migration wizard:

1. Navigate to **Pools** → **System Migration**, as shown in Figure 9-2. The System Migration pane provides access to the storage migration wizard and displays information about the migration progress.

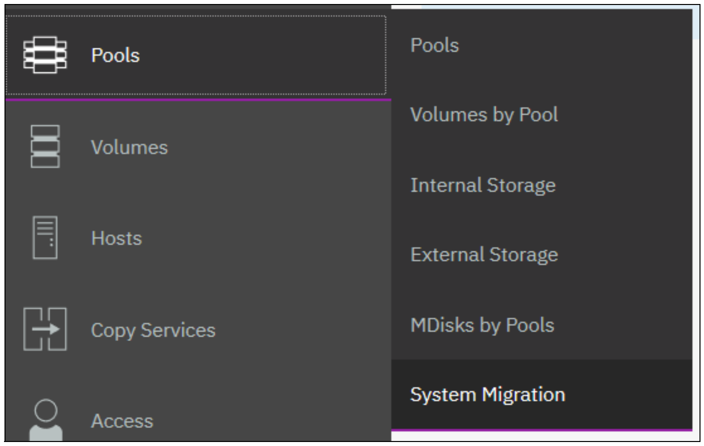


Figure 9-2 Navigating to Storage Migration

2. Click **Start New Migration** to begin the storage migration wizard, as shown in Figure 9-3.

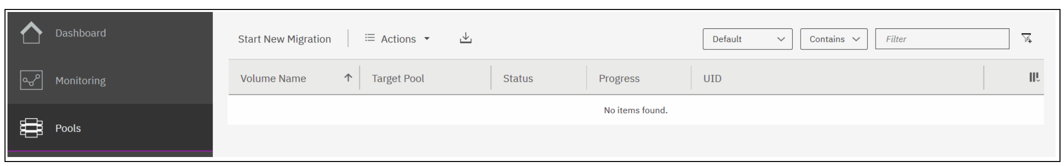


Figure 9-3 Starting a migration

> **Note:** Starting a new migration adds a volume to be migrated in the list displayed in the pane. After a volume is migrated, it remains in the list until you “finalize” the migration.

3. If both Fibre Channel and iSCSI external systems are detected, a dialog is shown asking you which protocol should be used. Select the type of attachment between the Storwize V7000 and the external system from which you want to migrate volumes and click **Next**. If only one type of attachment is detected, this dialog is not displayed.

4. When the wizard starts, you are prompted to verify the restrictions and prerequisites that are listed in Figure 9-4 on page 390. 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 (VIOS).

     • 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:

  - Storwize V7000 and the external storage system 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 9.1.2, “Prerequisites” on page 387.

If all restrictions are satisfied and prerequisites are met, select all of the options and click **Next**, as shown in Figure 9-4.

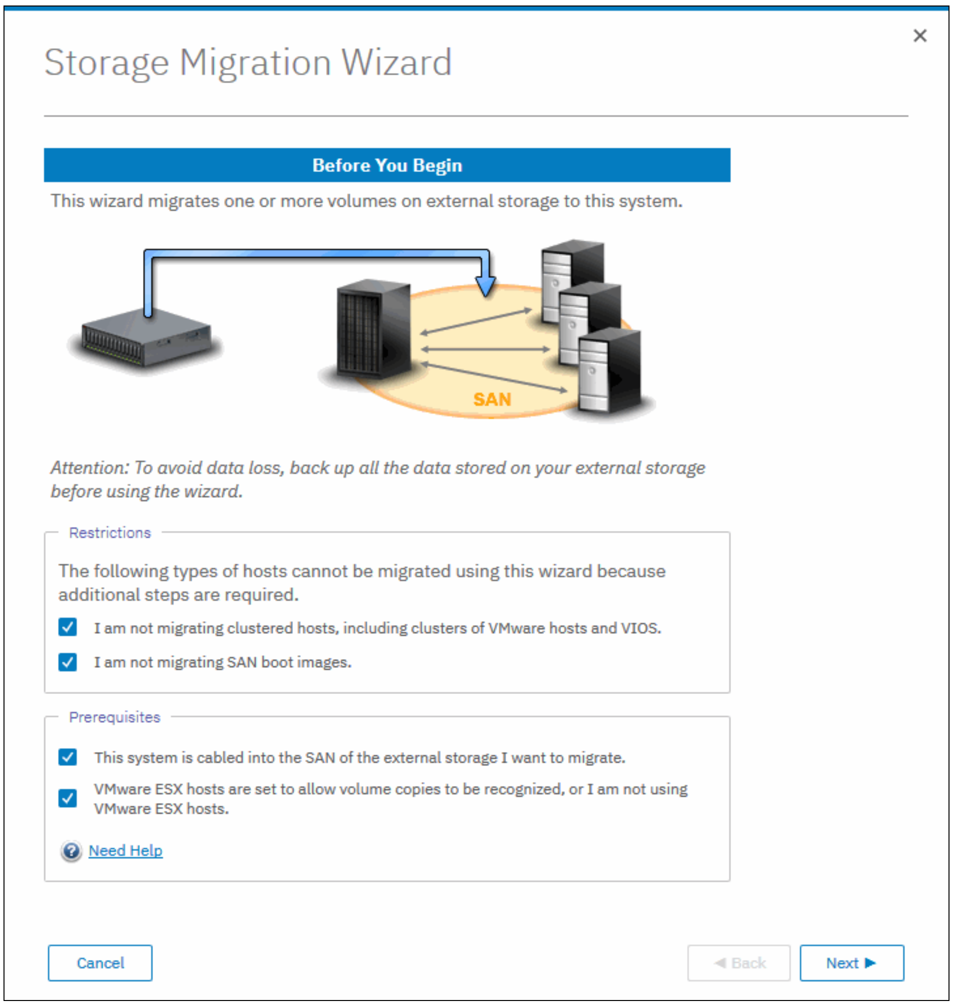


Figure 9-4 Restrictions and prerequisites confirmation

5. Prepare the environment migration by following the instructions that are shown in Figure 9-5.

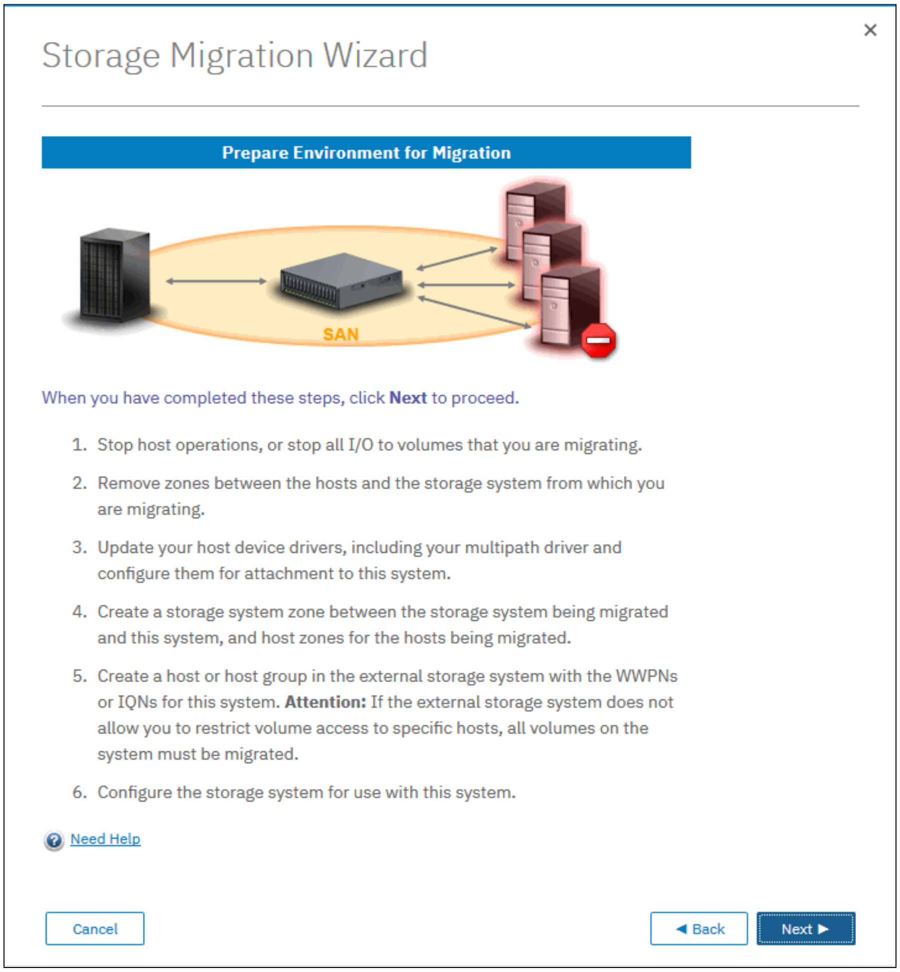


Figure 9-5 Preparing your environment for storage migration

The preparation phase includes the following steps:

a. Before migrating 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 system you are migrating.

c. Hosts usually do not support concurrent multipath drivers at the same time. You might need to remove drivers that are not compatible with the Storwize V7000, from the hosts and use the recommended device drivers. For more information about supported drivers, see the IBM SSIC.

d. If you are migrating external storage systems that connect to the system that uses Fibre Channel or Fibre Channel over Ethernet connections, ensure that you complete appropriate zoning changes to simplify migration. Use the following guidelines to ensure that zones are configured correctly for migration:

   - Zoning rules

     For every storage system, create one zone that contains this system's ports from every node and all external storage system ports, unless otherwise stated by the zoning guidelines for that storage system.

This system requires single-initiator zoning for all large configurations that contain more than 64 host objects. Each server Fibre Channel port must be in its own zone, which contains the Fibre Channel port and this system's ports. In configurations of fewer than 64 hosts, you can have up to 40 Fibre Channel ports in a host zone if the zone contains similar HBAs and operating systems.

- 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 Fibre Channel port.

– Because the Storwize V7000 should now be seen as a host from the external system to be migrated, you must define the Storwize V7000 as a host or host group by using the WWPNs or IQNs on the system to be migrated. Some legacy systems do not permit LUN-to-host mapping and would then present all the LUs to the Storwize V7000. In that case, all the LUs should be migrated.

6. If the previous preparation steps were followed, the Storwize V7000 is now seen as a host from the system to be migrated. LUs can then be mapped to the Storwize V7000. Map the external storage system by following the instructions that are shown in Figure 9-6.

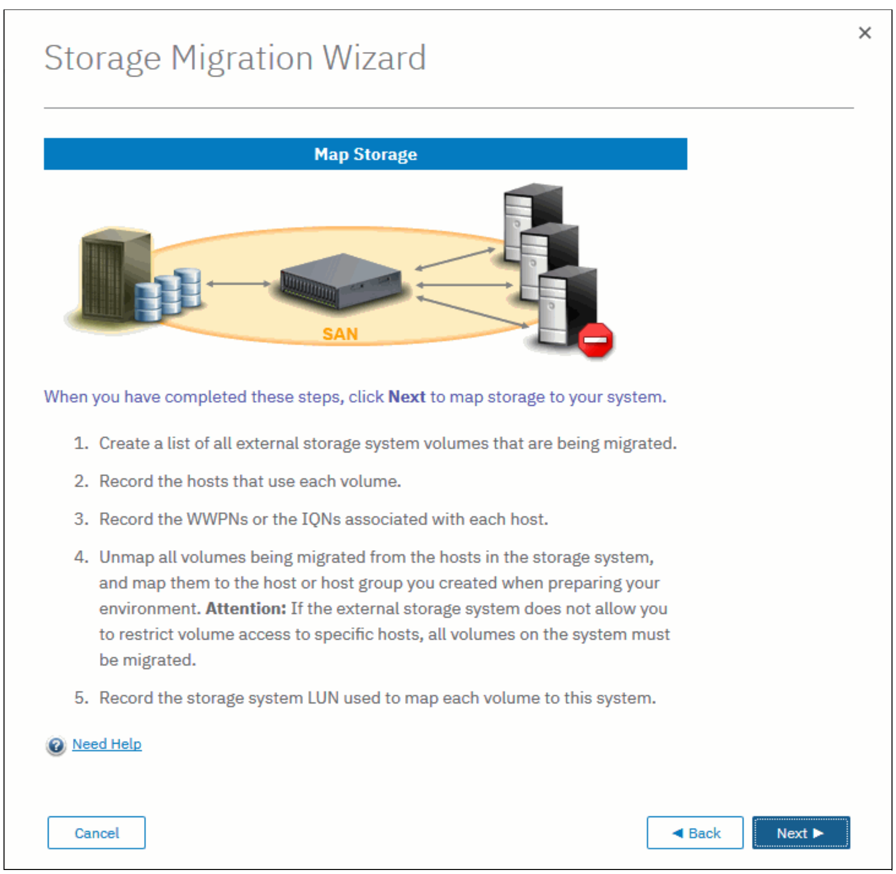


Figure 9-6 Steps to map the LUs to be migrated to the Storwize V7000

Before you migrate storage, record the hosts and their WWPNs or IQNs for each volume that is being migrated and the SCSI LUN when mapped the Storwize V7000.

Table 9-1 shows an example of a table that is used to capture information that relates to the external storage system LUs.

Table 9-1 Example table for capturing external LU information

Volume Name or ID	Hosts accessing this LUN	Host WWPNs or IQNs	SCSI LUN when mapped
1 IBM DB2® logs	DB2server	21000024FF2...	0
2 DB2 data	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 operating systems do not support changing the SCSI ID during the migration.

Click **Next** and wait for the system to discover external devices.

7. The next window shows all of 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 the previous step to trigger the discovery procedure again.

   Select the MDisks that you want to migrate, as shown in Figure 9-7. In this example, one MDisk (`mdisk3`) was found and is migrated. Detailed information about an MDisk is available by double-clicking it. To select multiple elements from the table, press **Shift** and then left-click or **Ctrl** and then left-click. Optionally, you can export the discovered MDisks list to a CSV file for further use by clicking **Export to CSV**.

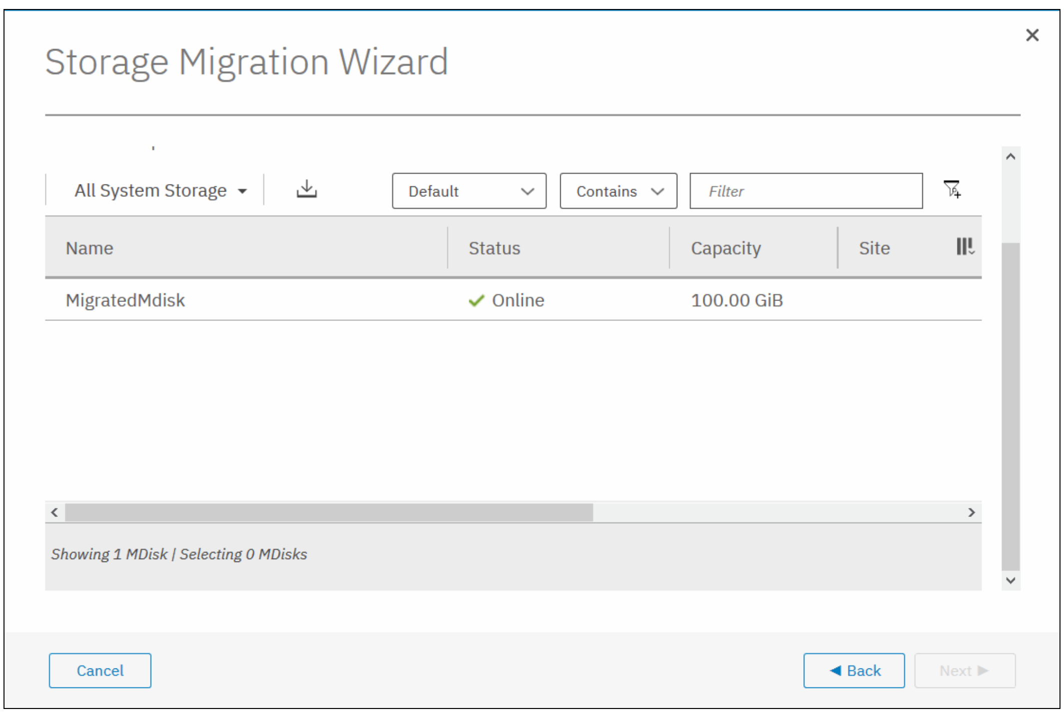


Figure 9-7 Discovering mapped LUs from external storage

Note: Select only the MDisks that are applicable to the current migration plan. After step 15 on page 399 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 called `MigrationPool_XXXX` and adds the imported MDisk to the storage pool as image-mode volumes.
9. The next page lists all of the hosts that are configured on the system and enables you to configure new hosts. This step is optional and can be bypassed by clicking **Next**. In this example, the host `ISCI_Host` is already configured, as shown in Figure 9-8. If no host is selected, you can create a host after the migration completes and map the imported volumes to it.

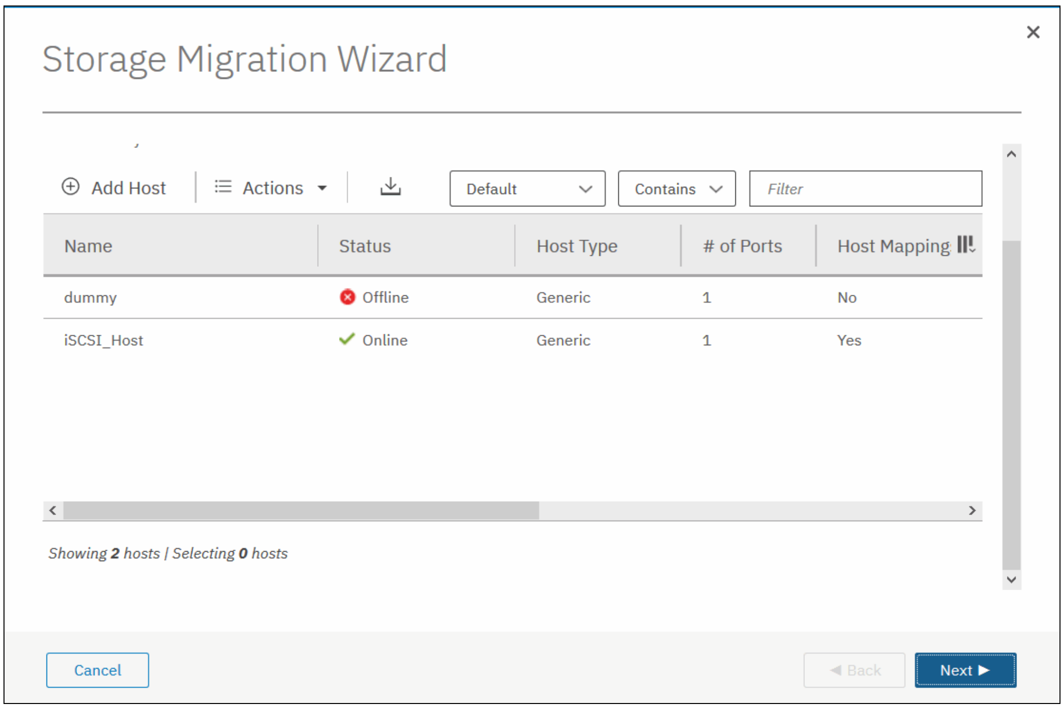


Figure 9-8 Listing of configured hosts to map the imported Volume to

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 9-9 shows the Add Host wizard with the details completed.

For more information about the Add Host wizard, see Chapter 8, “Hosts” on page 317.

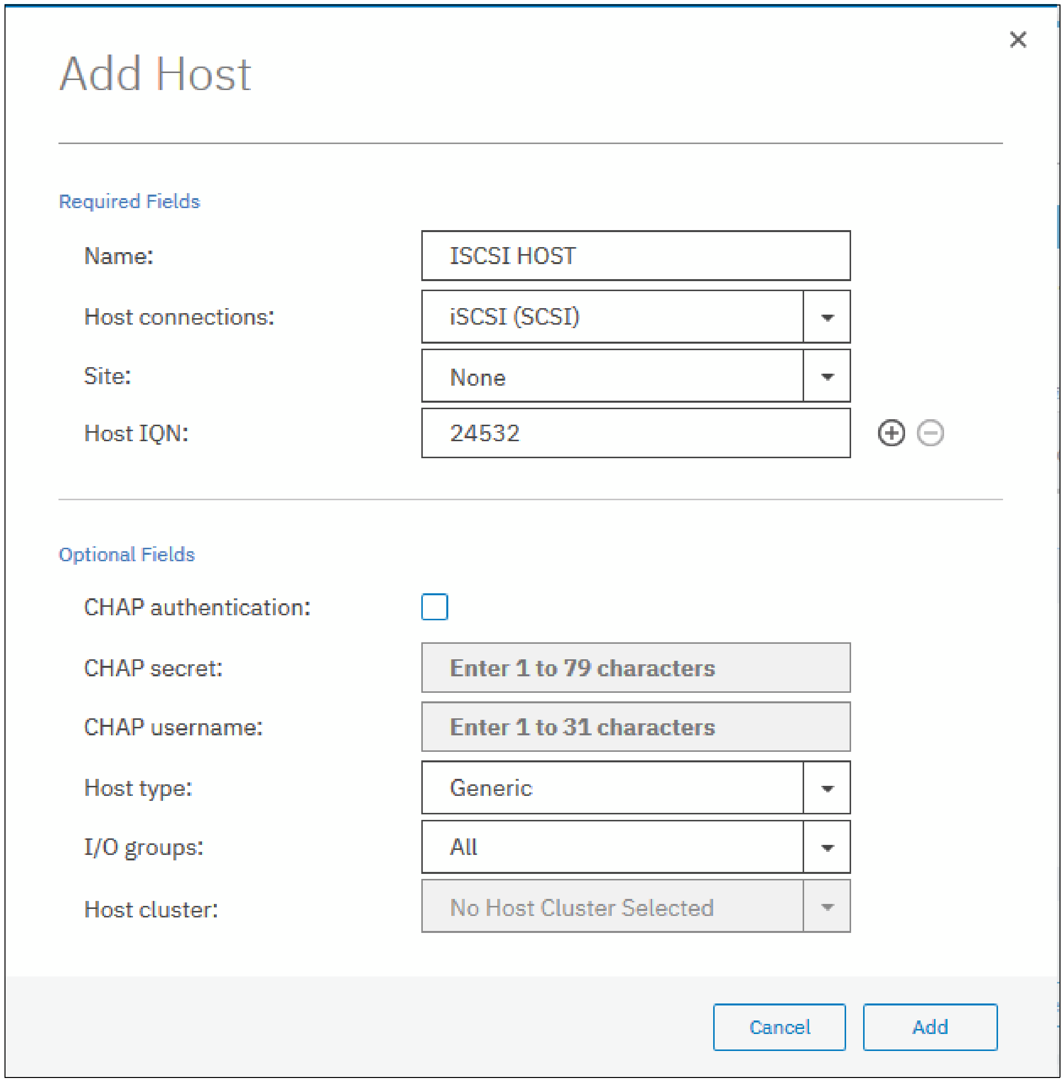


Figure 9-9 If not already defined, you can create a host during the migration process

11.Click **Add**. The host is created and is now listed in the Configure Hosts window, as shown in Figure 9-8 on page 394. Click **Next** to proceed.

12.The next window lists the new volumes and enables you to map them to hosts. 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.

13. Map the volumes to hosts by selecting the volumes and clicking **Map to Host**, as shown in Figure 9-10. This step is optional and can be bypassed by clicking **Next**.

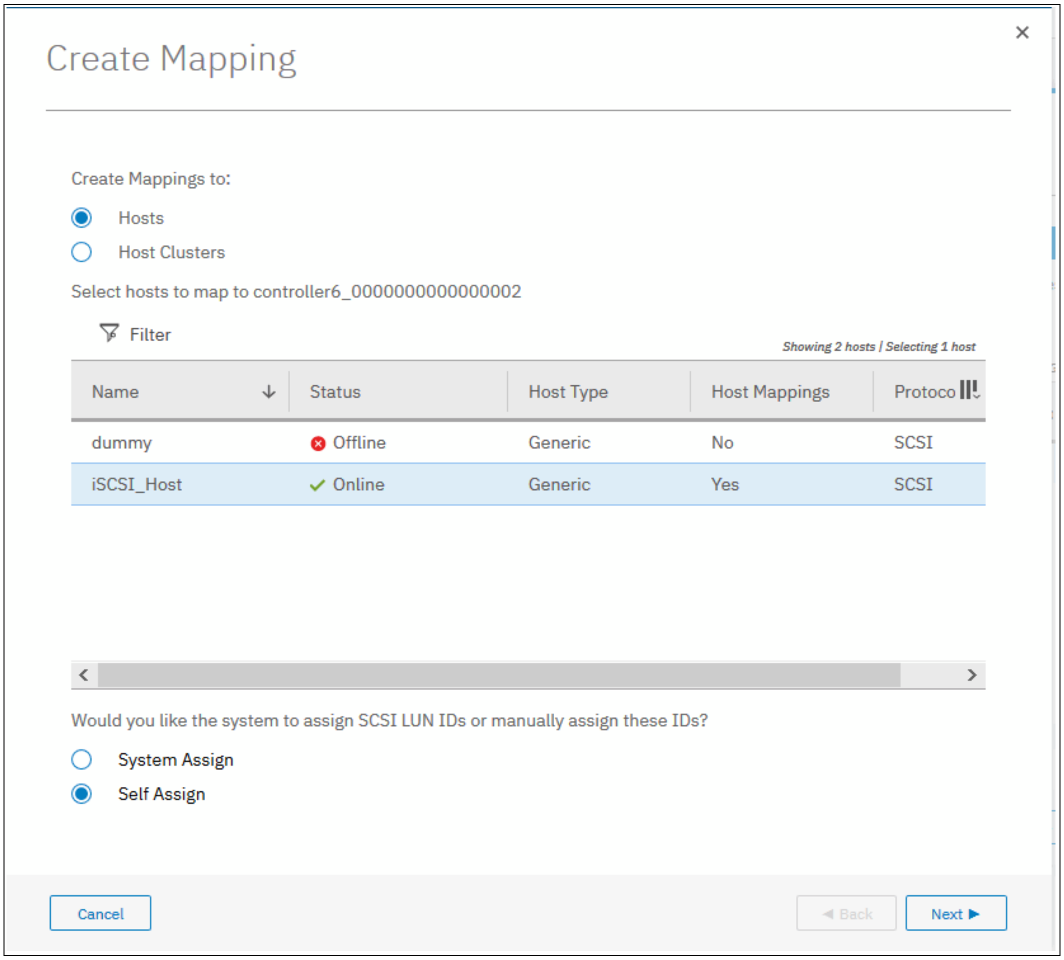


Figure 9-10 Select the host to map the new Volume to

You can manually assign a SCSI ID to the LUNs you are mapping. This technique is particularly 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 9-11.

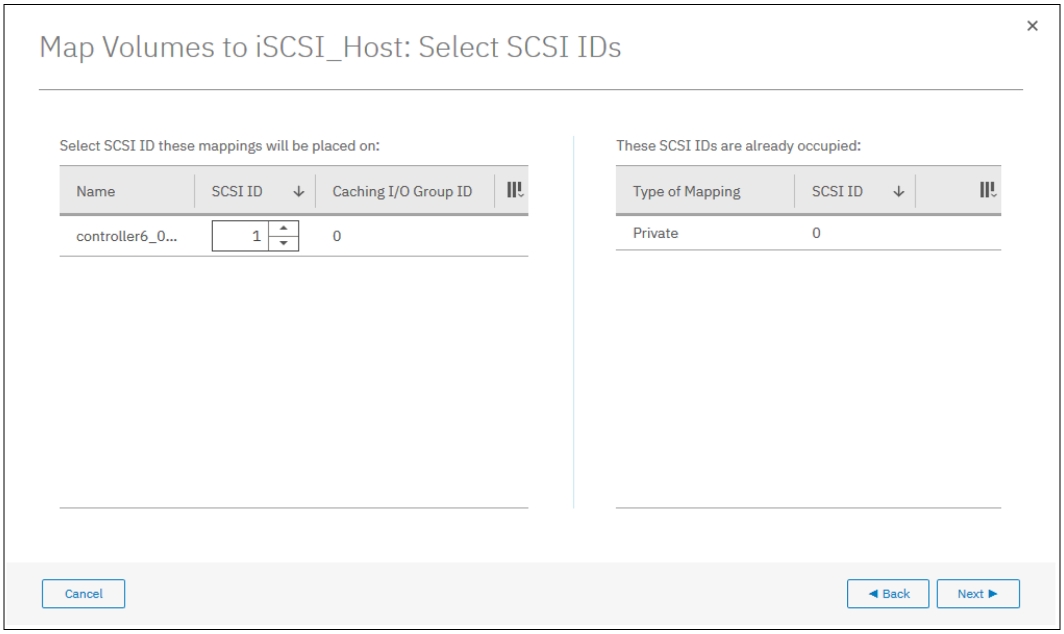


Figure 9-11 Manually assign a LUN SCSI ID to mapped Volume

When your LUN mapping is ready, click **Next**. A new dialog is displayed with a summary of the new and existing mappings, as shown in Figure 9-12.

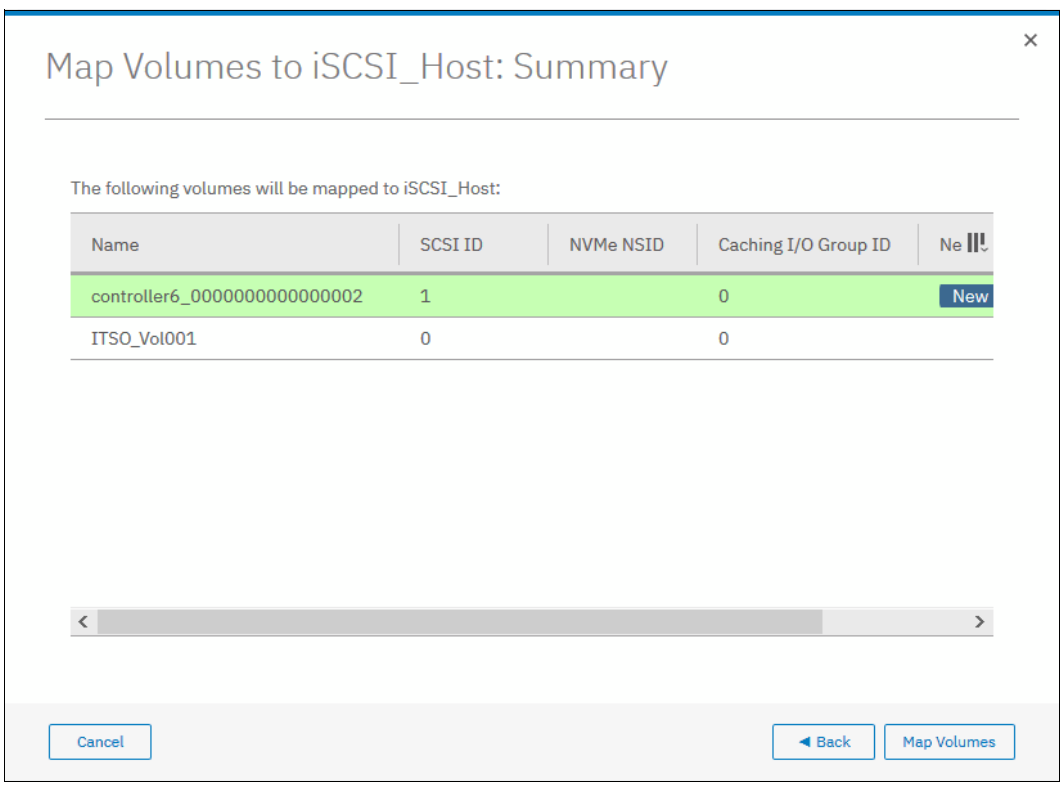


Figure 9-12 Volumes mapping summary before migration

Click **Map Volumes** and wait for the mappings to be created.

14. Select the storage pool that you want to migrate the imported volumes into. Ensure that the selected storage pool has enough space to accommodate the migrated volumes before you continue. This is an optional step. You can decide not to migrate to Storage pool and to leave the imported MDisk as an image-mode volume. This technique is not recommended because no volume mirroring are created. Therefore, no protection is available for the imported MDisk, and no data transfer occurs from the system to be migrated and the Storwize V7000. Click **Next**, as shown in Figure 9-13.

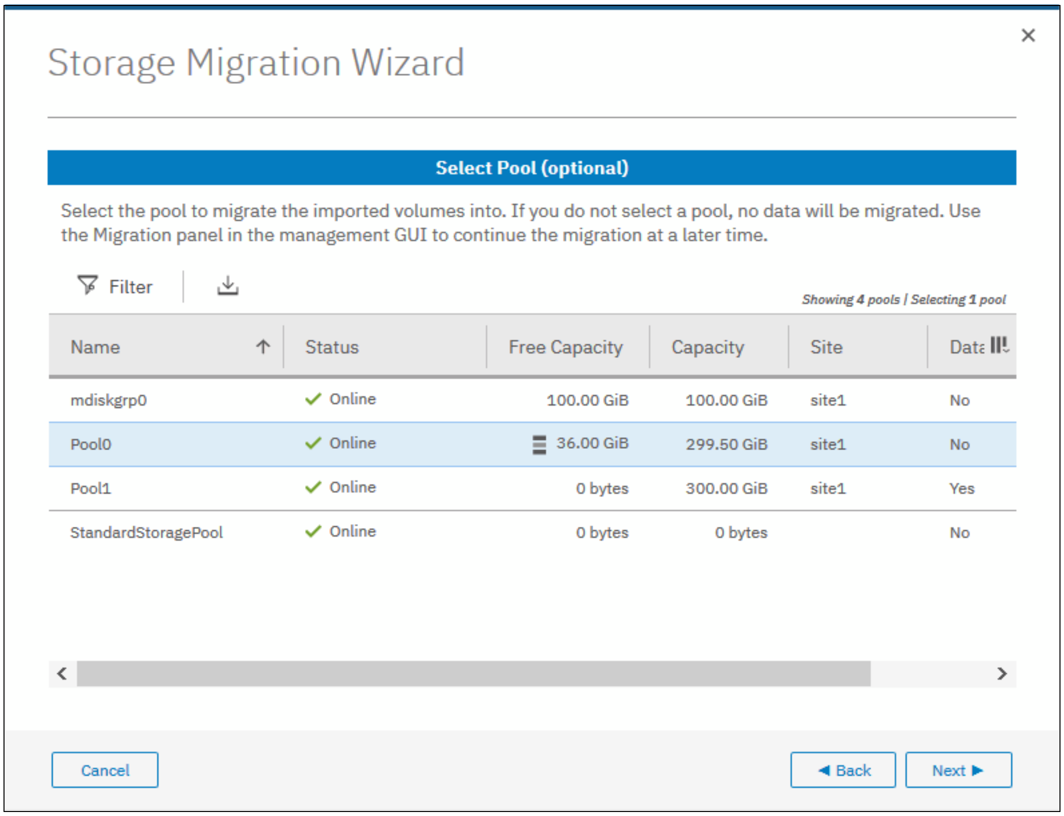


Figure 9-13 Select the pool to migrate the MDisk to

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 at the same time on both copies. Read I/Os are performed on the preferred copy.
>
> 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 9-14.

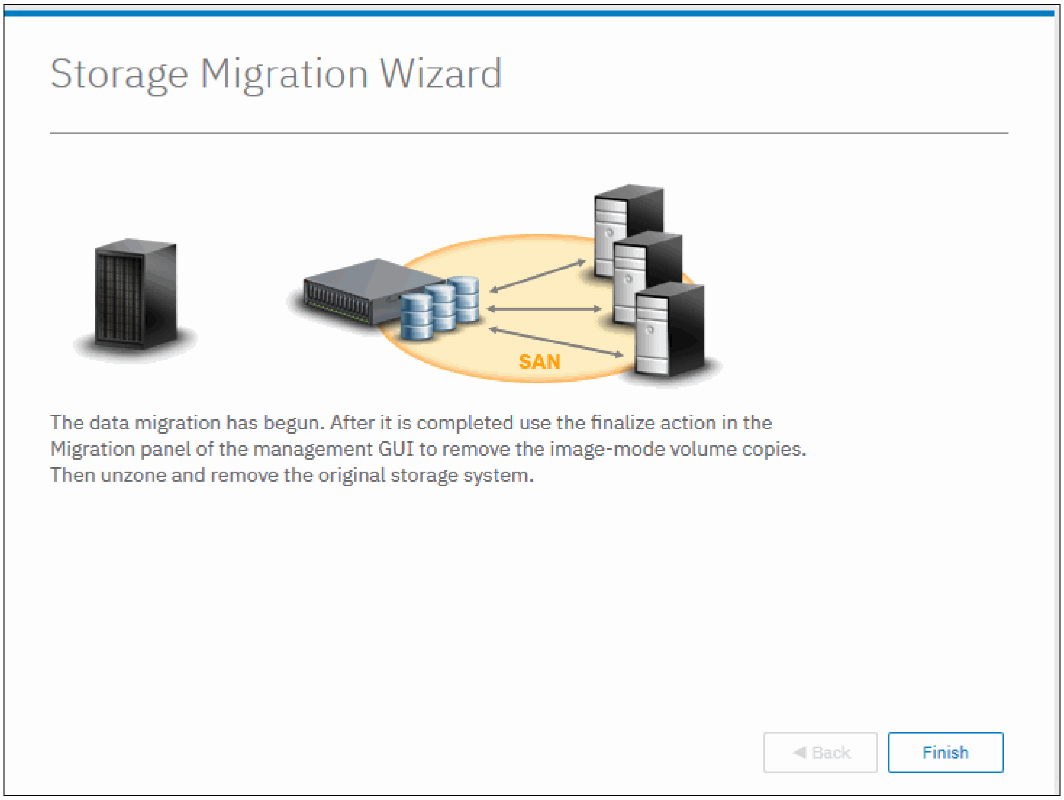


Figure 9-14 Migration is started

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 9-15. The target storage pool and the progress of the volume copy synchronization is also displayed there.

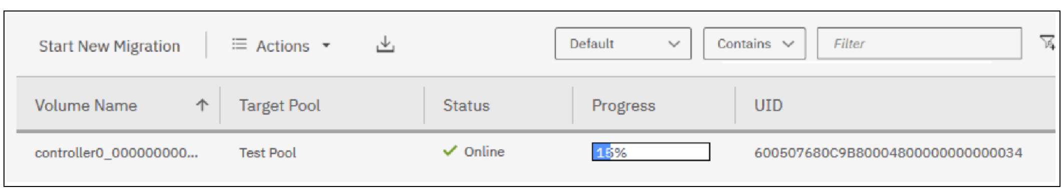


Figure 9-15 The ongoing Migration is listed in the Storage Migration window

16.If you want to check the progress by using the CLI, run the **`lsvdisksyncprogress`** command because the process is essentially a volume copy, as shown in Figure 9-16.

```
IBM_Storwize:ITSO-V7k:superuser>lsvdisksyncprogress
vdisk_id vdisk_name                      copy_id progress estimated_completion_time
26       controller0_0000000000000000 0 1_       7        181016175637
IBM_Storwize:ITSO-V7k:superuser>
```

Figure 9-16 Checking migration progress by using CLI

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 9-17 on page 400.

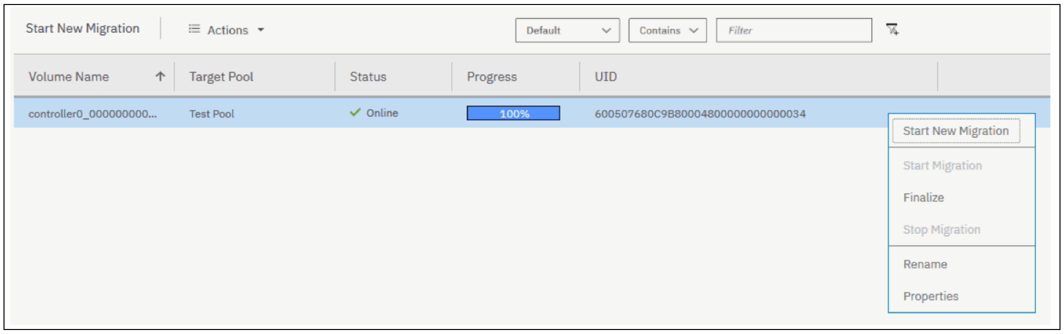


Figure 9-17 Finalizing a migration

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 9-18 shows the confirmation message.

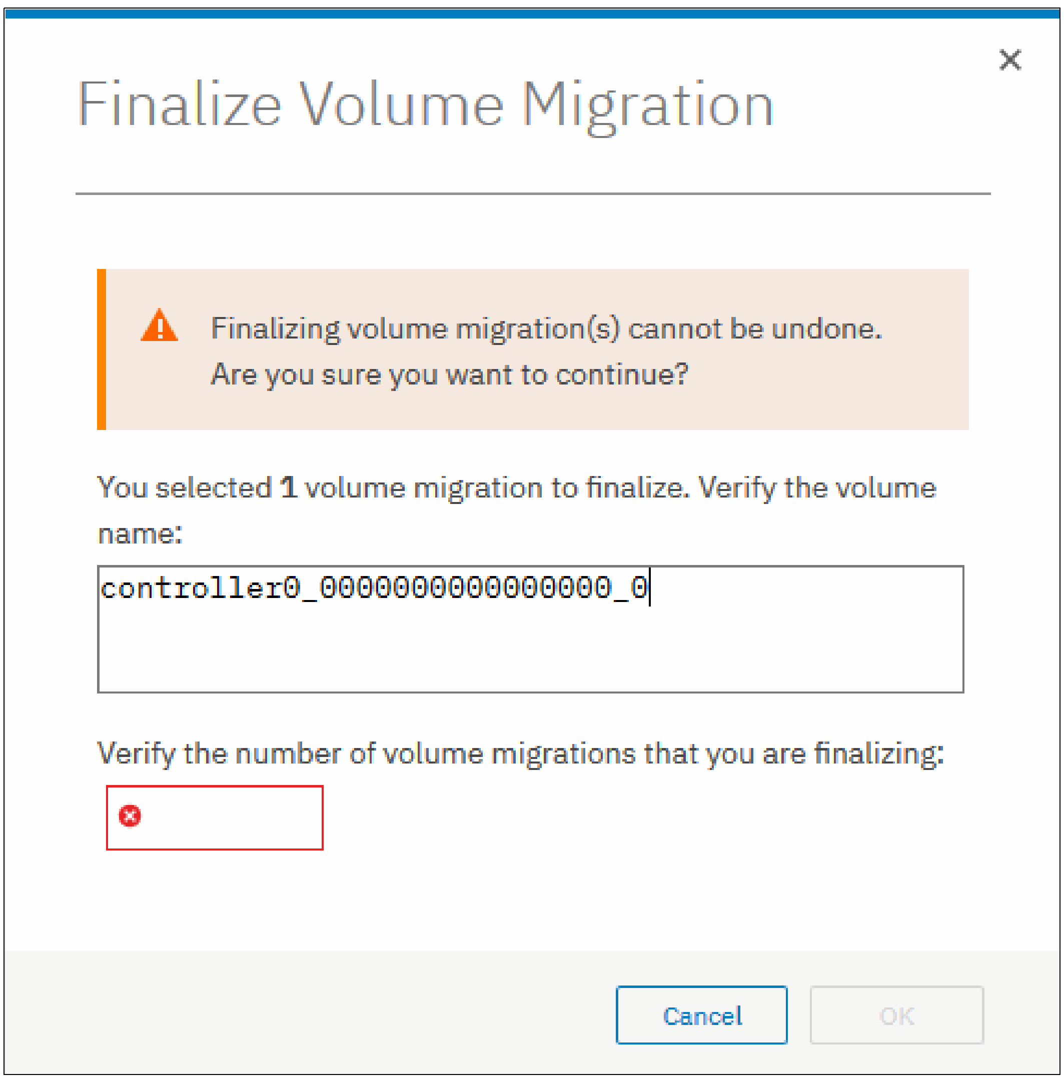


Figure 9-18 Migration finalization confirmation

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 navigating to **Pools** → **External Storage**, as shown in Figure 9-19. In the example, `mdisk3` was migrated and finalized. It appears as `unmanaged` in the external storage window.

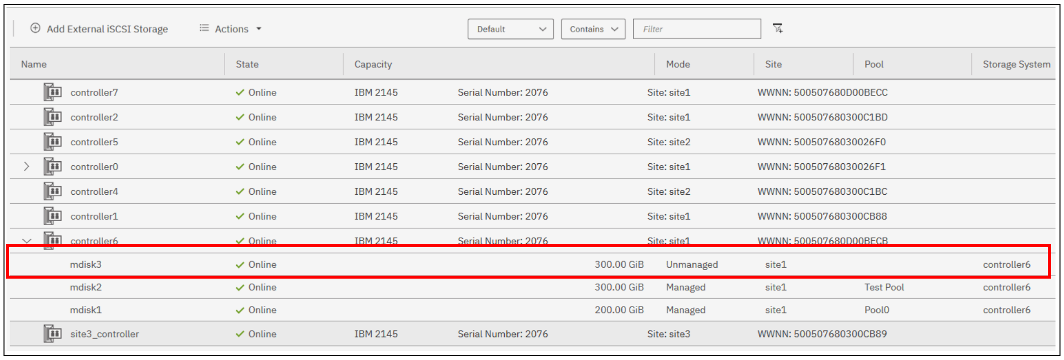


Figure 9-19 External Storage MDisks window

All the steps that are described in the Storage Migration wizard can be performed manually, but generally use the wizard as a guide.

Note: For a “real-life” demonstration of the storage migration capabilities offered with IBM Spectrum Virtualize, see this web page (log in required).

The demonstration includes three different step-by-step scenarios showing the integration of an IBM SAN Volume Controller cluster into an environment with one Microsoft Windows Server (image mode), one IBM AIX server (LVM mirroring), and one VMware ESXi server (storage vMotion).

10

Advanced features for storage efficiency

IBM Spectrum Virtualize running inside the IBM Storwize V7000 offers several functions for storage optimization and efficiency.

This chapter introduces the basic concepts of those functions. It also provides a short technical overview and implementation recommendations.

For more information about planning and configuration of storage efficiency features, see the following publications:

- *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521
- *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430
- *IBM Real-time Compression in IBM SAN Volume Controller and IBM Storwize V7000*, REDP-4859
- *Implementing IBM Real-time Compression in SAN Volume Controller and IBM Storwize V7000*, TIPS1083
- *Implementing IBM Easy Tier with IBM Real-time Compression*, TIPS1072

This chapter includes the following topics:

10.1 Easy Tier

IBM Spectrum Virtualize includes the IBM System Storage Easy Tier function. It enables automated subvolume data placement throughout different storage tiers, and automatically moves extents within the same storage tier to intelligently align the system with current workload requirements. It also optimizes the use of Flash drives or flash arrays.

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 to 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 significantly reduced. By servicing most of the application workload from the fastest storage, Easy Tier acts to accelerate application performance, and increase overall server utilization. This can reduce costs in servers and application licenses.

Note: Easy Tier is a licensed function, but it is included in base code. No actions are required to activate the Easy Tier license on IBM Storwize V7000.

10.1.1 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 three tiers, and each tier provides optimal performance for the extent, as shown in Figure 10-1.

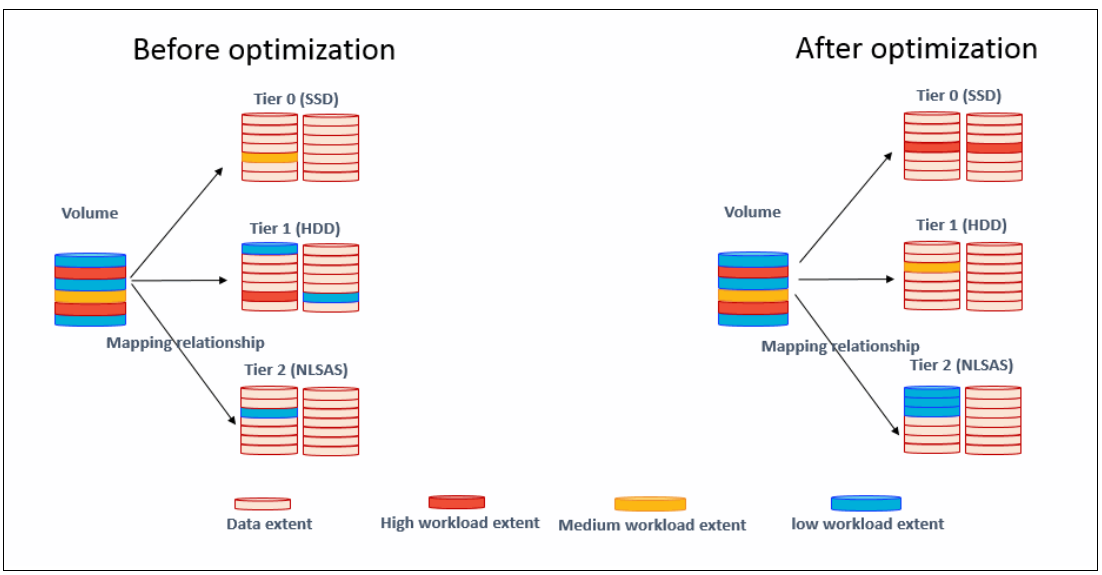


Figure 10-1 Easy Tier

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 one type of MDisk, Easy Tier enters balancing mode. With it, it moves extents from busy MDisks to less busy MDisks of the same tier.

Tiers of storage

The MDisks (external LUs or array type) that are presented to the Storwize V7000 are likely to have different performance attributes because of the type of disk or RAID array on which they are located.

Depending on performance, the system divides available storage into the following tiers:

- 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 (RI) flash drive technology. Tier 1 flash drives are lower-cost flash drives that typically offer capacities larger than enterprise class flash, but lower performance and write endurance characteristics.

- Enterprise tier

  Enterprise tier exists when the pool contains MDisks on enterprise-class hard disk drives, which are disk drives that are optimized for performance.

- Nearline tier

  Nearline tier exists when the pool has MDisks on nearline-class disks drives that are optimized for capacity.

For array type MDisks, the system automatically sets its tier because it knows the capabilities of array members, physical drives, or modules. External MDisks needs manual tier assignment when they are added to a storage pool.

> **Note:** The tier of MDisks mapped from certain types of IBM System Storage Enterprise Flash is fixed to tier0_flash, and cannot be changed.

Although IBM Storwize V7000 can distinguish four tiers, Easy Tier manages only a three tier storage architecture. MDisk tiers are mapped to Easy Tier tiers, depending on the pool configuration.

Figure 10-2 shows the possible combinations for the pool configuration with four MDisk tiers.

	EasyTier Tier (by configuration)													
	T0	T0+T1	T0+T1+T2	T0+T1+T2+T3	T0+T2	T0+T2+T3	T0+T3	T1	T1+T2	T1+T2+T3	T1+T3	T2	T2+T3	T3
T0 (Tier0 Flash)	1	1	1	1	1	1	1							
T1 (Tier1 Flash)		2	2	2				2	2	**1**	2			
T2 (Tier2 HDD)			**3**	2	2	2			**3**	2		2	2	
T3 (Tier3 NearLine)				3		3	**2**			3	3		3	3

Figure 10-2 Tier combinations

The table columns that are shown in Figure 10-2 on page 405 represent all of the possible pool configurations; the rows report in which Easy Tier tier each MDisk tier is mapped. For example, consider a pool with all the possible tiers configured that corresponds with the T0+T1+T2+T3 configuration in the table in Figure 10-2 on page 405. With this configuration, the T1 and T2 are mapped to the same Easy Tier tier. If there is no Tier 0 flash in a storage pool, Tier 1 flash is used as the highest performance tier.

For more information about planning and configuration considerations or best practices, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

Easy Tier actions

The Easy Tier function continuously monitors volumes for host I/O activity. It collects performance statistics for each extent, and derives averages for a rolling 24-hour period of I/O activity. Random and sequential I/O rate and bandwidth for reads and writes are collected, as well as I/O response time.

Different types of analytics are used to decide whether extent data migration is required. Once per day, Easy Tier analyze the statistics to work out which data should be sent to a higher performing tier or might be sent to a tier with lower performance. Four times per day, it analyzes the statistics to identify if any data needs to be rebalanced between managed disks in the same tier. Once every 5 minutes, Easy Tier checks the statistics to identify if any of the managed disks is overloaded.

All of these analysis phases generate a list of migrations that should be executed. The system then spends as long as needed running the migration plan.

The migration plan can consist of the following actions on volume extents (as shown in Figure 10-3):

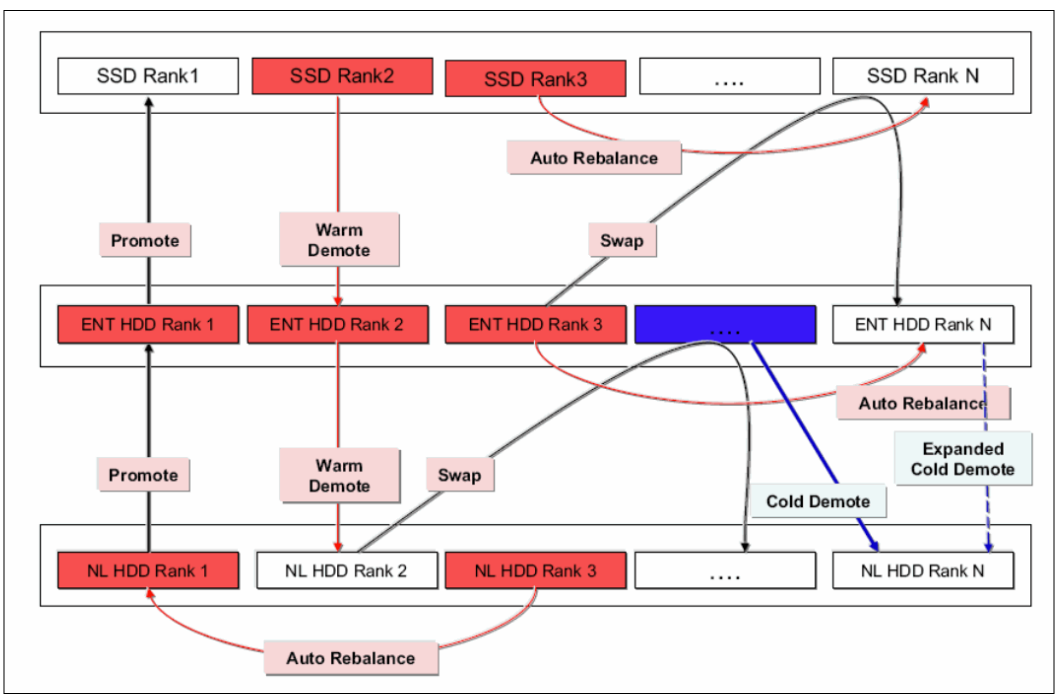


Figure 10-3 Actions on extents

- Promote

  Moves the hotter extents to the MDisks (ranks) of the higher-performance tier with available capacity. Promote occurs within adjacent tiers.

- Demote

  Demotes colder extents from the higher tier to the lower tier. Demote occurs within adjacent tiers.

- RB-Move (Auto Rebalance)

  Auto-rebalance, which is also known as intra-tier rebalancing, is a capability of Easy Tier that automatically rebalances the workload across all MDisks of a storage tier within a managed extent pool. It also automatically populates new ranks that were added to the pool.

- Warm demote

  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 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.

- Cold demote

  Easy Tier automatically locates and demotes inactive (or cold) extents that are on a higher performance tier to its adjacent lower-cost tier. In that way, it automatically frees extents on the higher storage tier before the extents on the lower tier become hot, and then helps the system to be more responsive to new hot data.

  Cold demote occurs between tiers 2 and 3 only.

- Expanded cold demote

  Demotes appropriate sequential workloads to the lowest tier to better use nearline tier bandwidth.

- Swap

  A swap moves a “hot” extent from a lower performance disk tier to a higher disk tier while simultaneously moving a “cold” extent from the higher disk tier to a lower performance disk tier.

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 guaranteed 1 GB every 5 minutes, and then gets whatever is left

> **Note:** Extent promotion or demotion only occurs between adjacent tiers. In a three-tier storage pool, Easy Tier does not move extents from a flash tier directly to nearline tier or vice versa without moving to the enterprise tier first.

The Easy Tier overload protection is designed to avoid overloading any type of MDisk with too much work. To achieve this, Easy Tier needs to have an indication of the maximum capability of a managed disk.

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. For a SAN-attached managed disk, the system cannot calculate the performance capabilities, so the system has a number of predefined levels that can be configured manually for each MDisk. This is called the Easy Tier load parameter (low, medium, high, very_high).

If you analyze the statistics and find that the system does not appear to be sending enough IOPS to your external SSD MDisk, you can always 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 only collects 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/Storage pool balancing

  In this mode, usage statistics are collected and extents. Extent migration is performed between tiers (if there is more than one pool in a tier). Also, auto-balance between MDisks of each tier is performed.

> **Note:** The auto-balance process automatically balances existing data when MDisks are added into a pool. However, it 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 Storwize V7000:

- 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” T0 or “slower” T2 tiers. If there are no free extents in T1, extents from the other tiers are used.
- 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 the other tier are used.
- Easy Tier monitors extent I/O activity of each copy of a mirrored volume. Easy Tier works with each copy independently of the other copy.

> **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.

- For compressed volumes on standard pools, only reads are analyzed by Easy Tier.
- 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`** CLI command on that volume.

10.1.2 Implementing and tuning Easy Tier

The Easy Tier function is enabled by default. It starts monitoring I/O activity immediately after storage pool and volumes are created, and starts extent migration as soon as the necessary I/O statistics are collected.

A few parameters can be adjusted. Also, Easy Tier can be switched 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, navigate to **Pools** → **External Storage**, select the MDisk (or MDisks) to add and click **Assign**.

> **Note:** The tier of MDisks 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 10-4.

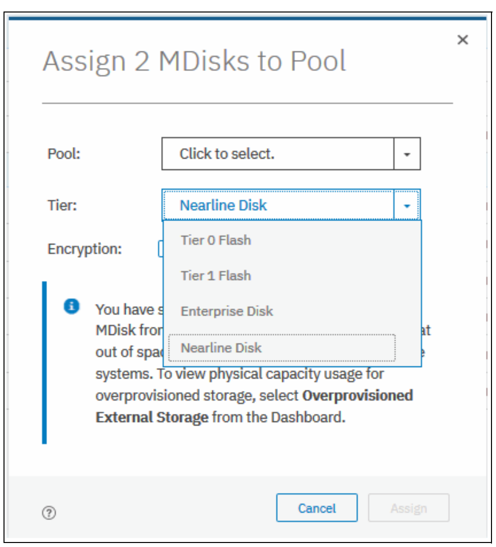


Figure 10-4 Choosing tier when assigning MDisks

To change the storage tier for an MDisk that is assigned, while in **Pools** → **External Storage**, right-click one or more selected MDisks and choose **Modify Tier**, as shown in Figure 10-5.

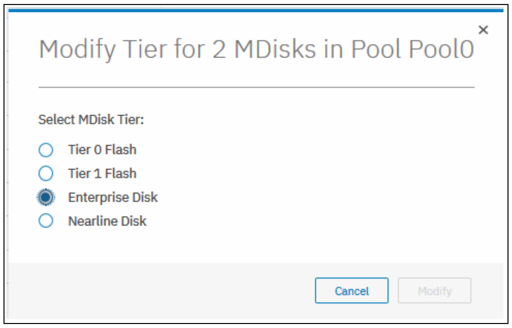


Figure 10-5 Changing MDisk tier

Note: Assigning a tier that does not match to a physical back-end storage type to an external MDisk is not supported by IBM and can lead to unpredictable consequences.

To determine what tier is assigned to an MDisk, on **Pools** → **External Storage**, select **Actions** → **Customize columns** and select **Tier**. This includes the current tier setting into a list of MDisk parameters that are shown on the **External Storage** pane. 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 10-6.

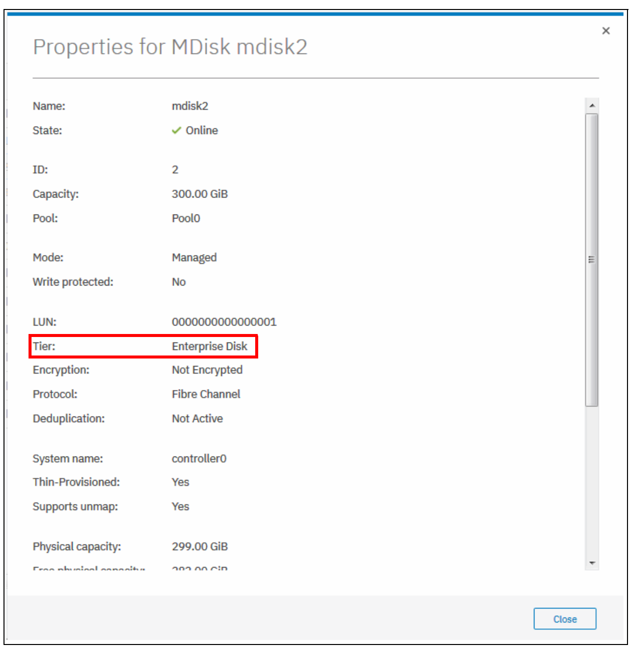


Figure 10-6 MDisk properties

To list MDisk parameters with the CLI, use the **`lsmdisk`** command. The current tier for each MDisk is shown. To change the external MDisk tier, use the **`chmdisk`** command with the **`-tier`** parameter, as shown in Example 10-1 on page 411.

Example 10-1 Listing and changing tiers for MDisks (partially shown)

```
IBM_Storwize:ITSOV7K: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_Storwize:ITSOV7K:superuser>chmdisk -tier tier1_flash mdisk2
IBM_Storwize:ITSOV7K: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 under-using an MDisk. If so, levels can be adjusted only by using the CLI. Use **chmdisk** with **-easytierload** parameter. To reset Easy Tier load to system-default for chosen MDisk, use **-easytier default**, as shown in Example 10-2.

Example 10-2 Changing Easy Tier load

```
IBM_Storwize:ITSOV7K:superuser>chmdisk -easytierload default mdisk2
IBM_Storwize:ITSOV7K:superuser>
IBM_Storwize:ITSOV7K:superuser>lsmdisk mdisk2 | grep tier
tier tier_enterprise
easy_tier_load high
IBM_Storwize:ITSOV7K:superuser>
```

Note: Adjust 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, use **lsmdisk** with MDisk name or ID as a parameter.

Storage pool settings

When storage pool (either standard pool or Data Reduction Pool) is created, Easy Tier is switched 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 moment later. This is not possible with the GUI, only with the system CLI.

To check current Easy Tier function state on a pool, navigate to **Pools** → **Pools**, right-click the selected pool, choose **Properties** and expand **View more details** section, as shown in Figure 10-7 on page 412.

Easy Tier can be in one of the following statuses:

- *active*

  Indicates that a pool is being managed by Easy Tier, and extent migrations between tiers can be performed. Performance-based pool balancing is also enabled.

  This is the expected state 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.

  This is the expected state for a pool with a single tier of storage.

- *inactive*

  Indicates that Easy Tier is inactive (switched off).

- *measured*

  Shows that Easy Tier statistics are being collected but no extent movement can be performed.

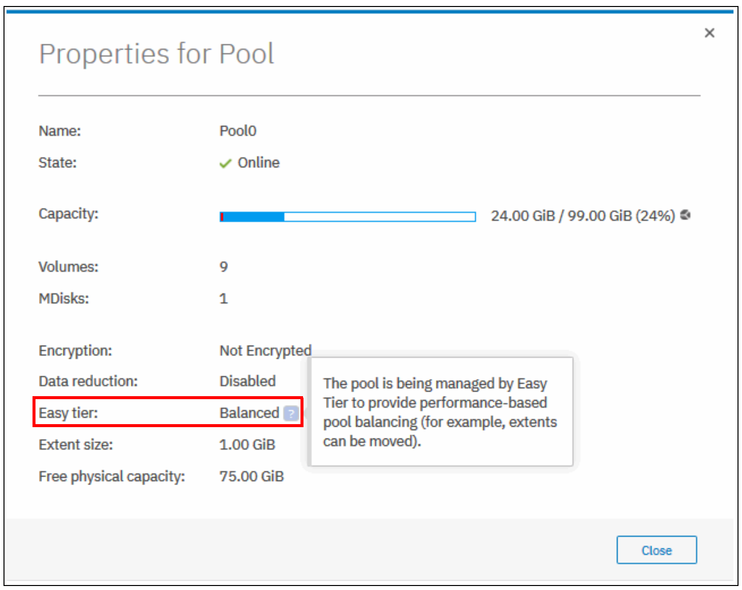


Figure 10-7 Pool properties

To find the status of the Easy Tier function on the pools with the CLI, use the **lsmdiskgrp** command without any parameters. To switch Easy Tier off or back on, use the **chmdiskgrp** command, as shown in Example 10-3. 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 10-3 Listing and changing Easy Tier status on pools

```
IBM_Storwize:ITSOV7K:superuser>lsmdiskgrp
id name  status mdisk_count ... easy_tier easy_tier_status
0  Pool0 online 1           ... auto      balanced
2  Pool1 online 3           ... auto      balanced
IBM_Storwize:ITSOV7K:superuser>chmdiskgrp -easytier measure Pool0
IBM_Storwize:ITSOV7K:superuser>chmdiskgrp -easytier auto Pool0
IBM_Storwize:ITSOV7K:superuser>
```

Volume settings

By default, each striped-type volume allows 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 data reduction pool cannot have Easy Tier switched off. It is possible to switch off Easy Tier only at a pool level.

This can be done only by using the CLI. Use the **lsvdisk** command to check and the **chvdisk** command to modify Easy Tier function status on a volume copy, as shown in Example 10-4.

Example 10-4 Checking and modifying Easy Tier settings on a volume

```
IBM_Storwize:ITSO-V7k:superuser>lsvdisk  vdisk0 |grep easy_tier
easy_tier on
easy_tier_status balanced
IBM_Storwize:ITSO-V7k:superuser>chvdisk -easytier off vdisk0
IBM_Storwize:ITSOV7K:superuser>
```

System-wide settings

There is a system-wide setting *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 accelerate mode, Easy Tier can move up to 48 GiB per 5 minutes, whereas 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, 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 recommended to keep Easy Tier acceleration mode off during normal system operation.

This setting can be changed online, but only by using the CLI. To turn on or off Easy Tier acceleration mode, use the **chsystem** command. Use the **lssystem** command to check its current state, as shown in Example 10-5.

Example 10-5 The chsystem command

```
IBM_Storwize:ITSOV7K:superuser>lssystem |grep easy_tier
easy_tier_acceleration off
IBM_Storwize:ITSOV7K:superuser>chsystem -easytieracceleration on
IBM_Storwize:ITSOV7K:superuser>
```

10.1.3 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.

The IBM Storage Tier Advisor Tool (STAT) is a Windows console application that can analyze heat data files that are produced by Easy Tier and produce a graphical display of the amount of “hot” data per volume and predictions of how more Solid-State Drive (T0) capacity, Enterprise Drive (T1), and Nearline Drive (T2) might benefit performance for the system and by storage pool.

IBM STAT can be downloaded from this IBM Support web page.

You can download the Storage Tier Advisor Tool 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 Storage Tier Advisor Tool is installed in `C:\Program Files\IBM\STAT\`.

On IBM Storwize V7000, 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 Storage Tier Advisor Tool started from a Windows command prompt console with the file specified as a parameter, as shown in Example 10-6.

Example 10-6 Running STAT in Windows command prompt

```
C:\Program Files (x86)\IBM\STAT>stat dpa_heat.7822DFF-1.181028.073824.data
```

The Storage Tier Advisor Tool creates a set of `.html` and `.csv` files that can be used for Easy Tier analysis.

To download a heat data file, open **Settings** → **Support** → **Support Package** → **Download Support Package** → **Download Existing Package**, as shown in Figure 10-8.

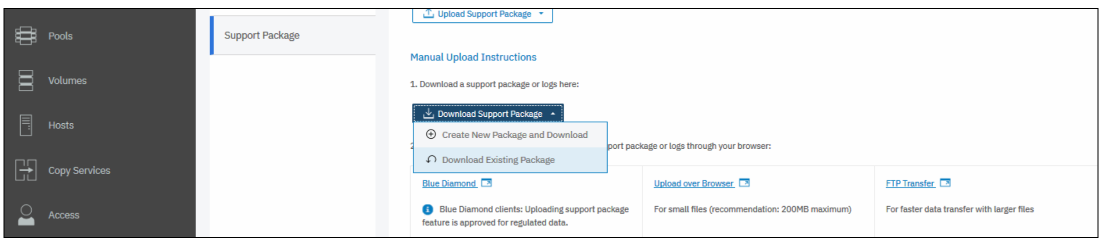


Figure 10-8 Download Easy Tier heat file: Download Support Package

A download window opens that shows all of 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 10-9. Save them in a convenient location (for example, to a subfolder that holds the STAT executable).

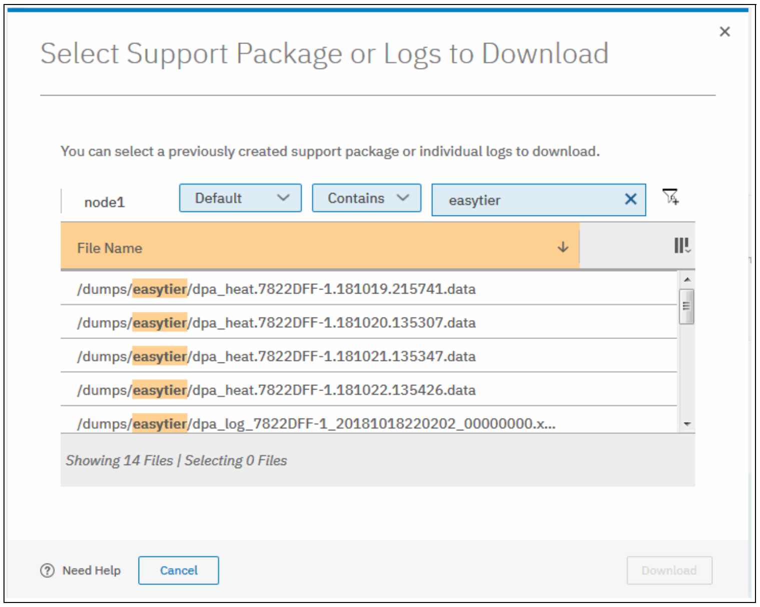


Figure 10-9 Downloading Easy Tier heat data file: dpa_heat files

You can also specify the output directory. 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.

For more information about how to interpret STAT tool output and CSV files analysis, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

10.2 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 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 (often as low as 10%) 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).

Thin provisioning presents more storage space to the hosts or servers that are connected to the storage system than is available on the storage system.

10.2.1 Concepts

Volumes can be configured as thin-provisioned or fully allocated. Both of them can be configured in standard pools and data reduction pools.

In IBM Storwize V7000, each volume has *virtual capacity* and *real capacity* parameters. *Virtual capacity* is the volume storage capacity that is available to a host Operating System (OS) and is used by it to create a file system. *Real capacity* is the storage capacity that is allocated to a volume from a pool, it shows the actual amount of space that is used on physical storage.

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

A thin-provisioned volume presents a different capacity to mapped hosts than the capacity that the volume uses in the storage pool, so *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.

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, 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 data reduction pools, the autoexpand feature is always enabled.

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.

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 Storwize V7000 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.

10.2.2 Implementation

For more information about creating thin-provisioned volumes, see Chapter 7, “Volumes” on page 241.

Metadata

In a standard pool, the system uses the real capacity to store data that is written to the volume, and metadata that describes the thin-provisioned configuration of the volume. Metadata uses volume real capacity and usually needs less than 0.5% of virtual capacity to store its data.

This means that if your host used 100% of virtual capacity, some extra space is required on your storage pool to store thin provisioning metadata. In a worst case, the real size of a thin provisioned volume can be 100.5% of its virtual capacity.

In a data reduction pool, metadata for a thin provisioned volume is stored separately from user data, and does not count in the volumes real capacity.

Volume parameters

When creating a thin provisioned volume in Custom mode, some of its parameters might be modified, as shown in Figure 10-10.

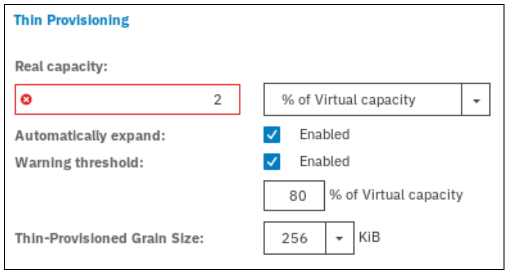


Figure 10-10 Thin VDisk parameters

In a DRP pool, thin provisioned volume fine-tuning is not required. Real capacity (rsize) value is ignored, and Grain Size is fixed to 8 KB.

When a thin-provisioned volume is created in a standard pool, Real capacity (rsize) defines both initial volume real capacity and an amount of contingency capacity, which is used by autoexpand.

Write I/Os to the grains of the thin volume in a standard pool (that were not previously written to) cause grains of the real capacity to be used to store metadata and user data. Write I/Os to the grains (that were previously written to) update the grain where data was written. The grain is defined when the volume is created, and can be 32 KiB, 64 KiB, 128 KiB, or 256 KiB.

Smaller granularities can save more space, but they have larger metadata directories. When you use thin-provisioning with FlashCopy, specify the same grain size for the thin-provisioned volume and FlashCopy.

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 virtual size.

10.3 Unmap

Spectrum Virtualize systems running V8.1.0 and later support the SCSI Unmap command. This enables hosts to notify the storage controller of capacity that is no longer required, which can improve capacity savings.

10.3.1 SCSI unmap command

Unmap is a set of SCSI primitives that allow hosts to indicate to a SCSI target that space allocated to a range of blocks on a target storage volume is no longer required. This command allows 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 within a file system. The storage controller can then 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, human intervention is usually needed on the storage controller. The SCSI Unmap feature is used to allow host operating systems to unprovision storage on the storage controller, which means that the resources can automatically be freed up in the storage pools and used for other purposes.

A SCSI unmappable volume is a volume that can have storage unprovision and space reclamation being triggered by the host operating system. IBM Storwize V7000 can pass the SCSI unmap command through to back-end storage controllers that support the function.

10.3.2 Back-end SCSI Unmap

The system can generate and send SCSI Unmap commands to specific back-end storage controllers.

This occurs when volumes are deleted, extents are migrated, or an Unmap command is received from the host. SCSI Unmap commands at the time of this writing are only sent to IBM A9000, IBM FlashSystem FS900 AE3, IBM FlashSystem FS9100, IBM Storwize, and Pure storage systems.

This helps prevent a thin-provisioning storage controller from running out of free capacity for write I/O requests. This means when you are using supported thin provisioned back-end storage, SCSI Unmap should normally be left enabled.

This feature is turned on by default. It is recommended to keep back-end unmap enabled, especially if a system is virtualizing a thin-provisioned storage controller or uses Flash Core Modules.

To verify that sending unmap commands to back-end is enabled, use the CLI command **lssystem**, as shown in Example 10-7.

Example 10-7 Verifying back-end unmap support status

```
IBM_Storwize:ITSOV7K:superuser>lssystem | grep backend_unmap
backend_unmap on
```

10.3.3 Host SCSI Unmap

The Spectrum Virtualize system might advertise support for SCSI Unmap to hosts. With it, some host types (for example, Windows, Linux, or VMware) then changes their behavior when creating a file system on a volume, issuing SCSI Unmap commands to the whole capacity of the volume. This causes the system to overwrite the whole capacity with zero-data, and the format completes when all of these writes 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.

Host Unmap commands can increase the free capacity that is reported by the Data Reduction Pool when received by thin-provisioned or compressed volumes. This does not apply to standard storage pools.

Also, IBM Storwize V7000 sends back-end SCSI Unmap commands to controllers that support them if host unmaps for corresponding blocks are received.

Host SCSI Unmap commands drive more I/O workload to back-end storage. In some circumstances (for example, volumes on a heavily loaded nearline SAS array), this 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 advertise support the SCSI Unmap command.

If you use Data Reduction Pools, thin-provisioned back-end that supports Unmap, or Flash Core Modules, it is recommended to turn on SCSI Unmap support.

If only standard pools are configured and back-end is traditional (fully provisioned), you might consider keeping host Unmap support switched off because you do not see any improvement.

To check and modify current setting for host SCSI Unmap support, use the **lssystem** and **chsystem** CLI commands, as shown in Example 10-8.

Example 10-8 Turning host unmap support on

```
IBM_Storwize:ITSOV7K:superuser>lssystem | grep host_unmap
host_unmap off
IBM_Storwize:ITSOV7K:superuser>chsystem -hostunmap on
IBM_Storwize:ITSOV7K: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.

10.3.4 Offload IO 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.

SCSI offload commands, such as Unmap and Xcopy, are used by hosts to format new file systems, or copy volumes without the host needing to read and then write data.

Some host types might request large amounts of I/O on the storage system by issuing Write Same/Unmap commands. If the underlying storage cannot handle the amount of I/O that is generated, the performance of volumes can be affected.

Spectrum Virtualize offload throttling limits the concurrent I/O that can be generated by such commands, which can prevent the MDisk overloading. This limits the rate at which host features, such as VMware VMotion, can copy data.

> **Note:** For systems that fee managing any nearline storage, it might be recommended to set the offload throttle to 100 MBps.

To implement offload throttle, you can use the **`mkthrottle`** command with **`-type offload`** parameter, or GUI, as shown in Figure 10-11.

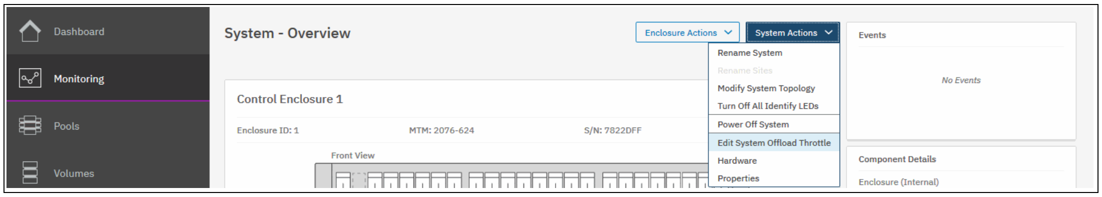


Figure 10-11 Setting offload throttle

10.4 Data Reduction Pools

Data Reduction Pools (DRP) increase infrastructure capacity use by using new efficiency functions and reducing storage costs. The pools enable you to automatically de-allocate and reclaim capacity of thin-provisioned volumes that contain deleted data and for the first time, enable this reclaimed capacity to be reused by other volumes. Data reduction pools support volume compression. Also, they support deduplication that can be configured with thin-provisioned and compressed volumes.

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

10.4.1 Introduction to DRPs

At its core, a DRP uses a Log Structured Array (LSA) to allocate capacity. A log structured array allows 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, an LSA allows you to allocate data sequentially and provide a directory that provides a lookup to match LBA with physical address within the array. Therefore, the volume 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 data reduction pools, 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 requests 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.

DRP technology allows you to create five types of volumes:

- Fully allocated

  This type provides no storage efficiency.

- Thin-provisioned

  This type provides some storage efficiency but no compression or deduplication. Volume capacity is allocated on demand as storage is first written.

- Thin and Compressed

  In addition to on-demand space allocation, data is compressed before being written to storage.

- Thin and Deduplicated

  In addition to on-demand space allocation, duplicates of data blocks are detected and are replaced with references to the first copy.

- Thin, Compressed, and Deduplicated

  These types provides maximum storage efficiency and capacity savings by combining thin, compressed, and deduplicated.

The following software and hardware requirements are needed for DRP compression and deduplication:

- Enabled Compression license
- V8.1.3.2 or higher
- Nodes must have at least 32 GB memory to support deduplication

Random Access Compression Engine (RACE) compression and DRP compressed volumes can coexist in the same I/O group. However, deduplication is not supported in the same I/O group as RACE compressed volumes.

10.4.2 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 are cheaper to cool. However, the cost of Flash storage is still higher than disk storage.

With technologies, such as DRP, the cost difference can be reduced to a point where an all Flash solution is feasible. The first benefit of DRP is in the form of storage savings because of deduplication. The deduplication process identifies unique data patterns, and stores the signature of the data for reference when writing new data. If the signature of the new data matches a signature, the new data is not written to disk, but instead a reference to the stored data is written. The same byte pattern can occur many times, which results in the amount of data that must be stored being greatly reduced.

The second benefit of DRP comes in the form of performance improvements because of compression. Although deduplication aims to identify the same data elsewhere in the storage pool and create references to the duplicate data instead of writing extra copies, compression is trying to reduce the size of the host data that is written.

Compression and deduplication are not mutually exclusive, one, both, or neither features can be enabled. If the volume is de-duplicated and compressed, data is de-duplicated first and then compressed. Therefore, deduplication references are created on the compressed data stored on the physical domain.

DRPs offer a new implementation of data compression that is integrated into the I/O stack. The new implementation makes better use of the resources of the system, and, similar to RACE, uses hardware accelerators for compression.

As with RACE, the new implementation uses the same Lempel-Ziv (LZ) based real-time compression and decompression algorithm. However, in contrast to RACE compression, DRP compression operates on smaller block sizes, which results in more performance gains.

The third benefit of DRP comes in the form of performance improvements because of Easy Tier.

The metadata of data reduction pools does not fit in RAM; therefore, it is stored on disk on metadata volumes, which are separate from data volumes. The metadata volumes of data reduction pools are small and frequently accessed. They are good candidates for promotion through Easy Tier. In contrast, data volumes are large, but because the metadata is stored separately, Easy Tier can accurately identify frequently used data. Performance gains are expected because Easy Tier promotes metadata to the fastest available storage tier.

DRPs support end-to-end SCSI Unmap functionality. Space that is freed from the hosts is a process called *unmap*. A host can issue a small file unmap (or a large chunk of unmap space if you are deleting a volume that is part of a data store on a host) and these result in the freeing of all the capacity allocated within that unmap. Similarly, deleting a volume at the DRP level frees all of the capacity back to the pool.

When a data reduction pool is created, the system monitors the pool for reclaimable capacity from host unmap operations. This capacity can be reclaimed by the system and redistributed into the pool. Create volumes that use thin provisioning or compression within the data reduction pool to maximize space within the pool.

10.4.3 Implementing DRP with compression and deduplication

The implementation process for DRPs is similar to standard pools, but has its own specifics.

Creating pools and volumes

To create a DRP, select **Data Reduction Enable** in the Create Pool dialog, which is available be clicking **Pools** → **Pools**. For more information about how to create a storage pool and populate it with MDisks, see Chapter 6, “Storage pools” on page 191.

There is a maximum number of four DRPs in a system. When this limit is reached, no other DRPs can be created.

A DRP uses metadata. Even when there are no volumes in the pool, some of the space of the pool is used to store the metadata. Regardless of the type of volumes the pool contains, metadata is always stored separate from customer data.

Metadata capacity depends on the total capacity of a storage pool and on a pool extent size. This should be considered when planning capacity.

> **Note:** If your DRP has a total capacity of less than 50 TB savings, you might need to decrease the extent size from the default for DRP (which is 4 GB) to 1 GB for optimal space.

The main purpose of DRPs is to be a fast and efficient container for volumes with data reduction capability. DRPs can also contain standard thick provisioned volumes.

To create a volume on a DRP, navigate to **Volumes** → **Volumes**, and click **Create Volumes**.

Figure 10-12 shows the Create Volumes window. 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.

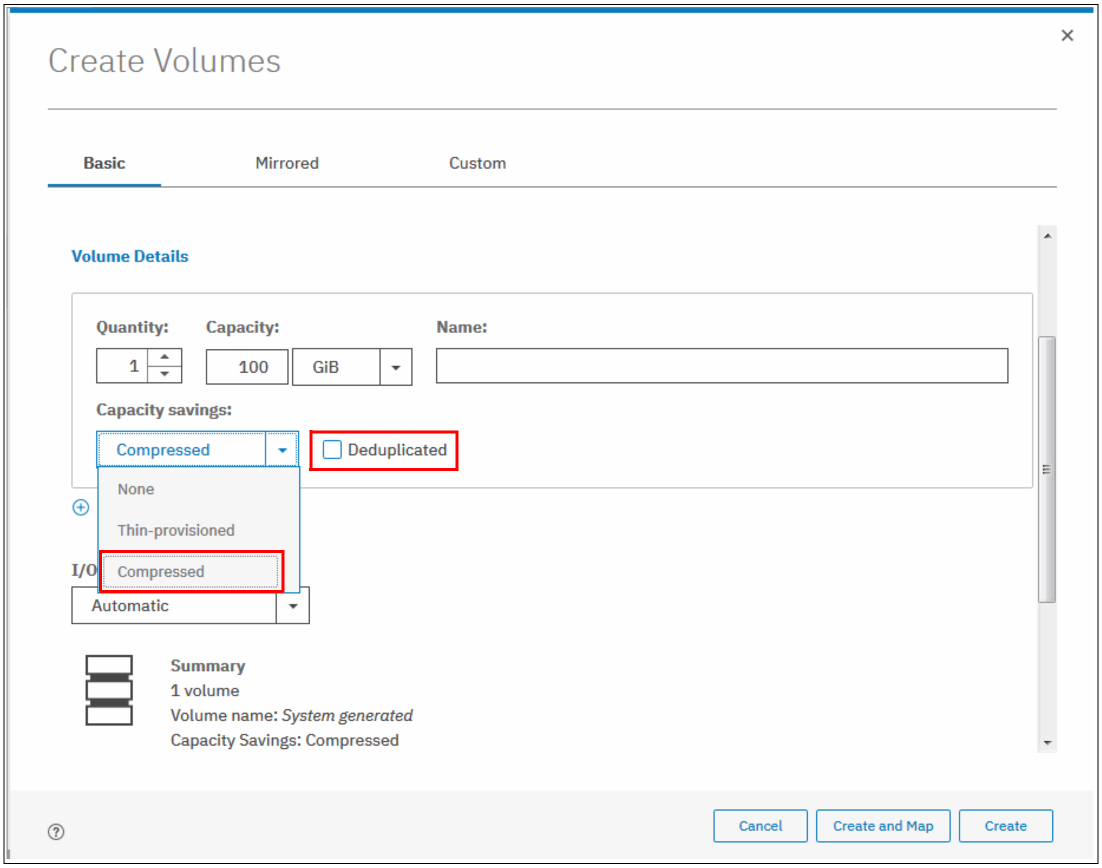


Figure 10-12 Create compressed volume

Capacity monitoring

Figure 10-13 shows the Volumes window with a list of volumes, which are created in DRP. Only Capacity (which is virtual capacity that is available to a host) is shown. Real capacity, Used capacity, and Compression savings show as Not Applicable for volumes with capacity savings. Only fully allocated volumes display those parameters.

Note: When using and reviewing Volumes in DRPs, be aware that no volume-copy level reporting is on used capacity.

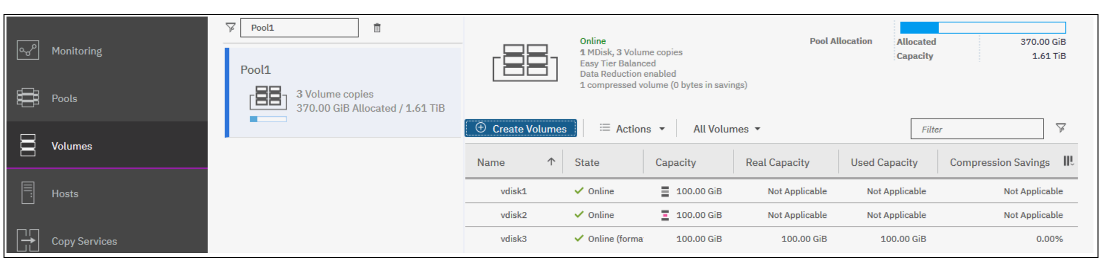


Figure 10-13 Volumes in DRP pool

The only parameter except virtual capacity that is available on a volume level to monitor thin, compressed, or deduplicated volume, is used_capacity_before_reduction. It indicates the total amount of data 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 data reduction pool. Capacity that is assigned to the volume in the tier is also not reported.

To find this value, use the **lsvdisk** command with volume name or ID as a parameter, as shown in Example 10-9. It shows a non-compressed thin-provisioned volume with virtual size 100 GiB on a DRP pool, which was mounted to a host, and had a 72 GiB file created on it.

Example 10-9 Volume in a DRP pool capacity monitoring

```
IBM_Storwize:ITSOV7K:superuser>lsvdisk vdisk1
id 15
name vdisk1
<...>
capacity 100.00GB
<...>
used_capacity
real_capacity
free_capacity

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 71.09GB
```

Capacity and space saving reporting is available in the storage pool views, and Out of space warning thresholds are configured at the storage pool level. You can check savings with the GUI Pools view, GUI Dashboard, or by using the CLI command **lsmdiskgrp** with the DRP pool ID or name as a parameter, as shown in Example 10-10.

Example 10-10 Pool savings monitoring

```
IBM_Storwize:ITSOV7K:superuser>lsmdiskgrp 3
id 3
name Pool1
<...>
capacity 8.64TB
free_capacity 7.53TB
virtual_capacity 1.79TB
used_capacity 882.38GB
real_capacity 883.95GB
overallocation 20
<...>
tier tier0_flash
tier_mdisk_count 0
```

```
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier tier1_flash
tier_mdisk_count 1
tier_capacity 8.64TB
tier_free_capacity 7.78TB
tier tier_enterprise
tier_mdisk_count 0
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier tier_nearline
tier_mdisk_count 0
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
<...>
data_reduction yes
used_capacity_before_reduction 1.30TB
used_capacity_after_reduction 793.38GB
overhead_capacity 89.00GB
deduplication_capacity_saving 10.80GB
reclaimable_capacity 0.00MB
physical_capacity 8.64TB
physical_free_capacity 7.78TB
shared_resources no
```

The output reports real capacity that is used on each of the storage tiers. Deduplication savings are also shown. That compression-related values show 0 MB because they belong to RtC (standard pool) compression.

For more information about every reported value, see IBM Knowledge Center.

Migrating to and from DRP

Although data can be migrated regardless of the nature of the new volumes, the type of these volumes determines which of the following migration strategies is:

- Migration strategy for any source volume on a standard pool to thin or compressed volume on DRP.

  If you need to migrate a fully allocated, thin-provisioned, or compressed volume from a standard pool to a fully allocated, thin-provisioned, or compressed volume in a DRP pool, create a second volume copy.

  To create a second copy, right-click the source volume and choose **Add Volume Copy**, as shown in Figure 10-14 on page 426. Choose target DRP for the second copy and **Capacity savings** type, None, Thin-provisioned, or Compressed.

  You can select **Deduplicated** only if there are no RtC compressed volumes in your I/O group.

  After you click **Add**, synchronization starts. The time synchronization takes to complete depends on the size of the volume and system performance.

You can increase the synchronization rate by right-clicking the volume and selecting **Modify Mirror Sync Rate**.

When copies are synchronized, Yes is displayed for both copies in the Synchronized column in the Volumes pane.

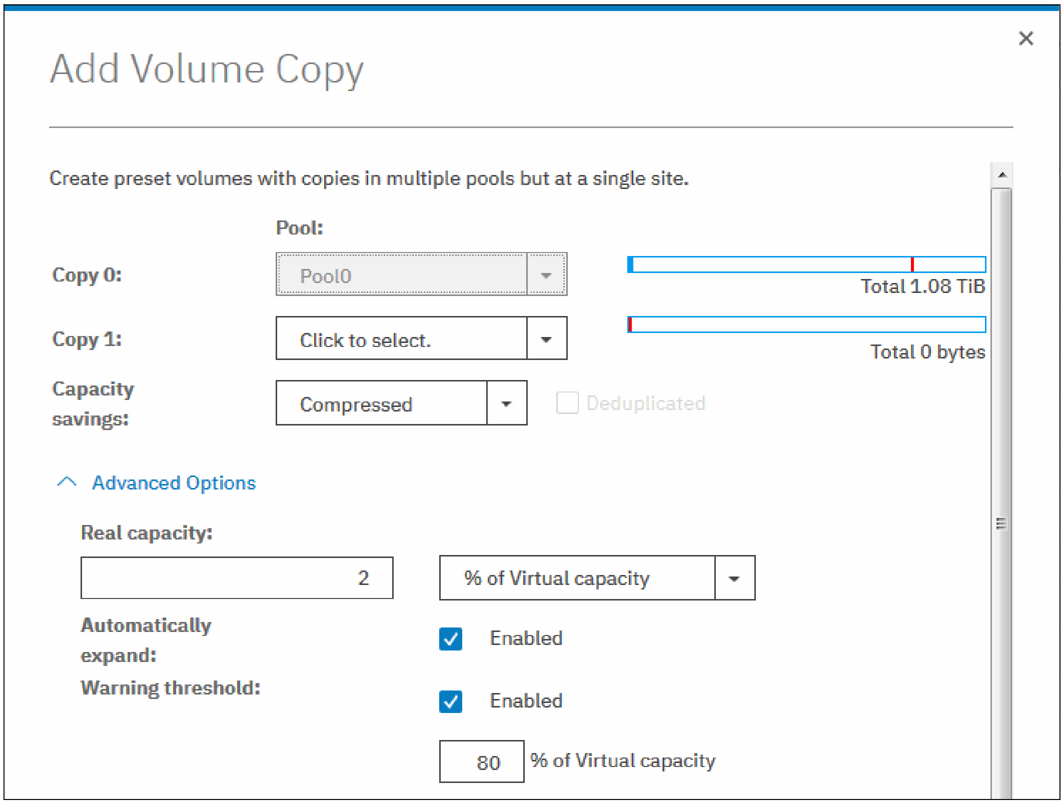


Figure 10-14 Add volume copy dialog

You can also track the synchronization process with Running tasks pane, as shown in Figure 10-15. After it reaches 100% and copies are in-sync, you can complete migration by deleting the source copy in a standard pool.

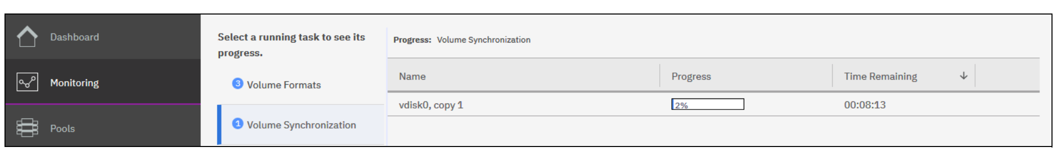


Figure 10-15 Synchronization progress

- Migration strategy for fully allocated, thin, or compressed VDisk on DRP to any type of volume on a standard pool.

  This process is the same as the previous method, but the target pool for a second copy is standard.

- Migration strategy for a compressed volume on a standard pool to thin and deduplicated or compressed and deduplicated volume on a DRP.

Deduplicated volumes cannot coexist in the same I/O group with compressed volumes in standard pools. To migrate a compressed volume from a standard pool to a thin-provisioned deduplicated or compressed deduplicated volume, a two-step procedure is required:

a. For a source compressed volume, create a second compressed, non-deduplicated copy in a DRP. Wait for synchronization to complete, and then, delete the source copy.

   Complete this process for all compressed volumes in all standard pools in an I/O group and verify that no more compressed volumes in standard pools are left.

b. For each compressed volume in a DRP pool, create a second copy on the same DRP pool, but select **Deduplicated**. Wait for synchronization, and delete the source copy to complete migration.

   Alternatively, right-click a compressed volume in a DRP, click **Modify Capacity Savings**, and select **Deduplicated**. A second copy is created and the source copy is deleted automatically after synchronization completes.

Garbage collection and volume deletion

DRP includes built-in services to enable garbage collection of unused blocks. 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 operating system, it is a process called *unmapping*. By it, hosts indicate that the allocated capacity is no longer required on a target volume. The freed space can be collected and reused on the system, without the reallocation of capacity on the storage.

Volume delete is a special case of Unmap that zeros an entire volume. There is a background Unmap process that sequentially walks the volume capacity, which emulates large Unmap requests. The purpose is to notify the garbage collection technology to free up the physical space.

Because this process can require some time, the volume deletion process in DRPS is asynchronous from the command. After you delete a volume by using the GUI or CLI, it goes into the deleting state, which can be seen by using the **`lsvdisk`** CLI command. After the Unmap process completes, the volume disappears.

Both host Unmaps and volume deletions increase unused space capacity. It is displayed by the system as `reclaimable_capacity`, which is shown by using the **`lsmdiskgrp`** command, as shown in Example 10-10 on page 424. Unused capacity is freed after it is reclaimed by garbage collection.

10.5 Compression with standard pools

Random Access Compression Engine (RACE) technology was first introduced in the IBM Real-time Compression Appliances. It is integrated into the IBM Storwize V7000 software stack as the IBM Real-time Compression (RtC) solution.

RACE or RtC is used for compression of the volume copies, which are allocated from standard pools. DRPs use a different IBM Storwize V7000 compression function.

For more information about RtC, see *IBM Real-time Compression in IBM SAN Volume Controller and IBM Storwize V7000*, REDP-4859.

10.5.1 Real-time Compression concepts

At a high level, the IBM RACE component compresses data that is written into the storage system dynamically. This compression occurs transparently, so Fibre Channel and iSCSI connected hosts are not aware of the compression. RACE is an inline compression technology, which means that each host write is compressed as it passes through IBM Spectrum Virtualize to the disks. This technology has a clear benefit over other compression technologies that are post-processing based.

These technologies do not provide immediate capacity savings. Therefore, they are not a good fit for primary storage workloads, such as databases and active data set applications.

RACE is based on the Lempel-Ziv lossless data compression algorithm and 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 storage pool. As part of its staging, the write request passes through the compression engine and is then stored in compressed format onto the storage pool. 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.

Capacity is saved when the data is written by the host because the host writes are smaller when they are written to the storage pool. IBM RtC is a self-tuning solution that adapts to the workload that runs on the system at any particular moment.

10.5.2 Implementing RtC compression

To create a compressed volume, choose **Capacity Savings** - **Compressed** in the Create Volumes window, as shown in Figure 10-16. For more information about creating volumes, see Chapter 7, “Volumes” on page 241.

Figure 10-16 Creating compressed VDisk

In addition to compressing data in real time, you can compress data sets (convert a volume to compressed). To do so, you must change the capacity savings settings of the volume by right-clicking it and selecting **Modify Capacity Settings**. In the menu, select **Compression** as the Capacity Savings option, as shown in Figure 10-17 on page 429.

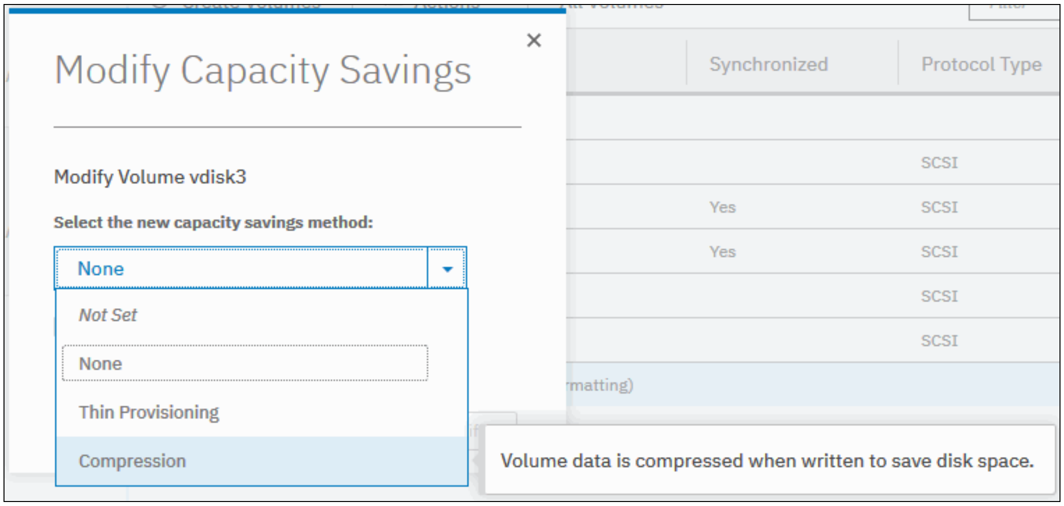


Figure 10-17 Selecting capacity setting

After the copies are fully synchronized, the original volume copy is deleted automatically.

As a result, compressed data is on the volume. This process is nondisruptive, so the data remains online and accessible by applications and users.

This capability enables clients to regain space from the storage pool, which can then be reused for other applications.

With the virtualization of external storage systems, the ability to compress stored data significantly enhances and accelerates the benefit to users. This capability enables them to see a tremendous return on their Storwize V7000 investment.

On the initial purchase of an Storwize V7000 with Real-time Compression, clients can defer their purchase of new storage. When storage is needed, IT purchases a lower amount of the required storage before compression.

For more information about volume migration for compressed volumes on standard pools to DRPs, see “Migrating to and from DRP” on page 425.

10.6 Saving estimation for compression and deduplication

This section provides information about the tools that are used for sizing the environment for compression and deduplication.

10.6.1 Evaluate compression savings by using IBM Comprestimator

IBM Comprestimator is an integrated GUI and CLI host-based utility that estimates the space savings that are achieved when compressed volumes are used for block devices. This utility provides a quick and easy view of showing the benefits of the use of compression. The utility performs read-only operations and therefore, does not affect the data that is being stored on device.

If the compression savings prove to be beneficial in your environment, volume mirroring can be used to convert volumes to compressed volumes in the data reduction pools.

To analyze all the volumes that are on the system, run the CLI `analyzevdiskbysystem` command.

The use of 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. Progress for analyzing of 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 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.

The command `lsvdiskanalysis` is used to display information for thin provisioning and compression estimation analysis report for all volumes.

You can also use the GUI to run analysis. Navigate to the **Volumes** pane, right-click any volume, and select **Space Savings** → **Estimate Compression Savings**, as shown in Figure 10-18.

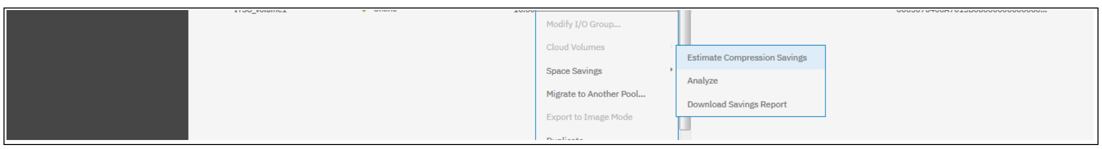


Figure 10-18 Compresstimator submenu

This action shows you the results of the latest estimation cycle and the date when it was completed. If no analysis was done yet, the system suggests running it.

To run or rerun on a single volume, select **Analyze** in **Space Savings** submenu.

If you are using a version of IBM Spectrum Virtualize that is older than V7.6 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. More information and the latest version of this utility, see this IBM Support web page.

Consider the following recommended best practices for using Comprestimator:

- Run the Comprestimator utility before implementing an IBM Spectrum Virtualize solution and DRPs.
- Download the latest version of the Comprestimator utility if you are not using one that is included in your IBM Spectrum Virtualize solution.
- Use Comprestimator to analyze volumes that contain as much active data as possible rather than volumes that are nearly empty or newly created. This use ensures more accuracy when sizing your environment for compression and DRPs.

Note: 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%.

10.6.2 Evaluating compression and deduplication

To help with the profiling and analysis of user workloads that must be migrated to the new system, IBM provides a highly accurate data reduction estimation tool that supports both deduplication and compression. The tool operates by scanning target workloads on any legacy array (from IBM or third party) and then merging all scan results to provide an integrated system level data reduction estimate.

The Data Reduction Estimator Tool (DRET) utility uses advanced mathematical and statistical algorithms to perform an analysis with low memory footprint. The utility runs on a host that can access the devices to be analyzed. It performs only read operations so it has no effect on the data 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 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, the 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. This 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 this IBM Support web page.

10.7 Data deduplication and compression on external storage

Starting from IBM Spectrum Virtualize V8.1.x, it supports over-provisioning on selected back-end controllers. This means that if back-end storage performs data deduplication or data compression on LUs provisioned from it, they still can be used as external MDisks on IBM Storwize V7000.

Thin-provisioned MDisks from controllers that are supported by this feature can be used as managed mode MDisks in IBM Storwize V7000 and added to storage pools (including DRPs).

Implementation steps for thin-provisioned MDisks are same as for fully allocated storage controllers. Extra caution is used when planning capacity for such configurations.

IBM Storwize V7000 detects if the MDisk is thin-provisioned, its total physical capacity, used, and remaining physical capacity. It detects if SCSI Unmap commands are supported by back-end.

By sending SCSI Unmap commands to thin-provisioned MDisks, IBM Storwize V7000 marks data that is no longer in use. Then, the garbage collection processes on the back-end can free unused capacity and reallocate it to free space.

At the time of this writing, the following back-end controllers are supported by thin-provisioned MDisks:

- IBM A9000 V12.1.0 and above
- FlashSystem 900 V1.4
- FlashSystem 9000 AE2 expansions
- IBM Storwize family system with code V8.1.0 and above
- Pure Storage

It is recommended to assign thin-provisioned MDisks to DRPs. To avoid overcommitting, assume a 1:1 compression ratio in back-end storage. Small, extra savings are realized from compressing metadata.

Fully allocated volumes that are above thin-provisioned MDisks configurations must be avoided or used with extreme caution because it can lead to overcommitting back-end storage.

It is not recommended to use “DRP above DRP”, when the back-end system belongs to the IBM Spectrum Virtualize family. With this configuration, implement fully allocated MDisks on the back-end and configure space efficiency features on IBM Storwize V7000 with DRPs.

The IBM Storwize V7000 storage administrator can monitor space use usage on thin-provisioned MDisks by using one of the following methods:

- GUI Dashboard, as shown in Figure 10-19.

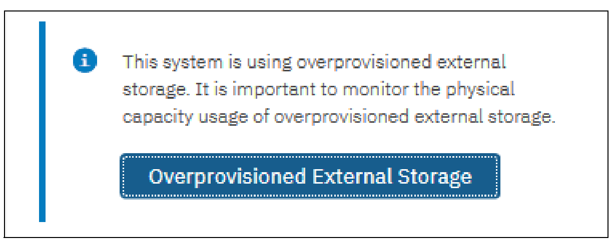


Figure 10-19 Dashboard button for overprovisioned storage monitoring

For more information, see Chapter 5, “Graphical user interface” on page 131.

- MDisk properties window, which opens by right-clicking an MDisk in **Pools** → **MDisks by Pools** and then selecting the **Properties** option, as shown in Figure 10-20.

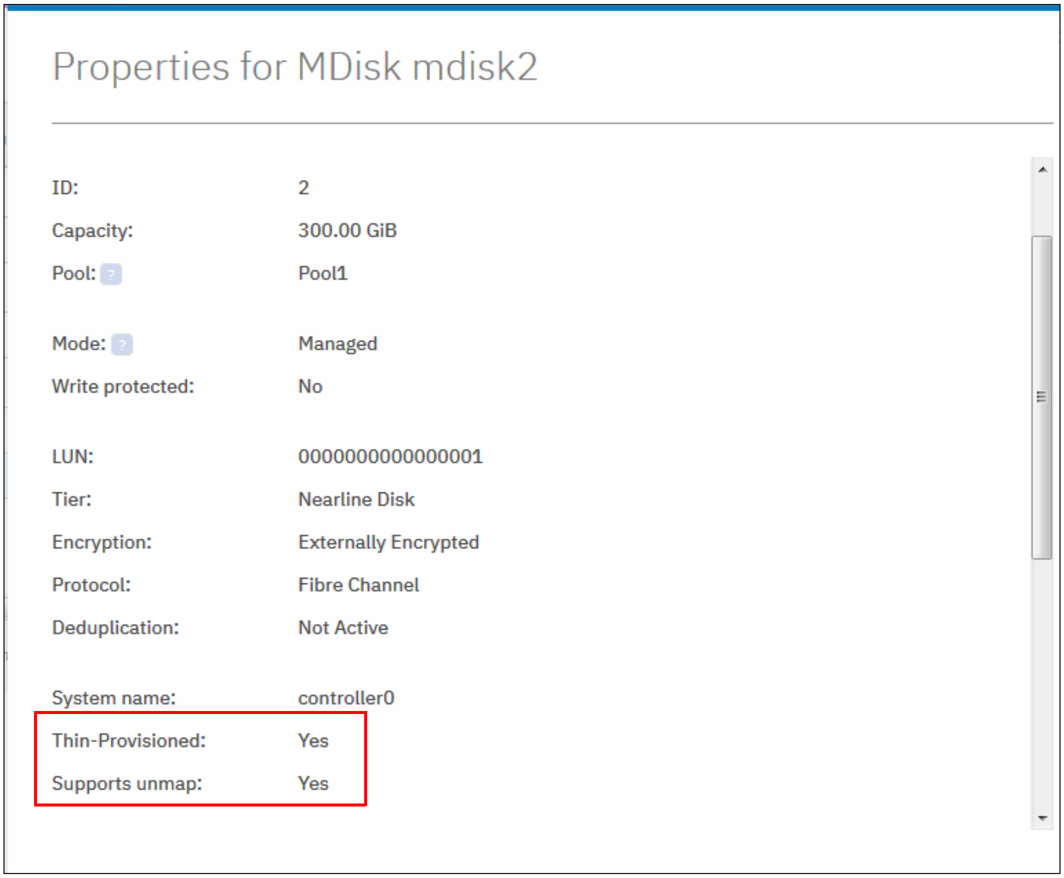


Figure 10-20 Thin-provisioned MDisk properties

- CLI command `lsmdisk` with MDisk name or ID as a parameter, as shown in Example 10-11.

Example 10-11 lsmdisk parameters for thin provisioned MDisks

```
IBM_Storwize:ITSOV7K: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 over-provisioning status and SCSI Unmap support for the selected MDisk are displayed.

IBM Storwize V7000 allocates `provisioning_group_id`, which is the identifier for the provisioning group that is affiliated with the MDisk. A provisioning group is an object that represents a set of managed disks that share physical resources. This is represented differently depending on the back-end storage, as shown in the following examples:

- A9000: The entire subsystem forms one provisioning group.
- Storwize V7000: The storage pool forms a provisioning group, which allows 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.

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 recommended to create multiple storage pools out of a MDisks in a single provisioning group.

11

Advanced Copy Services

This chapter describes the Advanced Copy Services, which are a group of functions that provide different methods of data copy. The storage software capabilities to support the interaction with hybrid clouds also are described. These functions are enabled by IBM Spectrum Virtualize software that runs inside IBM SAN Volume Controller and Storwize family products.

This chapter includes the following topics:

11.1 IBM FlashCopy

Through the IBM FlashCopy function of IBM Spectrum Virtualize, you can perform a *point-in-time copy* of one or more volumes. This section describes the inner workings of FlashCopy and provides details 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 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.

11.1.1 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:

- Rapidly creating consistent backups of dynamically changing data
- Rapidly creating consistent copies of production data to facilitate data movement or migration between hosts
- Rapidly creating copies of production data sets for application development and testing
- Rapidly creating copies of production data sets for auditing purposes and data mining
- Rapidly creating copies of production data sets 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 usually is managed as read-only *at the operating system 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:

- There is no need to concern about pairing mistakes. 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.

> **Preferred 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/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/archive as appropriate for your business.

In addition to the restore option, which 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 need to make the target available on a host. Is suggested to do not make the target available to the source host because seeing duplicates of disks causes problems for most host operating systems. 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 11.1.12, “Reverse FlashCopy” on page 457.

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 operating system 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.

11.1.2 FlashCopy principles and terminology

The FlashCopy function creates a point-in-time or time-zero (T0) copy of data that is stored on a *source volume* to a *target volume* by using a *copy on write* and *copy on-demand* 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 that their T0 is identical for all of them. A simple one-to-one FlashCopy mapping does not need to belong to a consistency group.

Figure 11-1 shows the basic terms that are used with FlashCopy. All elements are explained in this chapter.

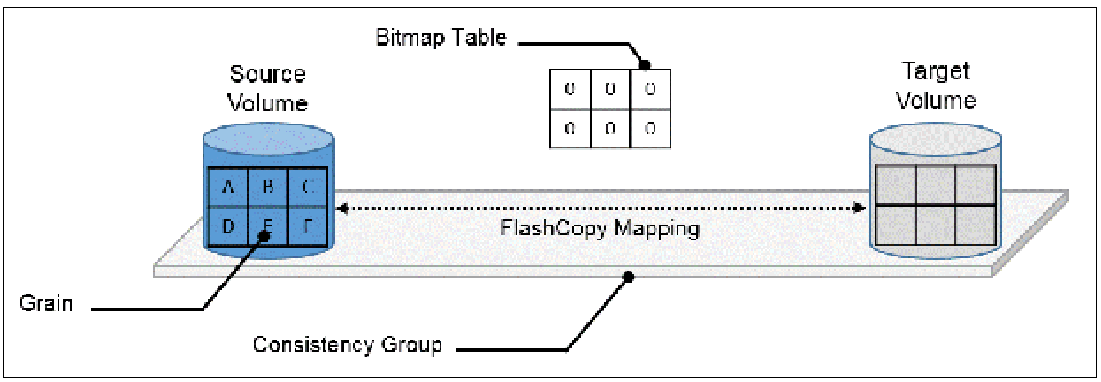


Figure 11-1 FlashCopy terminology

11.1.3 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 T0 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 point in time 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 point in time 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 these *actions* on a FlashCopy mapping:

- Create: Define a source and target, and set the properties of the mapping.
- Prepare: The system needs to be prepared before a FlashCopy copy starts. It basically 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.
- 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 T0 copy technology, you are bound by operating system and application requirements for interdependent data and the restriction to an entire volume.

The source and target volumes must belong to the same IBM SAN Volume Controller system, but they do not have to be in the same I/O Group or storage pool.

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, thanks to the use of Consistency Groups.

11.1.4 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 point-in-time 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 there is only one mapping in the Consistency Group. If a FlashCopy mapping is not part of any Consistency Group, it is referred 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 actual 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.

11.1.5 Crash consistent copy and hosts considerations

FlashCopy Consistency Groups do not provide application consistency. It only ensures 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 operating system. 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 operating system layer.

Both of these layers have various levels and methods of caching data to provide better speed. Because the Storwize V7000 and therefore, FlashCopy sit below these layers, *they are unaware of the cache at the application or operating system layers.*

To ensure the integrity of the copy that is made, it is necessary to flush the host operating system and application cache for any outstanding reads or writes before the FlashCopy operation is performed. Failing to flush the host operating system 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 operating systems 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 operating system (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.

Preferred 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 is possible because the database maintains strict control over I/O.

This method is as opposed to flushing data from both the application and the 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 Spectrum Protect Snapshot

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, it enables you 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 both IBM and non-IBM storage systems.

Note: For more information about how IBM Spectrum Protect Snapshot can help your business, see this IBM Knowledge Center web page.

11.1.6 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 referring to that grain is updated (for example, from "0" to "1"), as shown in Figure 11-2 on page 442.

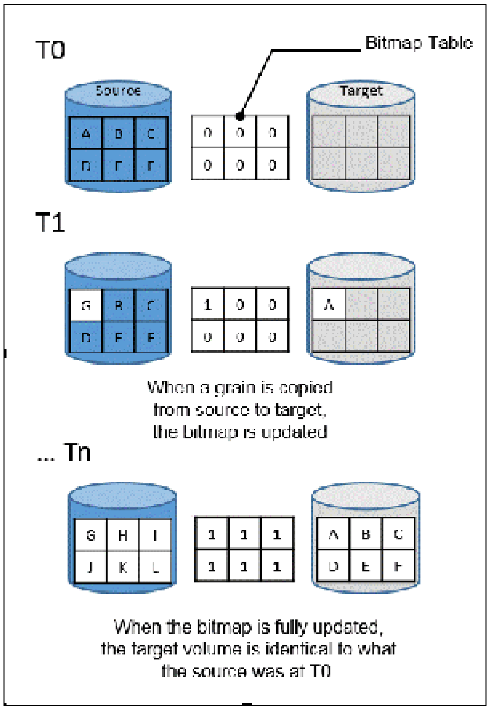


Figure 11-2   A simplified representation of grains and bitmap

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 creating a FlashCopy mapping 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, if the grain size parameter is not specified and one of the volumes is part of a FlashCopy mapping, the grain size of that mapping is used.

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. But other than in this situation, the default grain size is 256 KB.

Copy on Write and Copy on Demand

IBM Spectrum Virtualize FlashCopy uses Copy on Write (CoW) mechanism to copy data from a source volume to a target volume.

As shown in Figure 11-3 on page 443, 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.

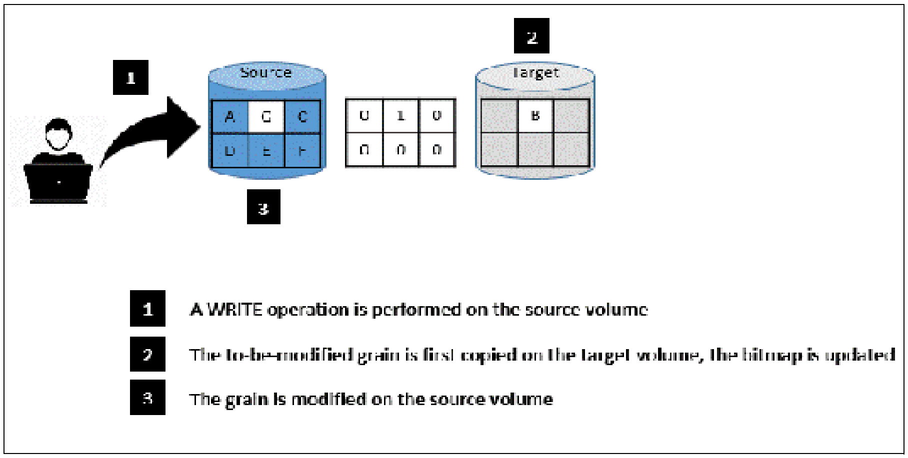


Figure 11-3 Copy on Write steps

With IBM FlashCopy, the target volume is immediately accessible for both Read *and* Write operations. Therefore, a target volume can be modified, even if it is part of a FlashCopy mapping. As shown in Figure 11-4, 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 (*Copy on Demand*). It is then modified with the new value. The bitmap is modified so the grain from the source is *not* copied again, even if it is changed or if a background copy is enabled.

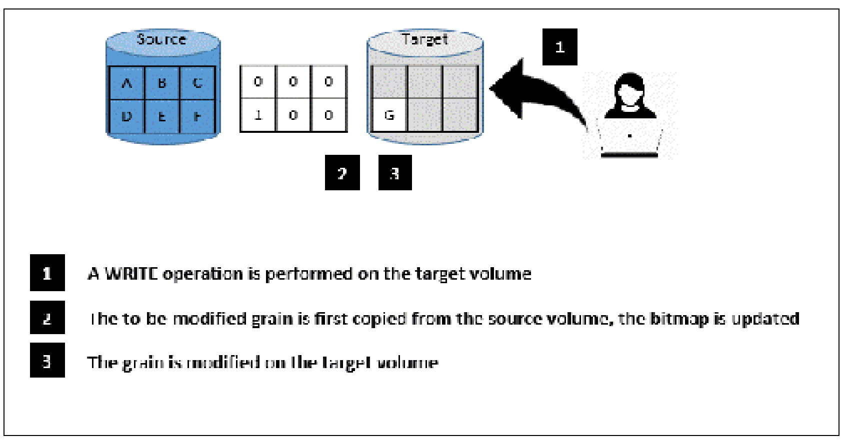


Figure 11-4 Copy on Demand steps

Note: If all the blocks of the grain to be modified are changed, there is no need to copy the source grain first. There is no copy on demand and it is directly modified.

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. Hereunder are the different I/O indirections.

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:

- 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 11-5.

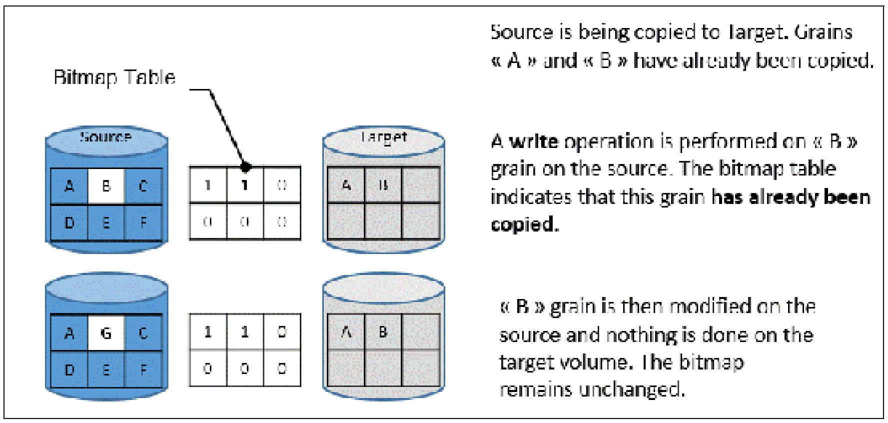


Figure 11-5 Modifying an already copied grain on the Source

- If the bitmap indicates that the grain was not yet copied, the grain is first copied on the target (copy on write), the bitmap table is updated, and the grain is modified on the source, as shown in Figure 11-6.

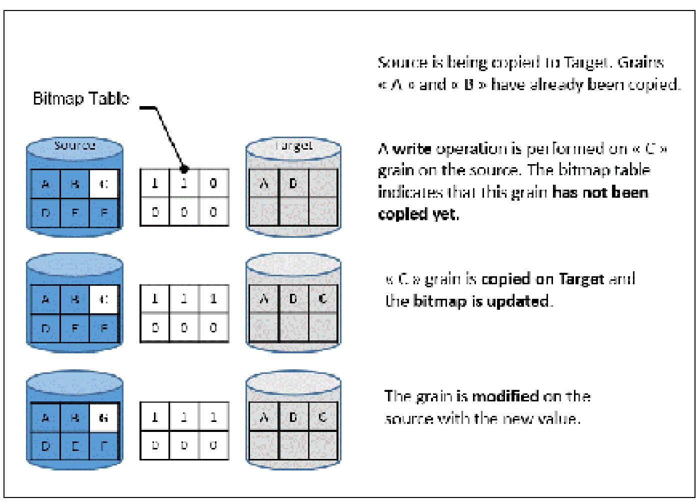


Figure 11-6 Modifying a non-copied grain on the source

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 is 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 11-7. The source volume remains unchanged.

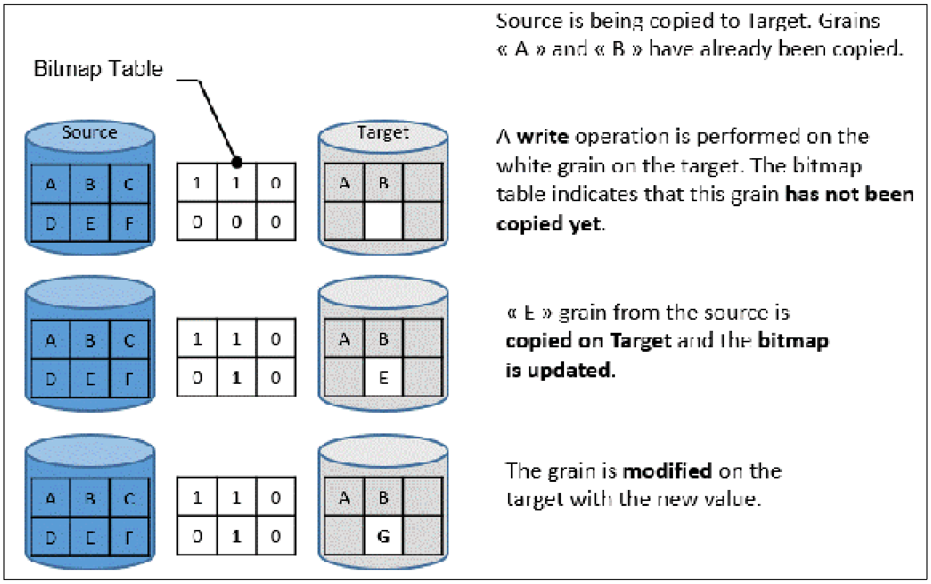


Figure 11-7 Modifying a non-copied grain on the target

> **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 11-8.

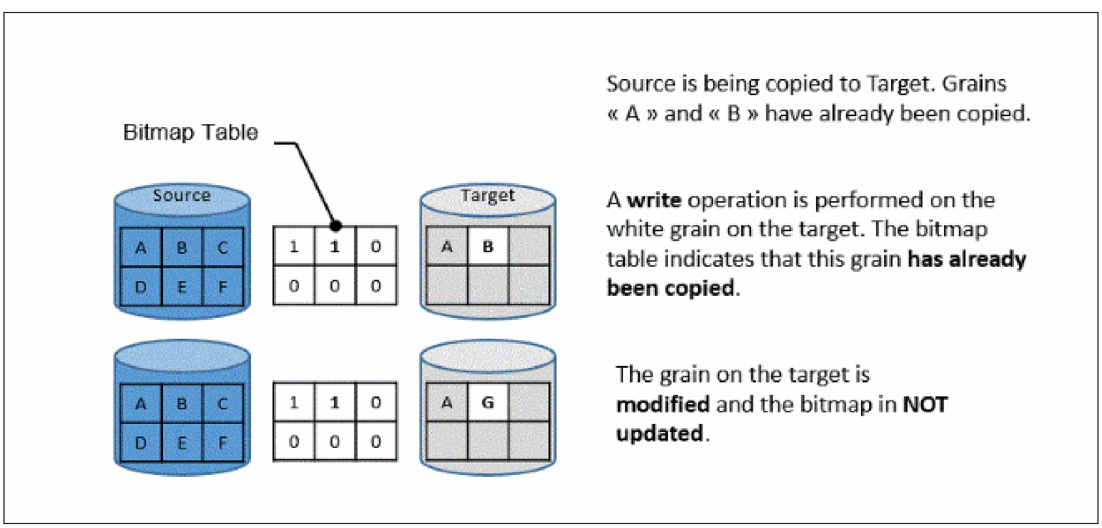


Figure 11-8 Modifying an already copied grain on the Target

Note: The bitmap is not updated in this case. Otherwise, it might be copied from the source later if a background copy is ongoing or if write operations are made on the source. That process overwrites 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:

- 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 11-9.
- 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 11-9.

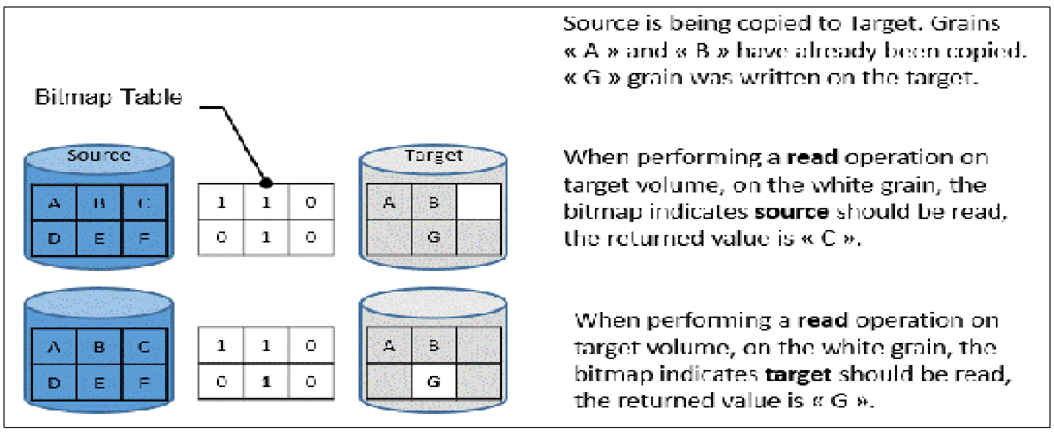


Figure 11-9 Reading a grain on target

If the source has multiple targets, the Indirection layer algorithm behaves differently on Target I/Os. For more information about multi-target operations, see 11.1.11, “Multiple target FlashCopy” on page 452.

11.1.7 Interaction with cache

IBM Spectrum Virtualize-based systems have their cache divided into upper and lower cache. Upper cache serves mostly as write cache and hides the write latency from the hosts and application. Lower cache is a read/write cache and optimizes I/O to and from disks. Figure 11-10 shows the IBM Spectrum Virtualize cache architecture.

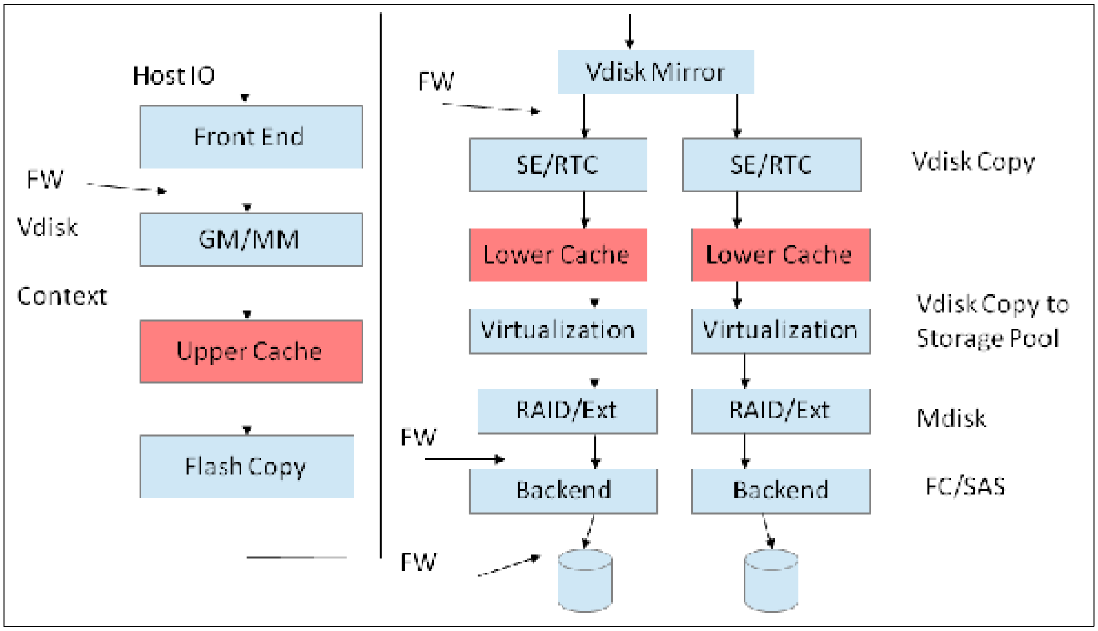


Figure 11-10 New cache architecture

This copy-on-write process introduces significant latency into write operations. To isolate the active application from this extra latency, the FlashCopy indirection layer is placed logically between upper and lower cache. Therefore, the extra latency that is introduced by the copy-on-write process is encountered only by the internal cache operations and not by the application.

The two-level cache provides more performance improvements to the FlashCopy mechanism. Because the FlashCopy layer is above lower cache in the IBM Spectrum Virtualize software stack, it can benefit from read prefetching and coalescing writes to back-end storage. Preparing FlashCopy benefits from the two-level cache because upper cache write data does not have to go directly to back-end storage, but to lower cache layer instead.

11.1.8 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 creating the FlashCopy mapping, 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 11-1.

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.

Table 11-1 Copy 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

The *grains per second* numbers represent the maximum number of grains that the Storwize V7000 copies per second. This amount assumes that the bandwidth to the managed disks (MDisks) can accommodate this rate.

If the Storwize V7000 cannot achieve these copy rates because of insufficient width 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 resides. This responsibility is moved to the other node in the I/O group if the node that performs the background and stopping copy fails.

11.1.9 Incremental FlashCopy

When a FlashCopy mapping is stopped, either because the entire source volume was copied onto the target volume or because 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.

Using 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 point-in-time), 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, for a particular grain, if the data changed on the source or target volumes, the grain from the source volume is copied to the target.

The advantages of Incremental FlashCopy are only useful if a previous full copy of the source volume was obtained. Incremental FlashCopy only helps with further recovery time objectives (RTOs, time needed to recover data from a previous state), it does not help with the initial RTO.

For example, as shown in Figure 11-11 on page 450, a FlashCopy mapping was defined between a source volume and a target volume with the **-incremental** option. Consider the following points:

- The mapping is started on 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 were copied and the target volume is a replica of the source volume at the beginning of Copy1. Although the mapping is stopped (thanks to the **-incremental** option), the bitmap is maintained.
- 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. There is no need to copy the other grains, as they were already 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.

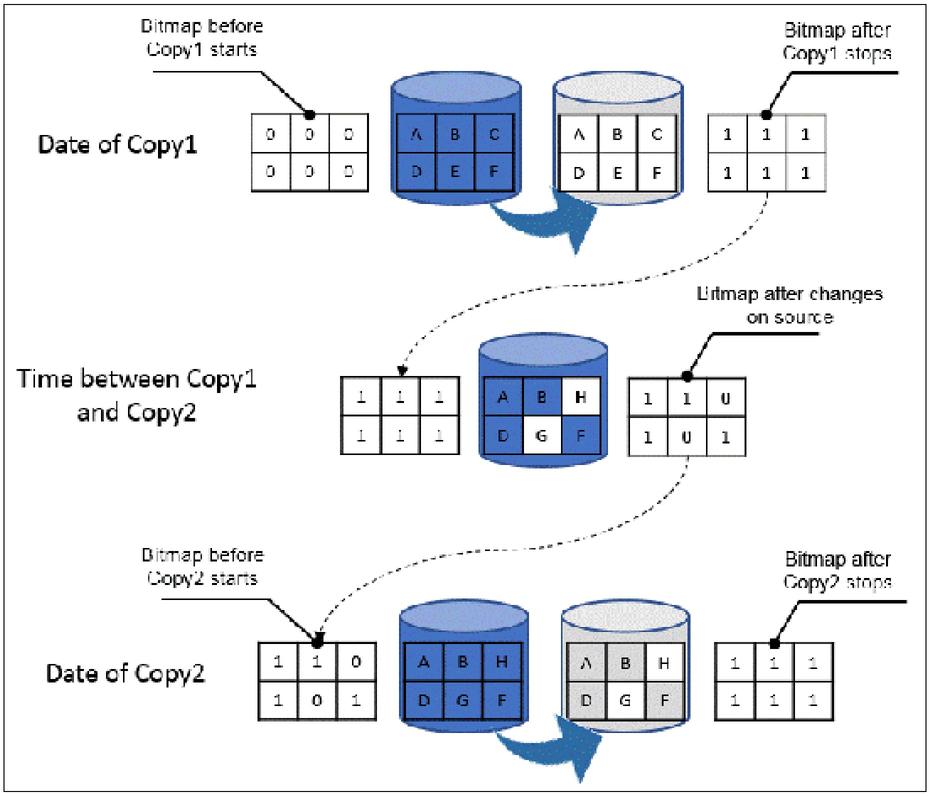


Figure 11-11 Incremental FlashCopy example

11.1.10 Starting FlashCopy mappings and Consistency Groups

You can perform the actions of preparing, starting, or stopping FlashCopy on a stand-alone mapping or a Consistency Group.

When using the CLI to perform FlashCopy on volumes, before you start a FlashCopy (regardless of the type and options specified), issue a **prestartfcmap** or **prestartfcconsistgrp** command. 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.

FlashCopy commands can then be issued to the FlashCopy Consistency Group and therefore, simultaneously for all of the FlashCopy mappings that are defined in the Consistency Group. For example, when a FlashCopy start command is issued 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 point-in-time copy that is consistent across all of the FlashCopy mappings that are contained in the Consistency Group.

Rather than using **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 both 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 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 not complete, 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 both 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 11-2 lists the various FlashCopy mapping states, and the corresponding states of the source and target volumes.

Table 11-2 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

11.1.11 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. But, a target volume can only have 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 11-12 on page 453 shows these different possibilities.

Every source-target relation is a FlashCopy mapping and is maintained with its own bitmap table. There is no consistency group bitmap table.

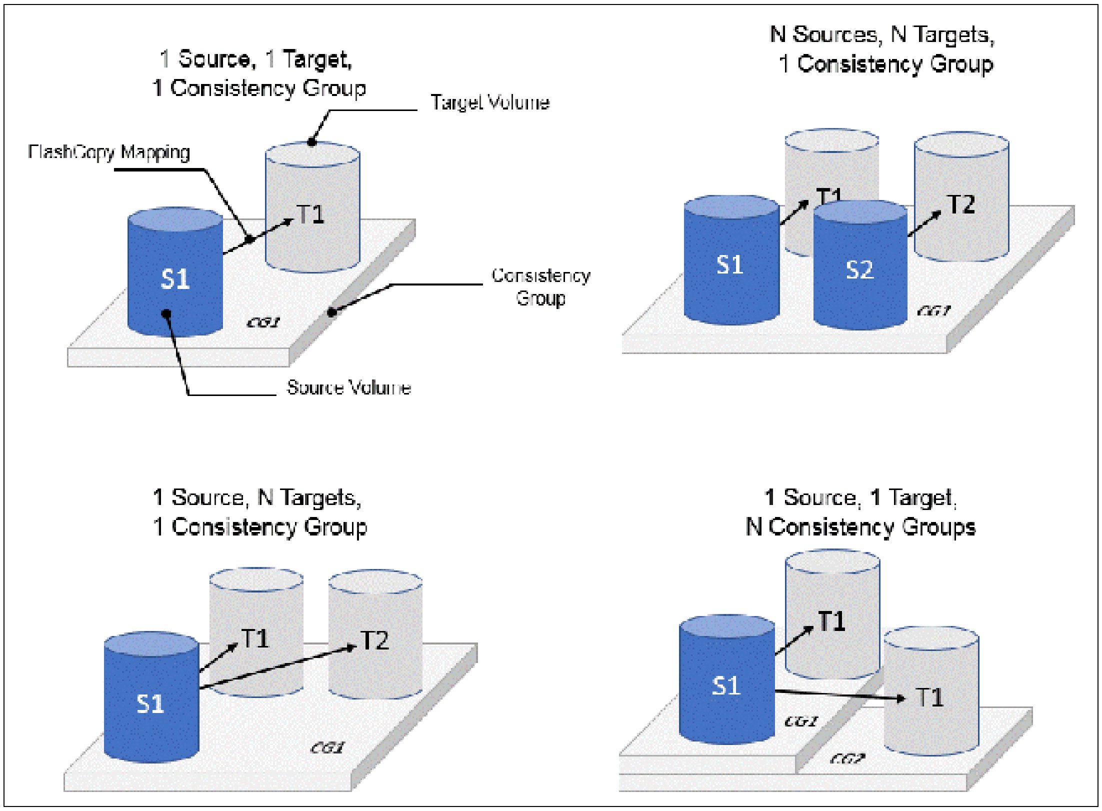


Figure 11-12   Consistency groups and mappings combinations

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. Therefore, multiple versions of a single volume are kept.

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 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 on the target volume because of the copy-on-write mechanism that is used by FlashCopy.

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). This generates a large workload that the system cannot handle and lead to a heavy performance impact on front-end operations.

To avoid any significant impact 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 11-13 shows an example of a source volume with three targets.

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.

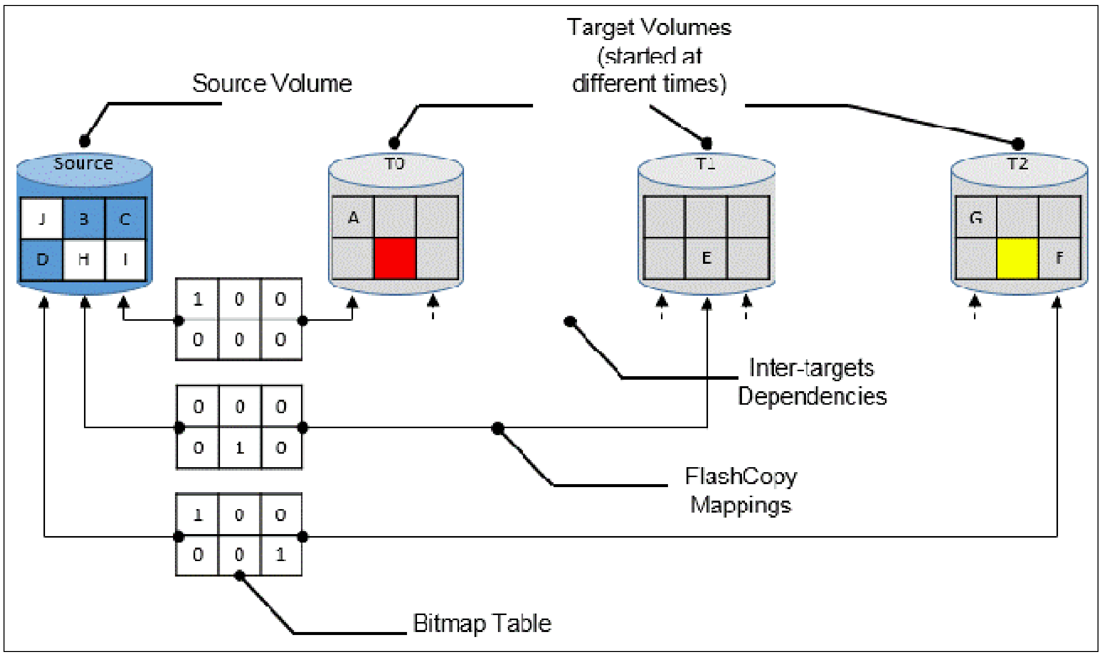


Figure 11-13 FlashCopy dependencies example

Source read with Multiple target FlashCopy

There is no specific behavior for read operations on source volumes when there are multiple targets for that volume. The data is always read from the source.

Source write with Multiple Target FlashCopy (Copy on Write)

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 11-3, for a source volume and three targets started at different times. In this example, there is no background copy. The “most recent” target is indicated with an asterisk.

Table 11-3 Sequence example of write IOs on a source with multiple targets

	Source Volume	Target T0	Target T1	Target T2
Time 0: mapping with T0 is started	A B C D E F	_ _ _* _ _ _	Not started	Not started
Time 1: change of “A” is made on source (->“G”)	G B C D E F	A _ _* _ _ _	Not started	Not started
Time 2: mapping with T1 is started	G B C D E F	A _ _ _ _ _	_ _ _* _ _ _	Not started

	Source Volume	Target T0	Target T1	Target T2
Time 3: change of “E” is made on source (->“H”)	G B C D H F	A _ _ _ _ _	_ _ _* _ E_	Not started
Time 4: mapping with T2 is started	G B C D H F	A _ _ _ _ _	_ _ _ _ E_	_ _ _* _ _ _
Time 5: change of “F” is made on source (->“I”)	G B C D H I	A _ _ _ _ _	_ _ _ _ E_	_ _ _* _ _ F
Time 6: change of “G” is made on source (->“J”)	J B C D H I	A _ _ _ _ _	_ _ _ _ E_	G _ _* _ _ F
Time 7: stop of Source-T2 mapping	J B C D H I	A _ _ _ _ _	G _ _ * _ E F	Stopped
Time 8: stop of Source-T1 mapping	J B C D H I	A _ _ * _ E F	Stopped	Stopped

* *“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:

- If the grain that is being read is copied from the source to the target, the read returns data from the target that is being 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, as shown in Figure 11-13 on page 454, if the yellow grain on T2 is read, it returns “H” because there is no newer target than T2. Therefore, the source is read.

  As another example, as shown in Figure 11-13 on page 454, if the red grain on T0 is read, it returns “E” because there are two newer targets for T0, and T1 is the oldest of them.

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 being 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, as shown in Figure 11-13 on page 454, if the grain “G” is changed on T2, then it needs to 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, as shown in Figure 11-13 on page 454, if the red grain on T0 is written, then it will first be copied from T1 (data “E”). After this copy is done, the write can be applied to the target.

Table 11-4 lists the indirection layer algorithm in a multi-target FlashCopy.

Table 11-4 Summary table of the FlashCopy indirection layer algorithm t

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 has 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, with Table 11-3 on page 454, if you stop the Source-T2 mapping on “Time 7,” 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 11-3 on page 454, 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 like 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 the FlashCopy mapping and can be changed dynamically. It is measured like the copyrate property, but both properties are independent. Table 11-5 lists the relationship of the cleaning rate values to the attempted number of grains to be split per second.

Table 11-5 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

11.1.12 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 therefore multiple rollback points.

A key advantage of the IBM Spectrum Virtualize Multiple Target Reverse FlashCopy function is that the reverse FlashCopy does not destroy the original target. This feature enables processes that are using the target, such as a tape backup or tests, to continue uninterrupted.

IBM Spectrum Virtualize also provides the ability to 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 11-14 shows an example of Reverse FlashCopy with a simple FlashCopy mapping (single target).

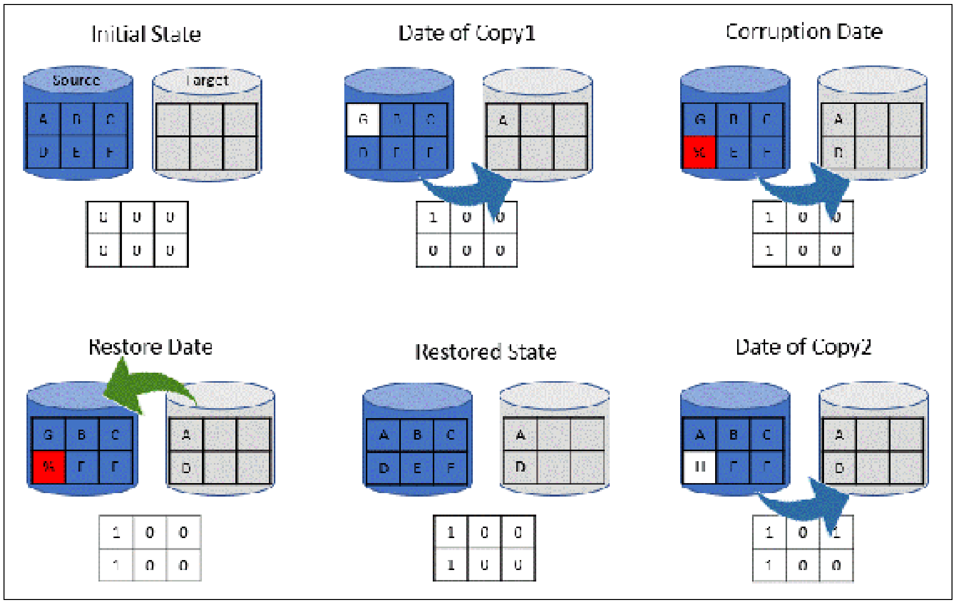


Figure 11-14   A reverse FlashCopy example for data restoration

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 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 these steps:

1. Creating a mapping from target to source (if not 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 bitmap table instead, with its updated bits.
2. Start the new mapping. Because of the 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.

11.1.13 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 with the CLI, use 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 11-1 shows the size of the ITSO-TSTRS001 volume. The ITSO-TSTRS002 volume is then created, which specifies the same size.

Example 11-1 Listing the size of a volume in bytes and creating a volume of equal size

```
IBM_Storwize:ITSO-V7k:superuser>lsvdisk -bytes ITSO-TSTRS001
id 19
name ITSO-TSTRS001
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name Pool0
capacity 1073741824
type striped
......

IBM_Storwize:ITSO-V7k:superuser>mkvdisk -mdiskgrp 0 -iogrp 0 -size 1073741824
-unit b -name ITSO-TSTRS002
Virtual Disk, id [20], successfully created
IBM_Storwize:ITSO-V7k:superuser>

IBM_Storwize:ITSO-V7k:superuser>lsvdisk -delim " "
19 ITSO-TSTRS001 0 io_grp0 online 0 Pool0 1.00GB striped
6005076400A7015B0800000000000016 0 1 not_empty 0 no 0 0 Pool0 yes no 19
ITSO-TSTRS001
20 ITSO-TSTRS002 0 io_grp0 online 0 Pool0 1.00GB striped
6005076400A7015B0800000000000017 0 1 not_empty 0 no 0 0 Pool0 yes no 20
ITSO-TSTRS002
IBM_Storwize:ITSO-V7k:superuser>
```

Tip: Alternatively, you can use the **expandvdisksize** and **shrinkvdisksize** volume commands to modify the size of the volume.

These actions must be performed before a mapping is created.

11.1.14 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 11-6 on page 460.

Overview of a FlashCopy sequence of events: The following tasks show the FlashCopy sequence:

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 11-6 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: ► 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 **prestartfcmap** or **prestartfcconsistgrp** 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 **prestartfcmap** or **prestartfcconsistgrp** command places the FlashCopy mapping into the Preparing state. The **prestartfcmap** or **prestartfcconsistgrp** 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	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 using the **`startfcmap`** or **`startfcconsistgrp`** command. The following actions occur during the running of the **`startfcmap`** command or the **`startfcconsistgrp`** command: ▶ 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. As part of the **`startfcmap`** or **`startfcconsistgrp`** command, read and write caching is enabled for the source and target volumes.
Modify	The following FlashCopy mapping properties can be modified: ▶ 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	The following separate mechanisms can be used to stop a FlashCopy mapping: ▶ Issue a command ▶ An I/O error occurred
Delete	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.
Flush failed	If the flush of data from the cache cannot be completed, the FlashCopy mapping enters the `stopped` state.
Copy complete	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.
Bitmap online/offline	The node failed.

11.1.15 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 that is used on the storage system) is lower, as shown on Figure 11-15. 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.

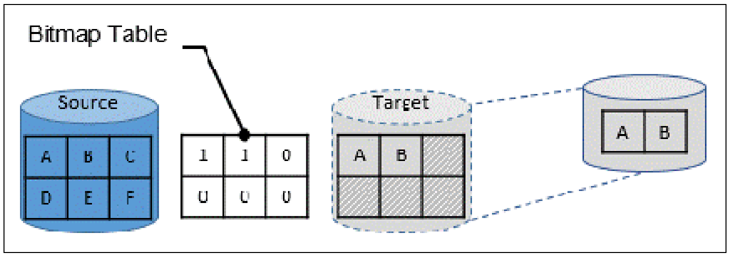


Figure 11-15 Thin-provisioned target volume

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.

11.1.16 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, there is a lock for each grain. The 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 already 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.

11.1.17 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.

This section describes these scenarios.

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 SAN Volume Controller 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 happens, the managed disks are said to be *Path Offline* on certain nodes.

Other nodes: Other nodes might see the managed disks as Online because their connection to the managed disks is still up.

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.

11.1.18 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.

11.1.19 Interoperation with Metro Mirror and Global Mirror

A volume can be part of any copy relationship (FlashCopy, Metro Mirror, or Remote Mirror). Therefore, FlashCopy can work with MM/GM to provide better protection of the data.

For example, you can perform a Metro Mirror 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, Metro Mirror, or Remote Mirror if it is set to Transparent Cloud Tiering function.

Table 11-7 on page 465 lists the supported combinations of FlashCopy and remote copy. In the table, *remote copy* refers to Metro Mirror and Global Mirror.

Table 11-7 FlashCopy and remote copy interaction

Component	Remote copy primary site	Remote copy secondary site
FlashCopy Source	Supported	Supported latency: When the FlashCopy relationship is in the `preparing` and `prepared` states, the cache at the remote copy secondary site operates in write-through mode. This process adds latency to the latent remote copy relationship.
FlashCopy Target	This is a supported combination and has the following restrictions: ► Issuing a **`stop -force`** might cause the remote copy relationship to be fully resynchronized. ► Code level must be 6.2.*x* or later. ► I/O Group must be the same.	This is a supported combination with the major restriction that the FlashCopy mapping cannot be `copying`, `stopping`, or `suspended`. Otherwise, the restrictions are the same as at the remote copy primary site.

11.1.20 FlashCopy attributes and limitations

The FlashCopy function in IBM Spectrum Virtualize features the following attributes:

- The target is the time-zero 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 operating system, 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.
- The source and target volumes must be on the same Storwize V7000 system.
- The source and target volumes do not need to be in the same I/O Group or storage pool.
- 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*). But a target volume can only have 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 Storwize V7000 I/O Group to prevent a single point of failure.
- FlashCopy mapping and Consistency Groups can be automatically withdrawn after the completion of the background copy.
- Thin-provisioned FlashCopy (or Snapshot in the graphical user interface (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.

IBM FlashCopy limitations for IBM Spectrum Virtualize V8.2 are listed in Table 11-8.

Table 11-8 FlashCopy limitations in V8.2

Property	Maximum number
FlashCopy mappings per system	5000
FlashCopy targets per source	256
FlashCopy mappings per consistency group	512
FlashCopy consistency groups per system	255
Total FlashCopy volume capacity per I/O group	4096 TiB

11.2 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.

11.2.1 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 copy-on-write point-in-time copy. 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 11-16, the source volume user can still work on the original data volume (such as 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 need to remain linked to the source. Whenever the link is broken (FlashCopy mapping stopped or deleted), the target volumes become unusable.

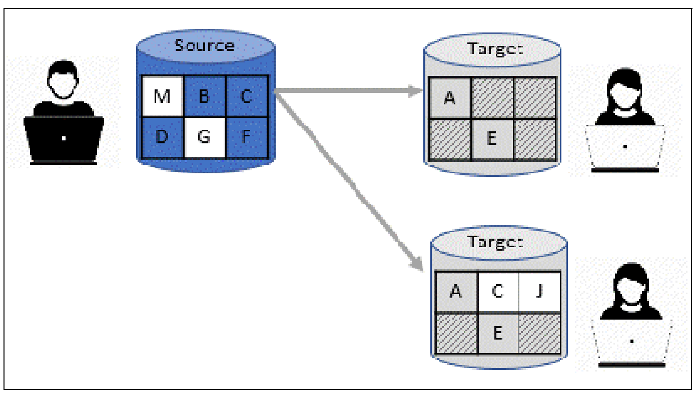


Figure 11-16 FlashCopy snapshot preset example

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, there is no expectation that it is refreshed or that there is any other need to reference the original production data again. If the source is thin-provisioned, the target is thin-provisioned for the auto-create target.

Backup
The backup preset creates an incremental point-in-time 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 due to 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 new 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. FlashCopy backup preset, for example, does not schedule a regular copy of your volumes. It over-writes 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.

11.2.2 FlashCopy window

This section describes the tasks that you can perform at a FlashCopy level by using the IBM Spectrum Virtualize GUI.

When using the IBM Spectrum Virtualize GUI, FlashCopy components can be seen in different windows. Three windows are related to FlashCopy and are reachable through the **Copy Services** menu, as shown in Figure 11-17.

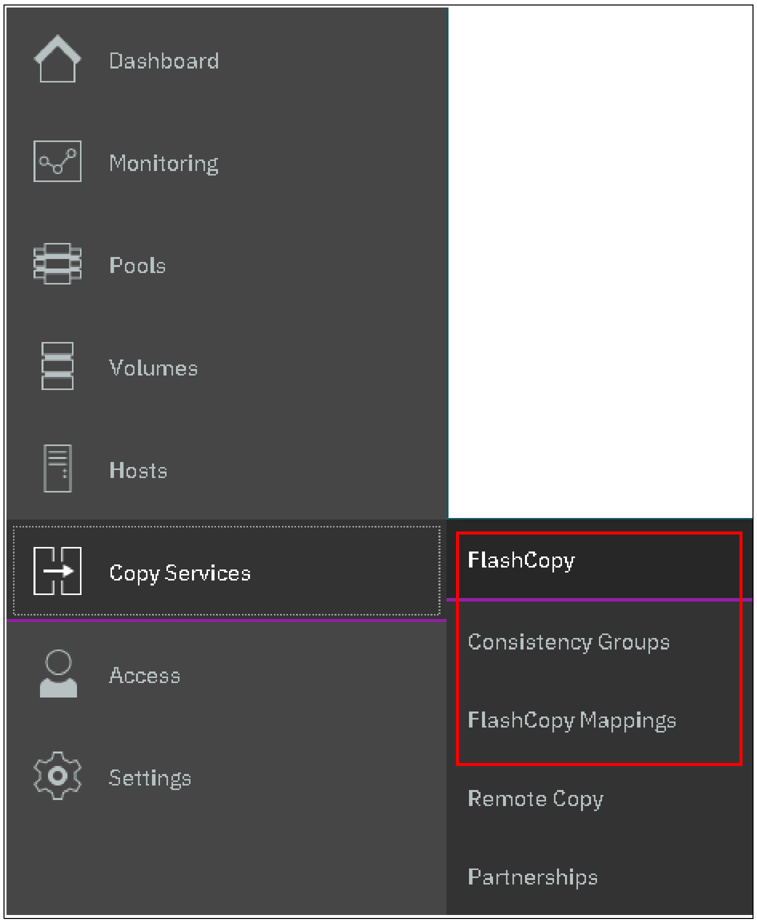


Figure 11-17 Copy Services menu

The FlashCopy window is accessible by clicking **Copy Services** → **FlashCopy**. All of the volumes that are defined in the system are displayed. Volumes that are part of a FlashCopy mapping appear as shown in Figure 11-18. By clicking a source volume, you can display the list of its target volumes.

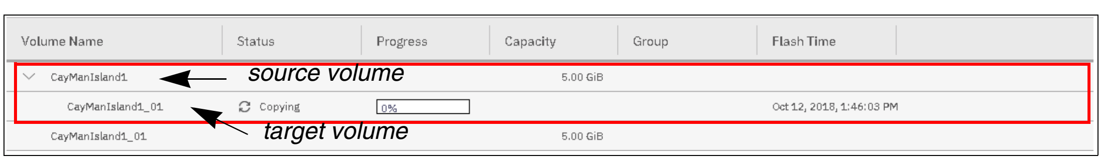


Figure 11-18 Source and target volumes displayed in the FlashCopy window

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):

- The Consistency Group window is accessible by clicking **Copy Services** → **Consistency Groups**. Use the Consistency Groups window, as shown in Figure 11-19, to list the FlashCopy mappings that are part of consistency groups and part of no consistency groups.

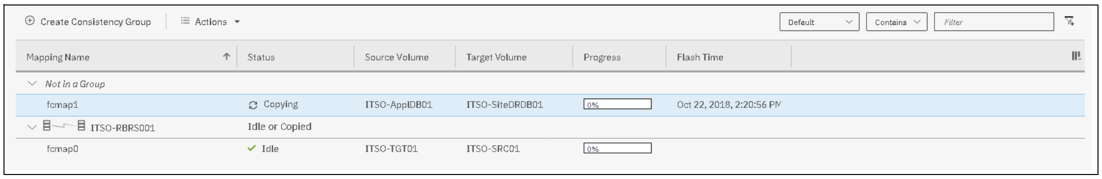


Figure 11-19 Consistency Groups window

- The FlashCopy Mappings window is accessible by clicking **Copy Services** → **FlashCopy Mappings**. Use the FlashCopy Mappings window, as shown in Figure 11-20, to display the list of mappings between source volumes and target volumes.

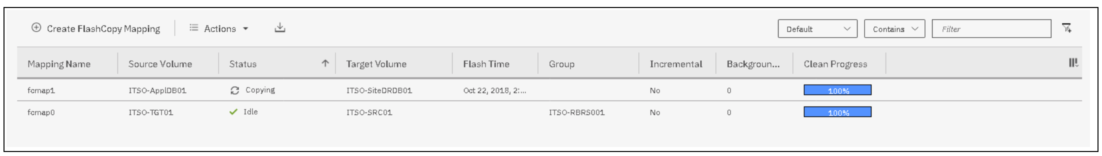


Figure 11-20 FlashCopy mapping pane

11.2.3 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 11-21 on page 471. Select the volume for which you want to create the FlashCopy mapping. Right-click the volume or click the **Actions** menu.

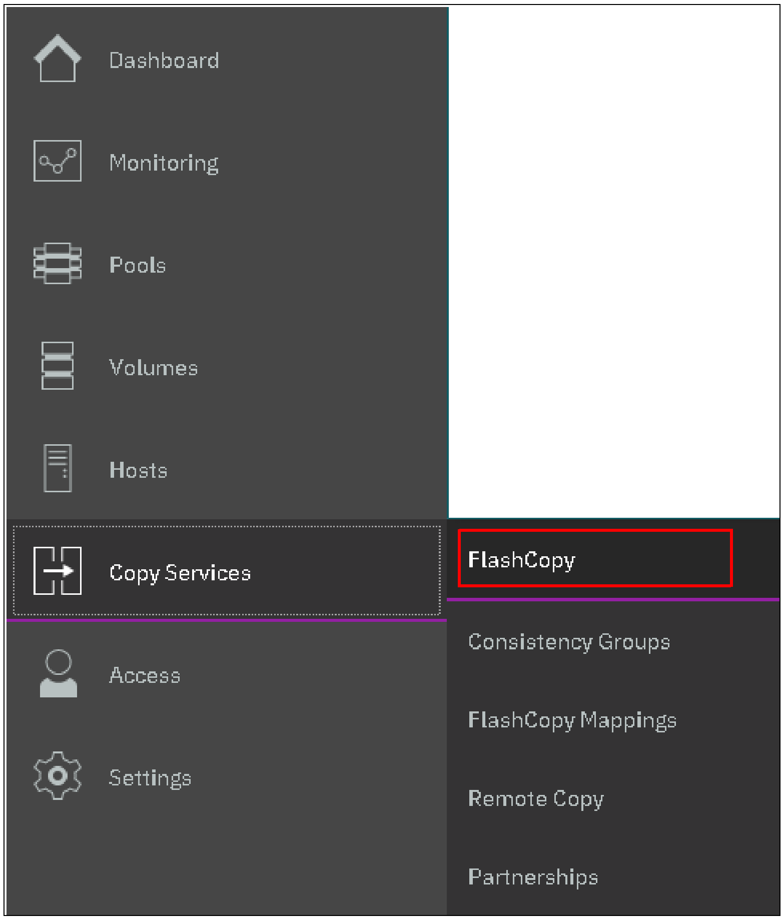


Figure 11-21 FlashCopy window

Multiple FlashCopy mappings: To create multiple FlashCopy mappings at once, select multiple volumes by holding down **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 471.
- If you want the system to create the target volumes for you, see "Creating a FlashCopy mapping and target volumes" on page 475.

Creating a FlashCopy mapping with existing target Volumes

Complete the following steps to use 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 already 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 11-22 on page 472.

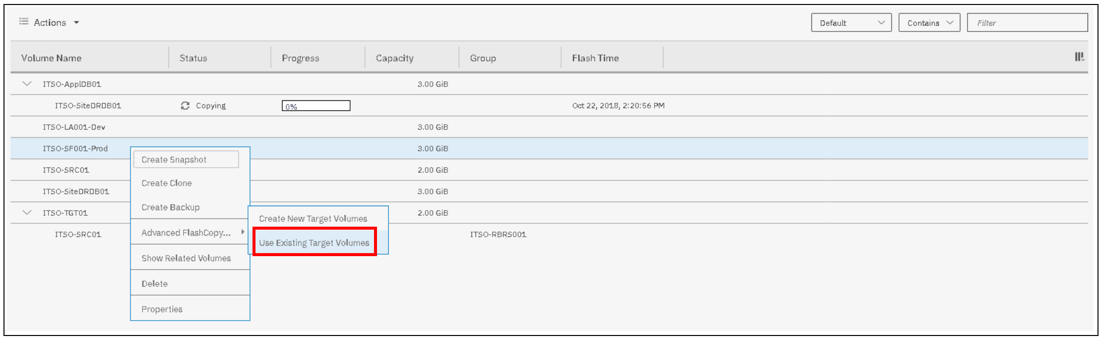


Figure 11-22 Creating a FlashCopy mapping with an existing target

The Create FlashCopy Mapping window opens as shown in Figure 11-23. 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.

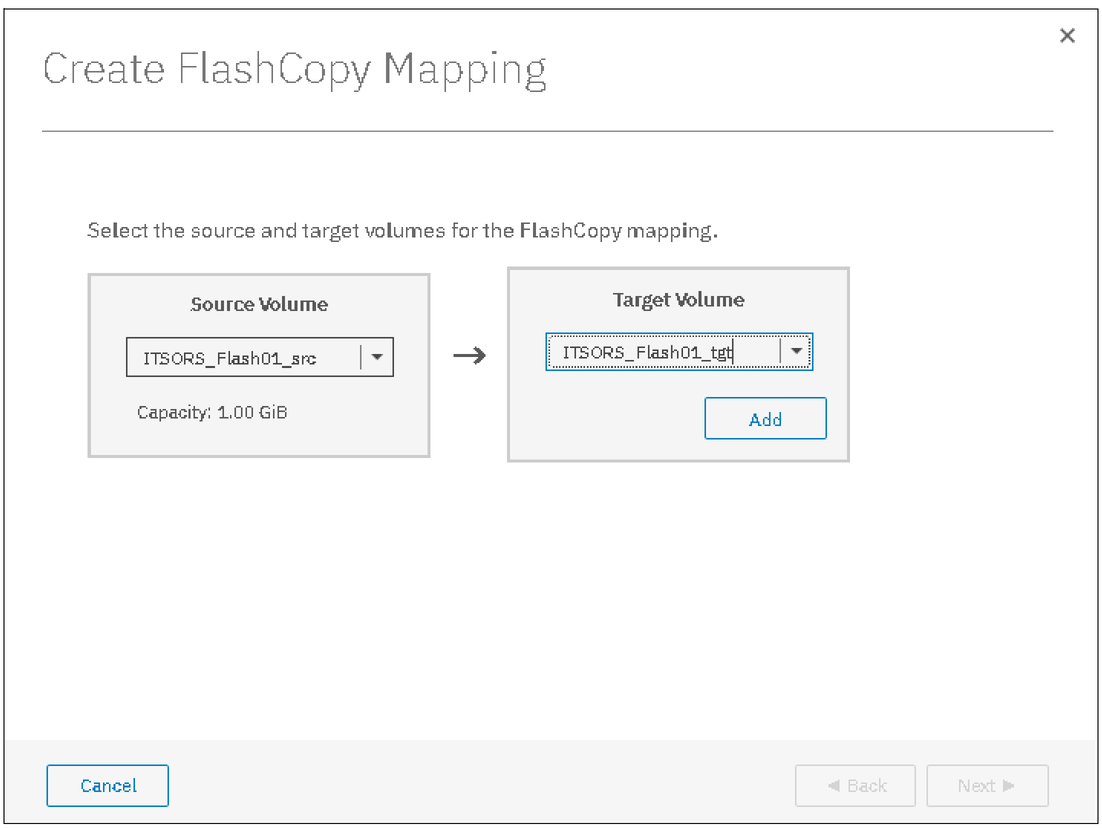


Figure 11-23 Selecting source and target for a FlashCopy mapping

To remove a mapping that was created, click , as shown in Figure 11-24 on page 473.

2. Click **Next** after you create all of the mappings that you need, as shown in Figure 11-24.

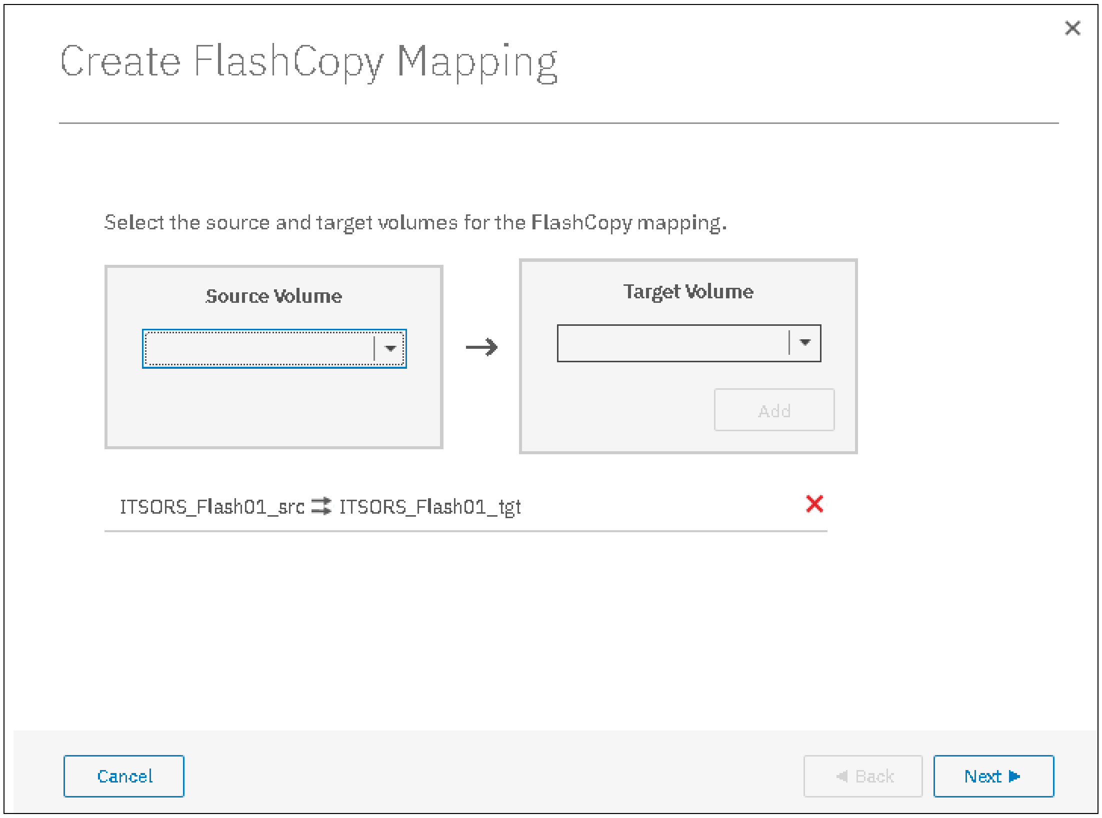


Figure 11-24 Viewing source and target at creation time

3. In the next window, select one FlashCopy preset. The GUI provides the following presets to simplify common FlashCopy operations, as shown in Figure 11-25 on page 474. For more information about the presets, see 11.2.1, “FlashCopy presets” on page 466:
   - Snapshot: Creates a point-in-time snapshot copy of the source volume.
   - Clone: Creates a point-in-time 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 that it is thin-provisioned first. Using the Snapshot preset does not make the system check whether the target volume is thin-provisioned.

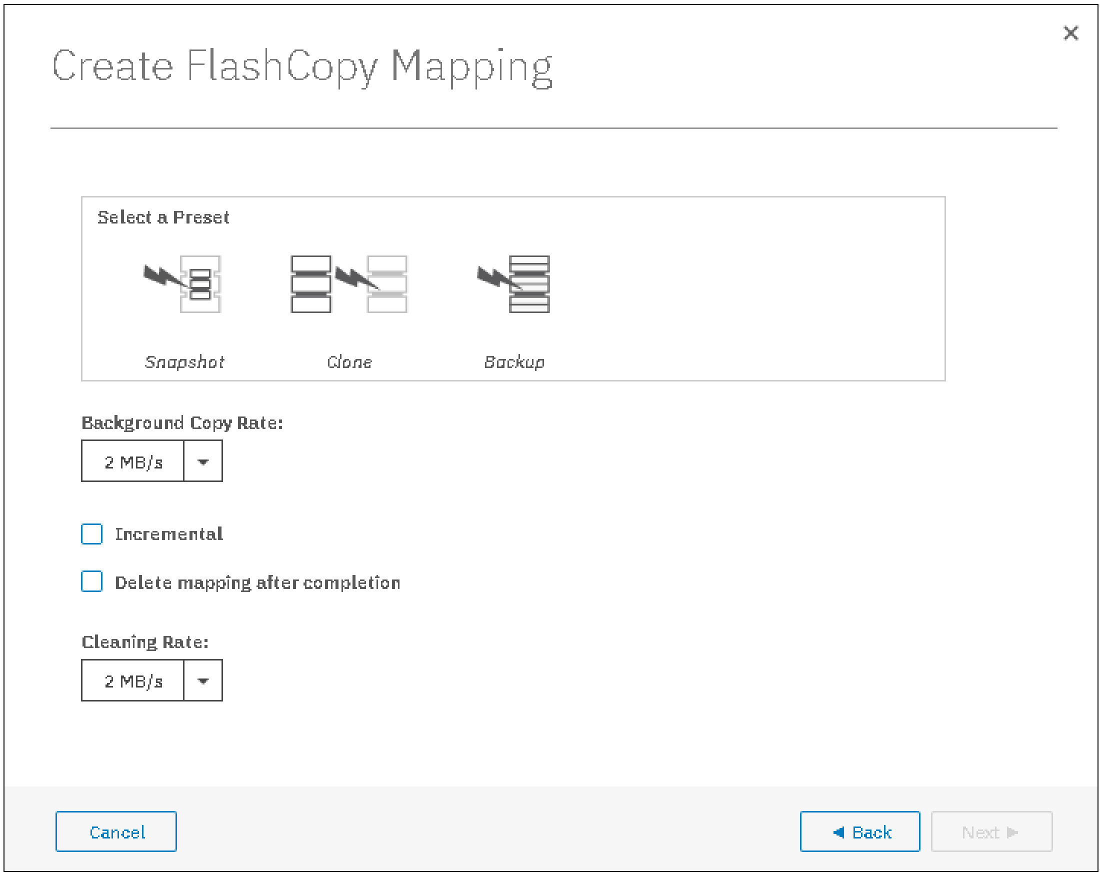


Figure 11-25 FlashCopy mapping preset selection

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 change is not recommended for the following reasons:

– For example, 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.

– As another example, 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 Cleaning Rate options, see Table 11-1 on page 448 or Table 11-5 on page 457.

When your FlashCopy mapping setup is ready, click **Next**.

4. You can choose whether to add the mappings to a Consistency Group as shown in Figure 11-26.

   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.

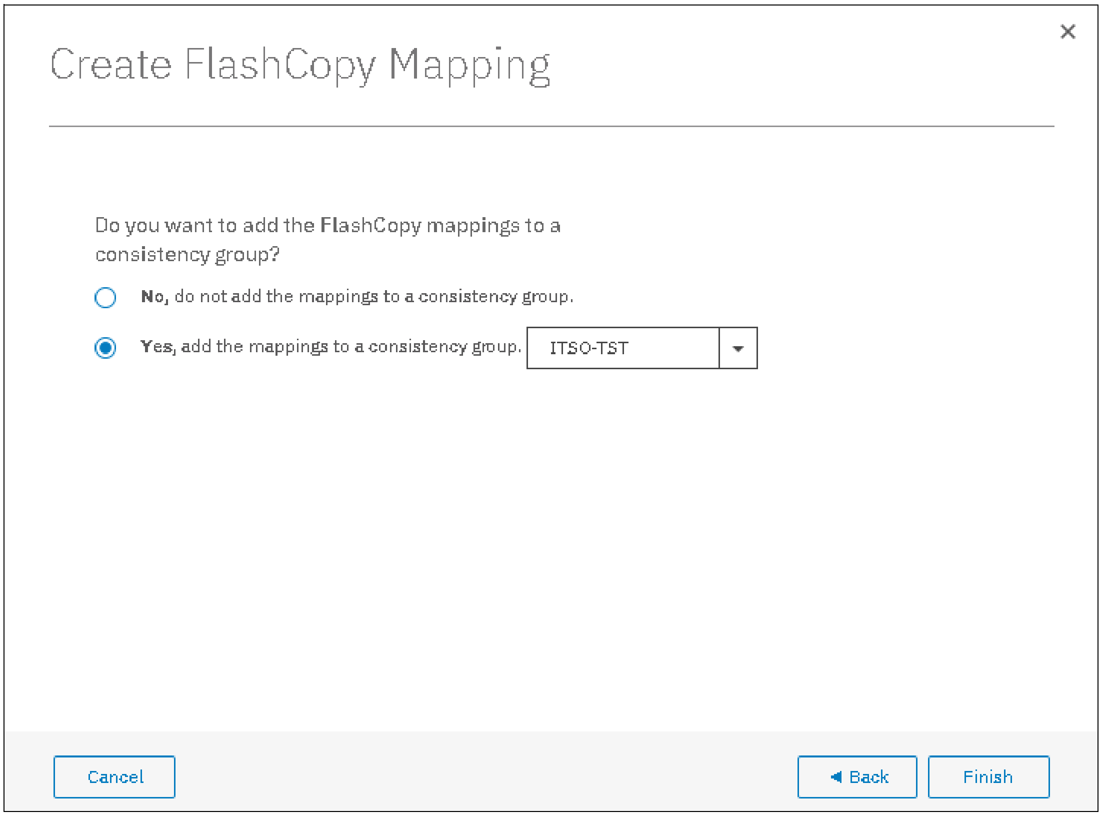


Figure 11-26   Select or not a Consistency Group for the FlashCopy mapping

5. 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 what to do at this stage, 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 need to 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 11-27 on page 476.

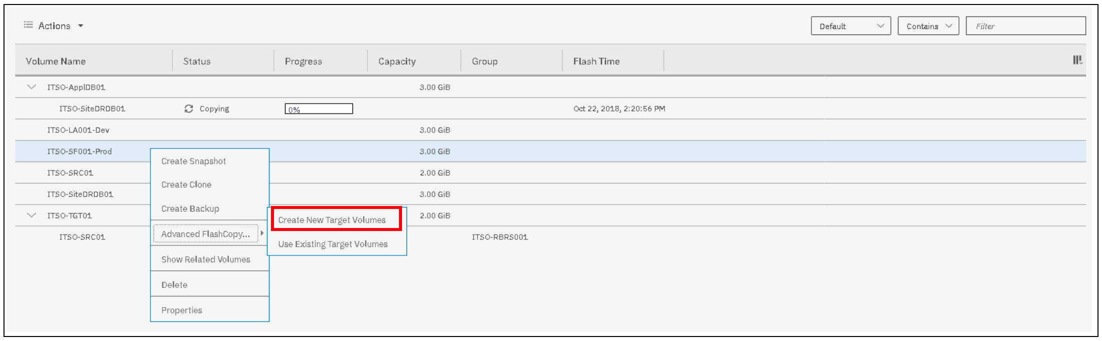


Figure 11-27 Creating a FlashCopy mapping and creating targets

2. In the next window, select one FlashCopy preset. The GUI provides the following presets to simplify common FlashCopy operations, as shown in Figure 11-28. For more information about the presets, see 11.2.1, “FlashCopy presets” on page 466:
   - Snapshot: Creates a point-in-time snapshot copy of the source volume.
   - Clone: Creates a point-in-time 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 probably want the target volume to be defined as thin-provisioned to save space on your system. If you use an existing target, ensure that it is thin-provisioned first. Using the **Snapshot** preset does not make the system check whether the target volume is thin-provisioned.

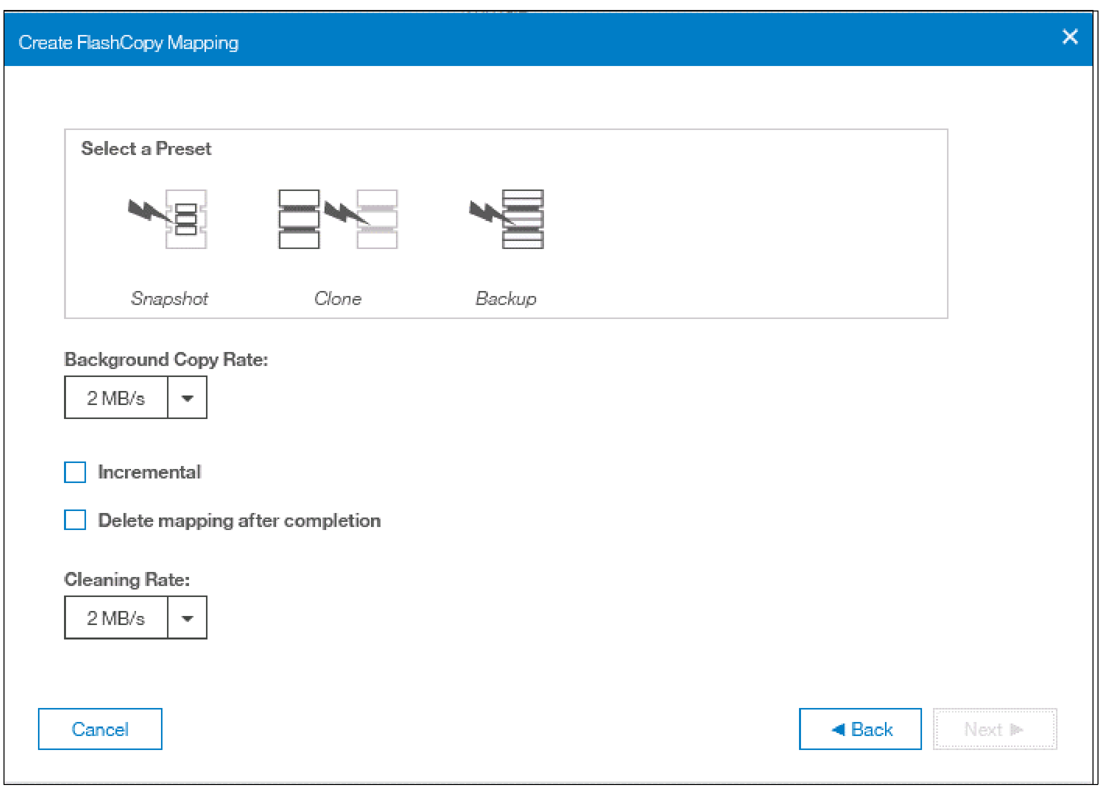


Figure 11-28 FlashCopy mapping preset selection

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 change is not recommended for the following reasons:

– For example, 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.

– As another example, 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 Cleaning Rate options, see Table 11-1 on page 448 or see Table 11-5 on page 457.

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 11-29.

   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.

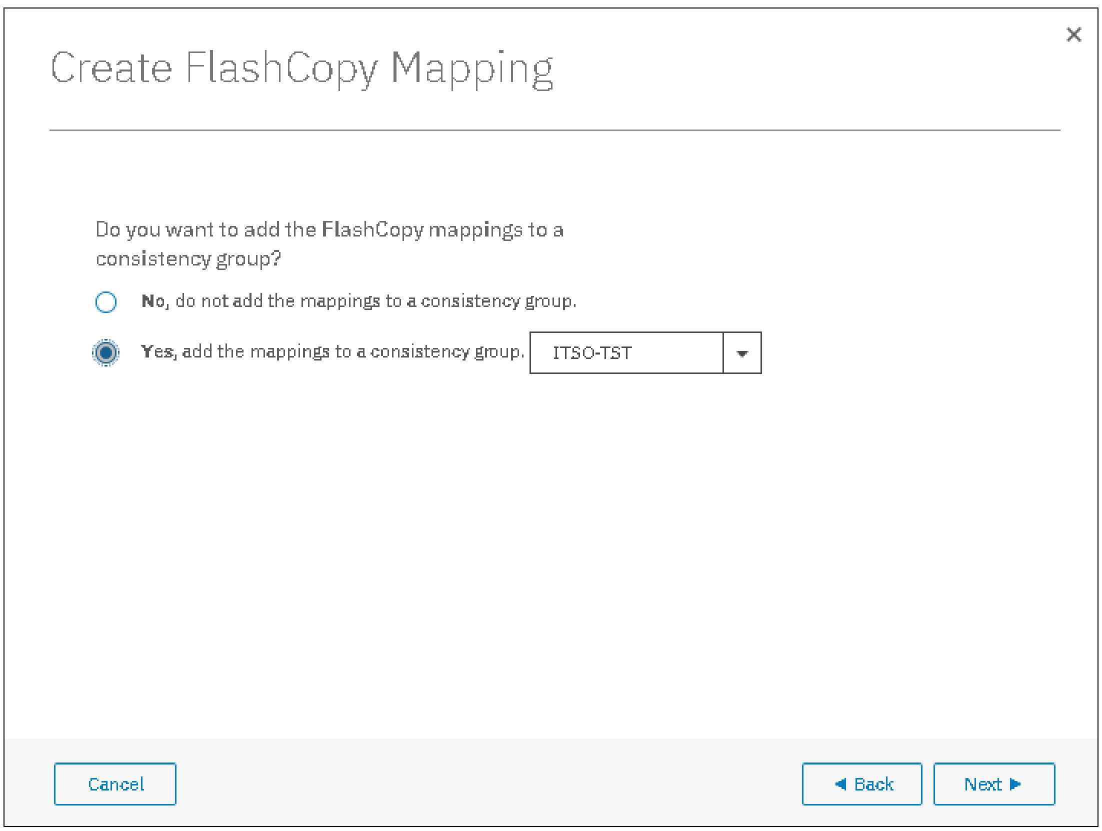


Figure 11-29 Select a Consistency Group for the FlashCopy mapping

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 what to do at this stage, you can change it later. Click **Next**.

5. In this step, the system asks the user to select the pool that is used to automatically create new targets, as shown in Figure 11-30. Click **Next**.

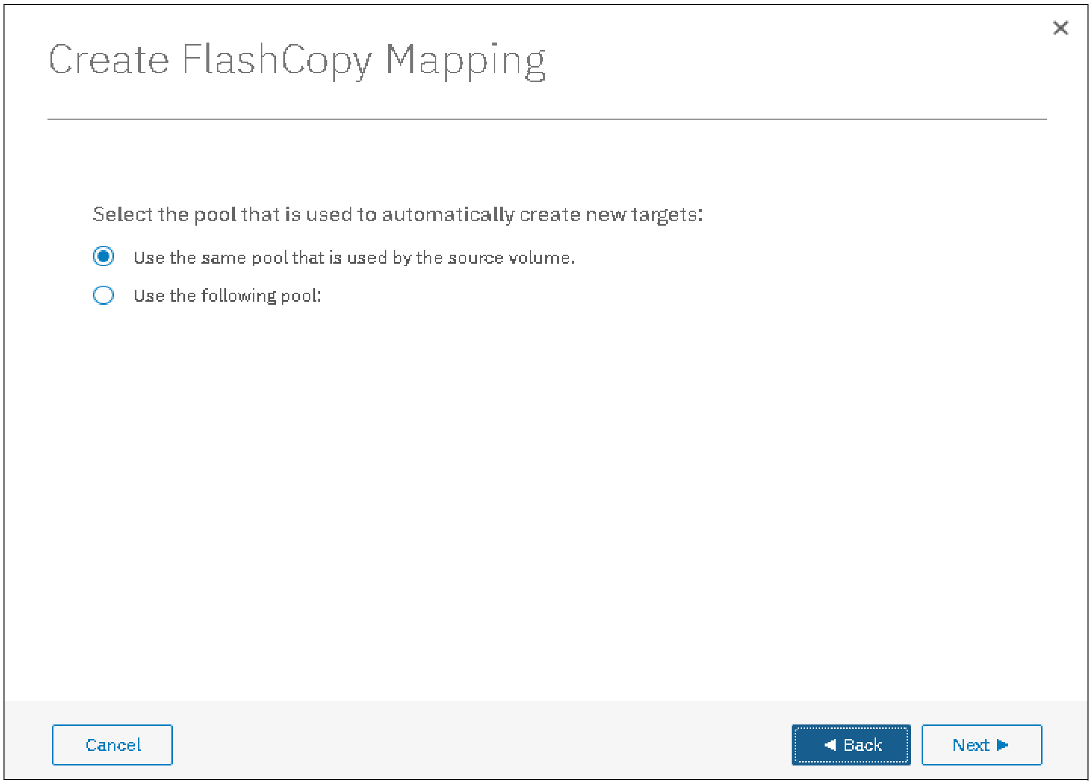


Figure 11-30 Select the pool

6. In this step, the system asks the user how to define the new volumes that are created, as shown in Figure 11-31 on page 479. It can be None, Thin-provisioned, or Inherit from source volume. If selected Inherit from source volume, the system checks the type of the source volume and then creates a target of the same type. Click **Finish**.

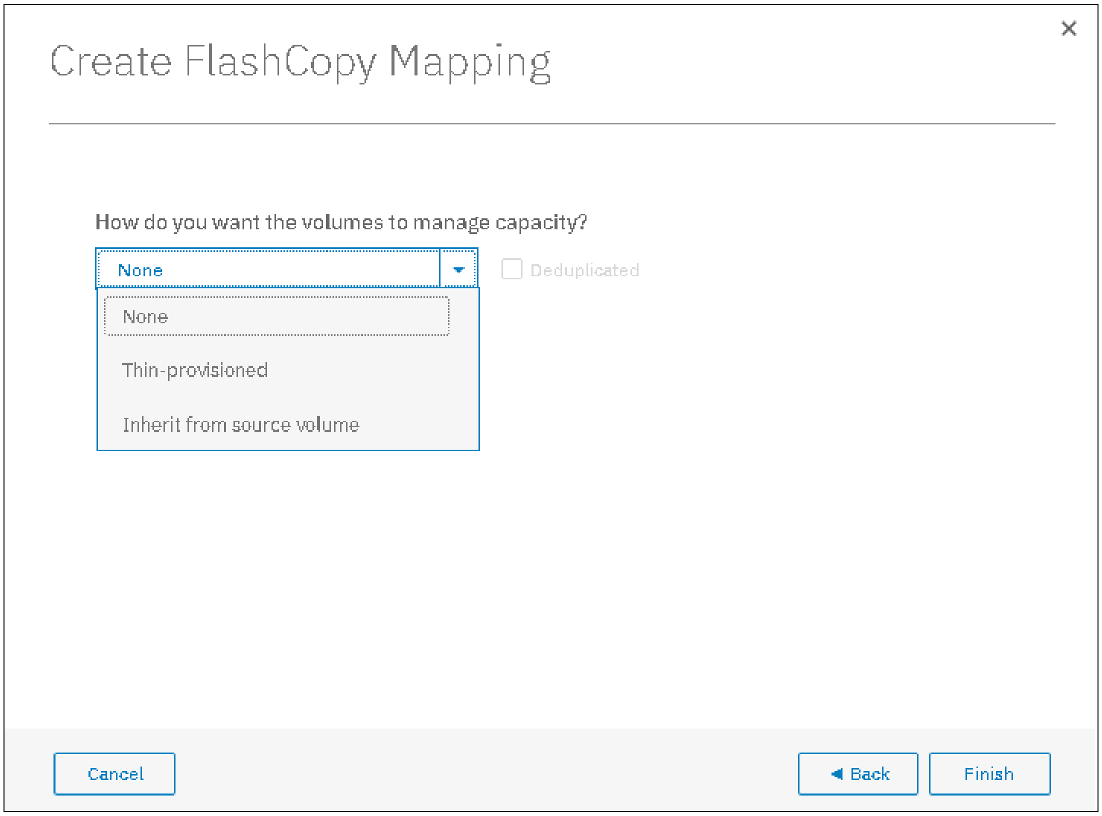


Figure 11-31   Select the type of volumes for the created targets

Note: If you selected multiple source volumes to create FlashCopy mappings, if you select **Inherit properties from source Volume**, it applies to each 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.

11.2.4 Single-click snapshot

The *snapshot* creates a point-in-time 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 11-32.

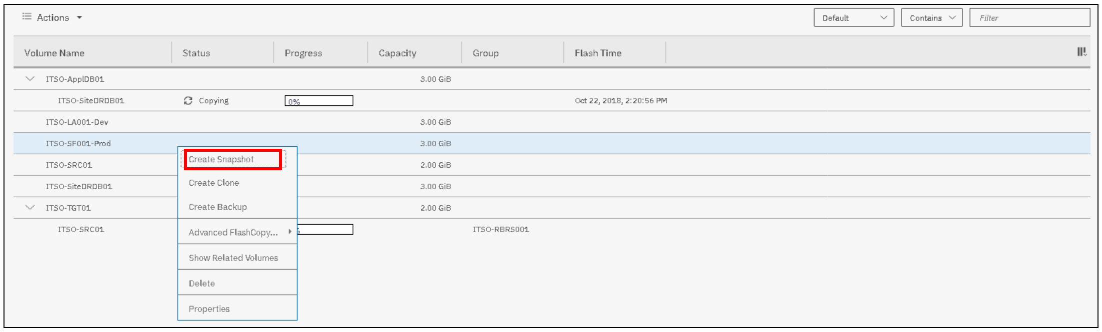


Figure 11-32 Single-click snapshot creation and start

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 11-33.

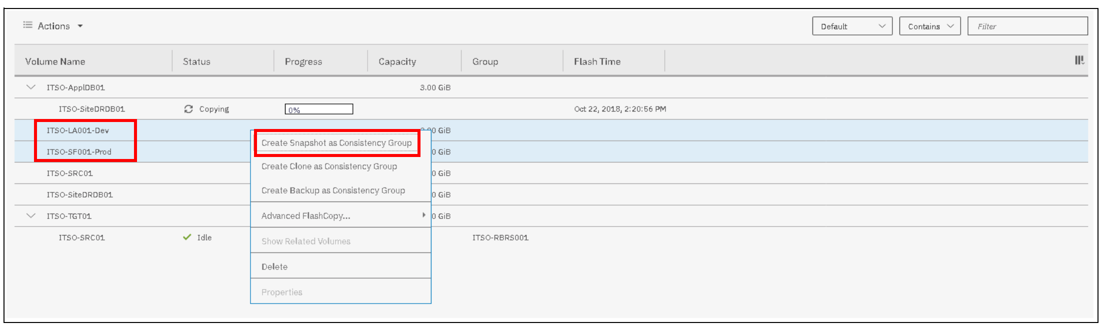


Figure 11-33 Selection single-click snapshot creation and start

4. For each selected source volume, the following actions occur:
   – A FlashCopy mapping is automatically created. It is named by default `fcmapXX`.
   – 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. Consistency groups are named by default `fccstgrpX`.

The newly created consistency group is automatically started.

11.2.5 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 11-34.

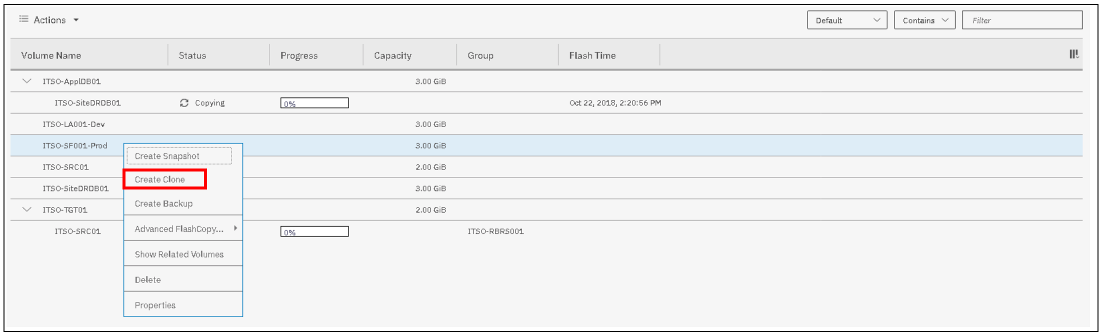


Figure 11-34 Single-click clone creation and start

3. You can select multiple volumes at a time, which creates many snapshots automatically. The system then automatically groups the FlashCopy mappings in a new consistency group, as shown in Figure 11-35.

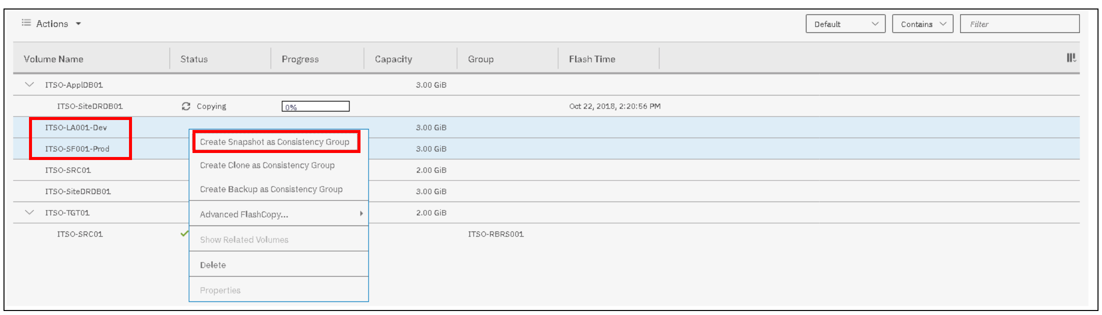


Figure 11-35 Selection single-click clone creation and start

4. For each selected source volume, the following actions occur:
   - A FlashCopy mapping is automatically created. It is named by default `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. Consistency groups are named by default `fccstgrpX`.
   - The newly created consistency group is automatically started.

11.2.6 Single-click backup

The backup creates a point-in-time 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 11-36.

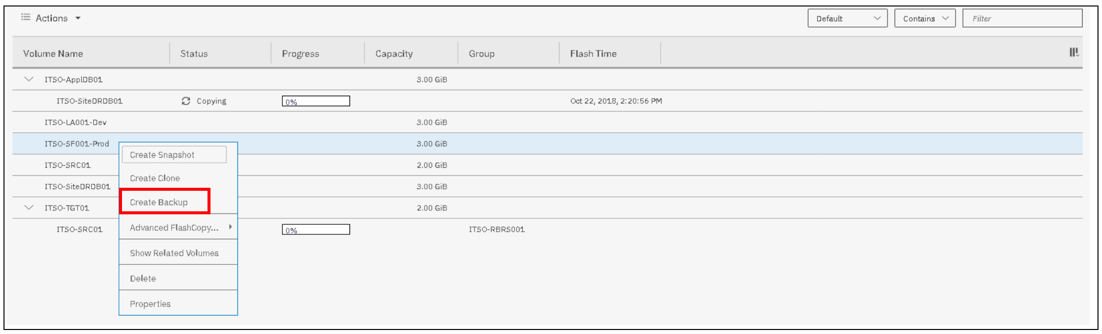


Figure 11-36 Single-click backup creation and start

3. You can select multiple volumes at a time, which creates many snapshots automatically. The system then automatically groups the FlashCopy mappings in a new consistency group, as shown Figure 11-37.

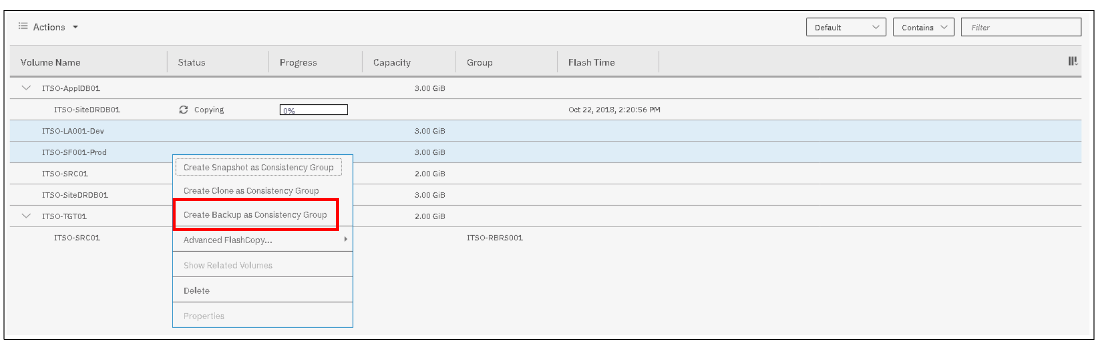


Figure 11-37 Selection single-click backup creation and start

4. For each selected source volume, the following actions occur:
   - A FlashCopy mapping is automatically created. It is named by default `fcmapXX`.
   - 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. Consistency groups are named by default `fccstgrpX`.
   - The newly created consistency group is automatically started.

11.2.7 Creating a FlashCopy 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**. Click **Create Consistency Group**, as shown in Figure 11-38 on page 483.

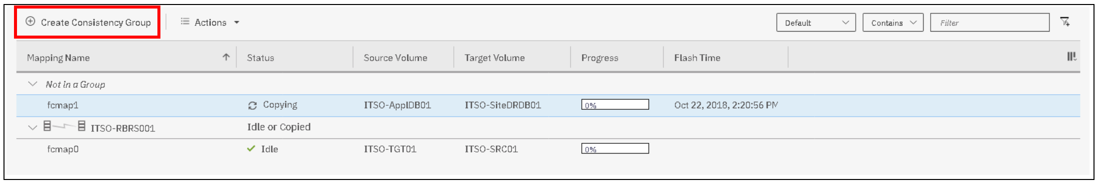


Figure 11-38 Creating a consistency group

2. Enter the FlashCopy Consistency Group name that you want to use and click **Create**, as shown in Figure 11-39.

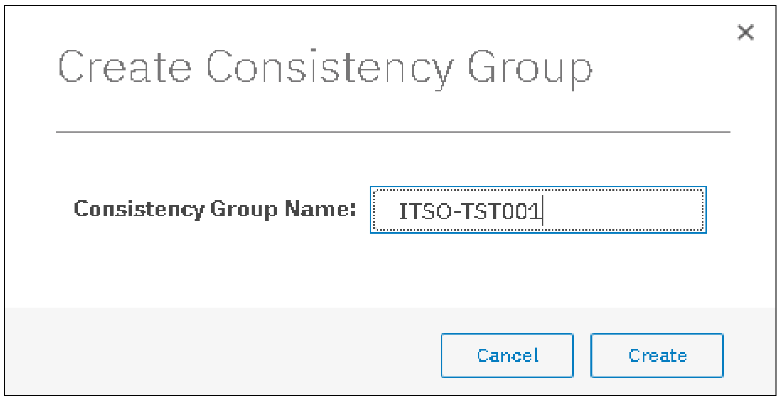


Figure 11-39 Enter the name of new consistency group

> **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.

11.2.8 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 created previously.
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 11-40.

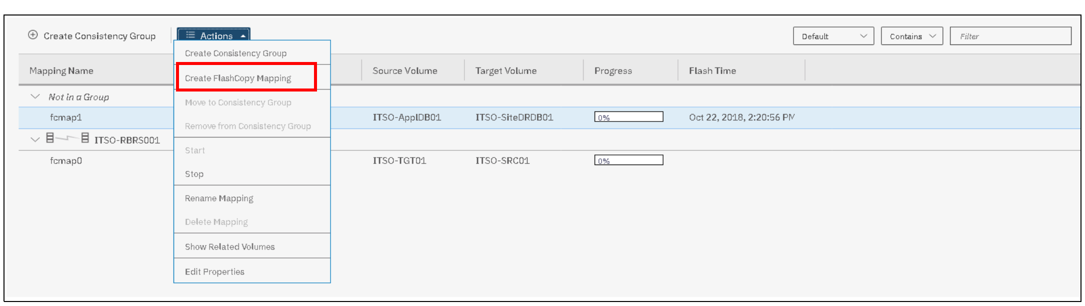


Figure 11-40 Creating a FlashCopy mapping

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 11-41.

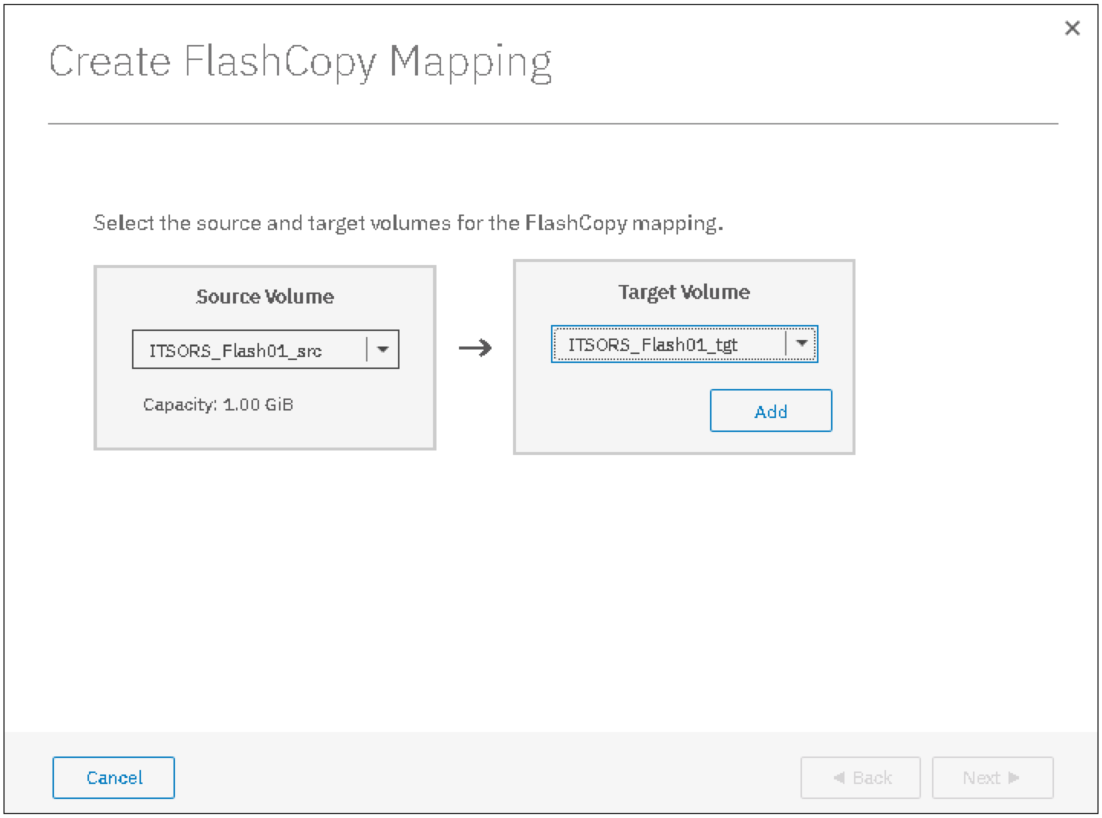


Figure 11-41 Select source and target volumes for the FlashCopy mapping

Repeat this step to create other mappings. To remove a mapping that was created, click . Click **Next.**

> **Important:** The source and target volumes must be of equal size. Therefore, only the targets with the appropriate size are shown for a source volume.
>
> Volumes that are target volumes in another FlashCopy mapping cannot be the 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 11-42 on page 485. For more information about the presets, see 11.2.1, “FlashCopy presets” on page 466:
   – Snapshot: Creates a point-in-time snapshot copy of the source volume.
   – Clone: Creates a point-in-time 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.

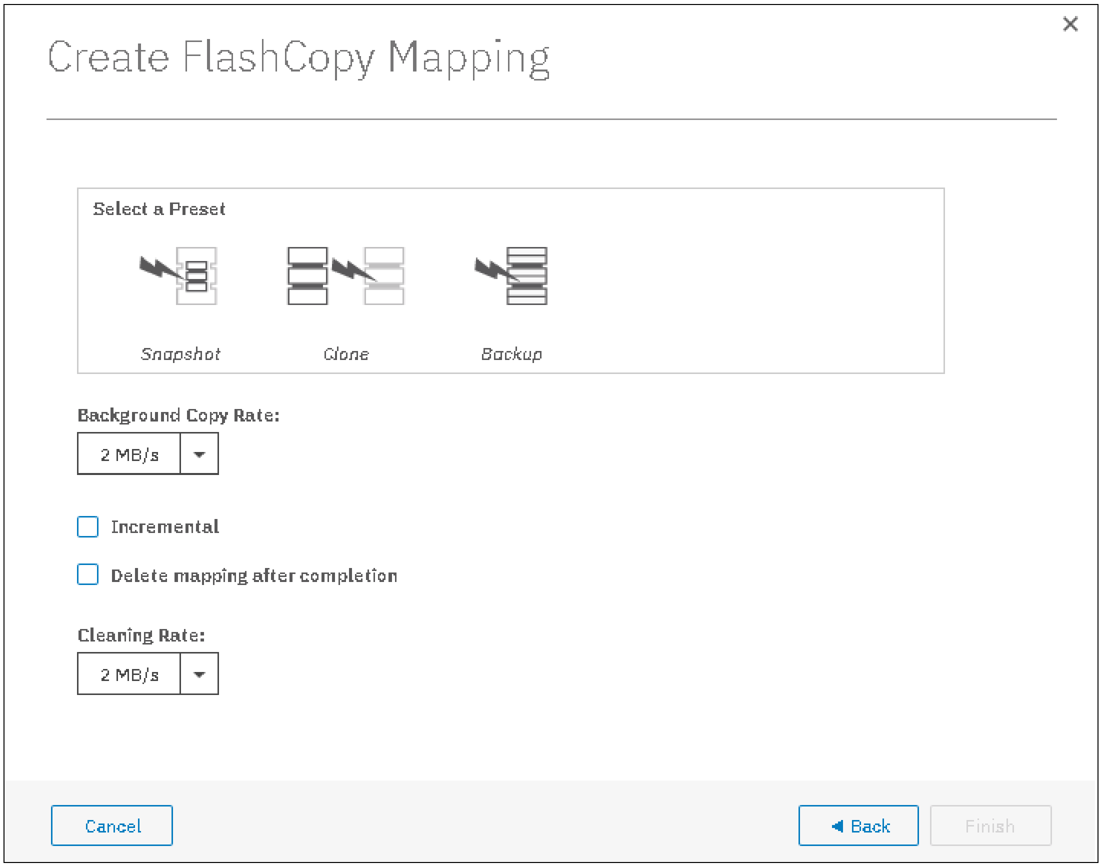


Figure 11-42 FlashCopy mapping preset selection

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:

- For example, 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.
- As another example, if you select the **Snapshot** preset but then change the **Background Copy Rate**, you have with a full copy of your source volume.

For more information about the Background Copy Rate and Cleaning Rate options, see Table 11-1 on page 448 or see Table 11-5 on page 457.

5. When your FlashCopy mapping setup is ready, click **Finish**.

11.2.9 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 11-43 on page 486. Also, depending on which window you are inside Copy Services, you can right-click the mappings and select **Show Related Volumes**.

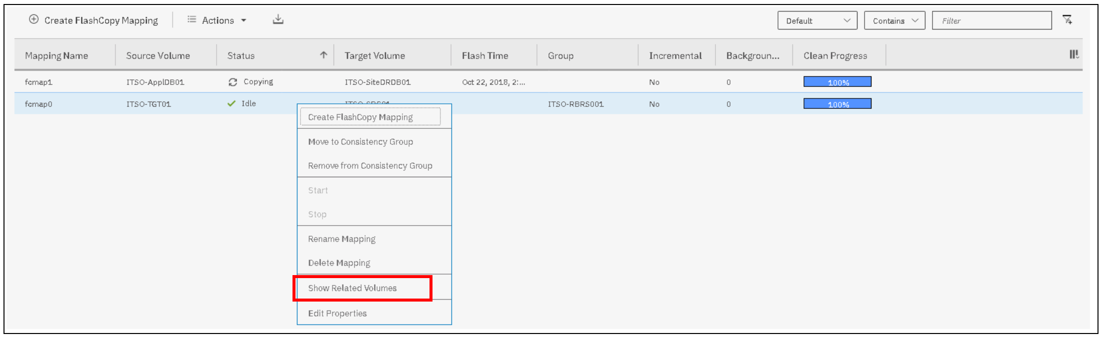


Figure 11-43 Showing related volumes for a mapping, a consistency group or another volume

3. In the related volumes window, you can see the related mapping for a volume, as shown in Figure 11-44. Click one of these volumes to see its properties.

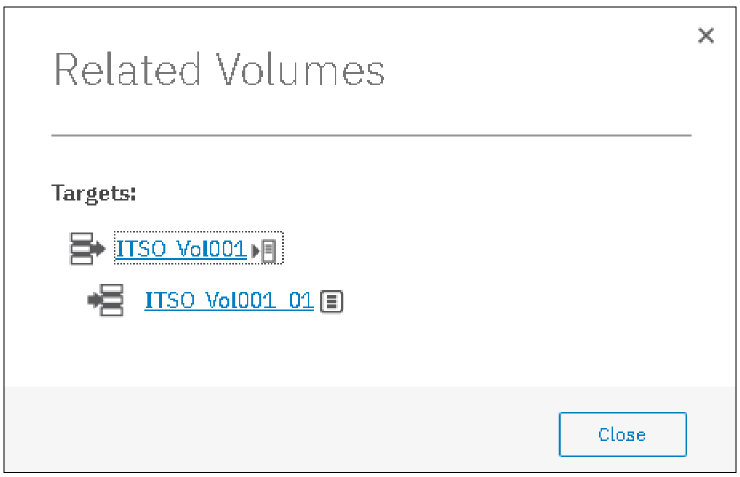


Figure 11-44 Showing related volumes list

11.2.10 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 11-45.

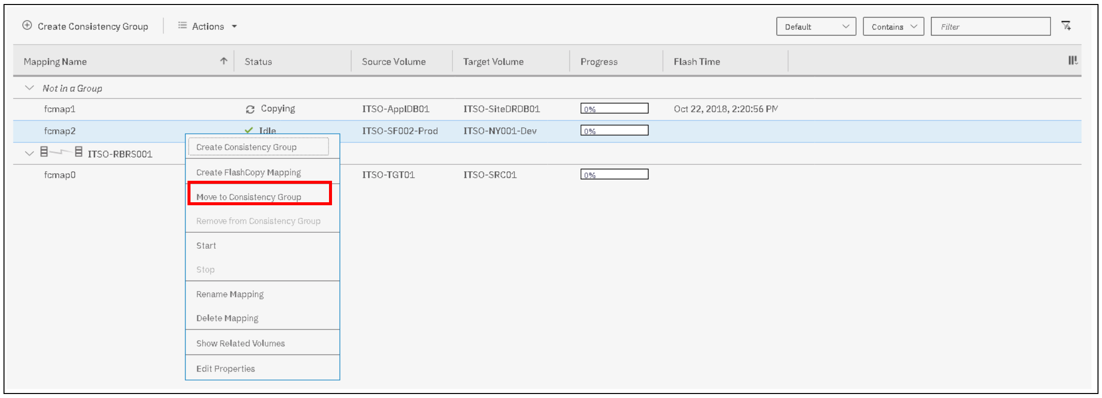


Figure 11-45 Moving a FlashCopy mapping to a consistency group

> **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 11-46.

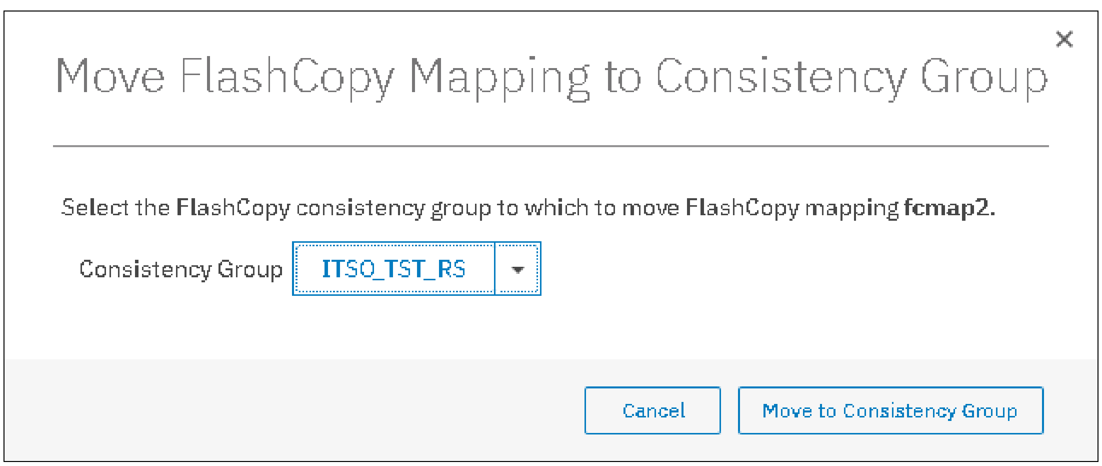


Figure 11-46 Selecting the consistency group where to move the FlashCopy mapping

4. Click **Move to Consistency Group** to confirm your changes.

11.2.11 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 11-47.

> **Note:** Only FlashCopy mappings that belong to a consistency group can be removed.

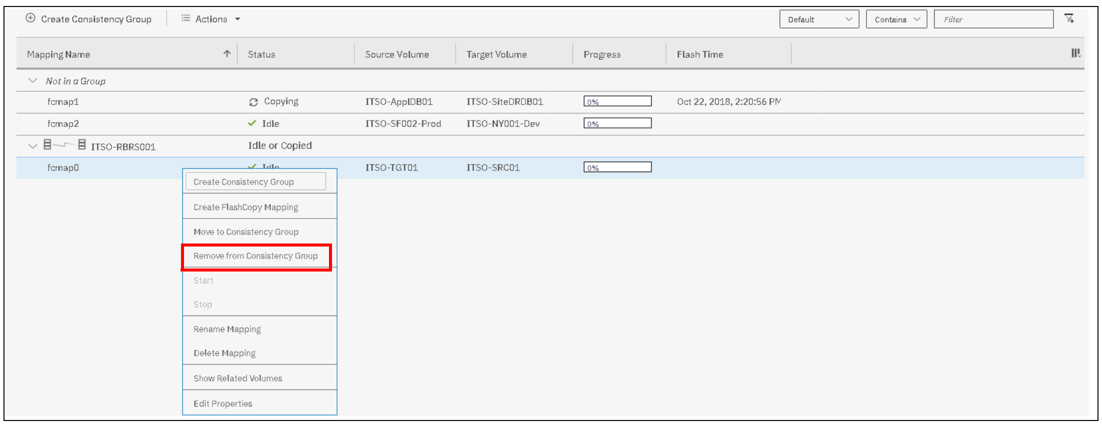


Figure 11-47 Removing FlashCopy mappings from a consistency group

3. In the Remove FlashCopy Mapping from Consistency Group window, click **Remove**, as shown in Figure 11-48.

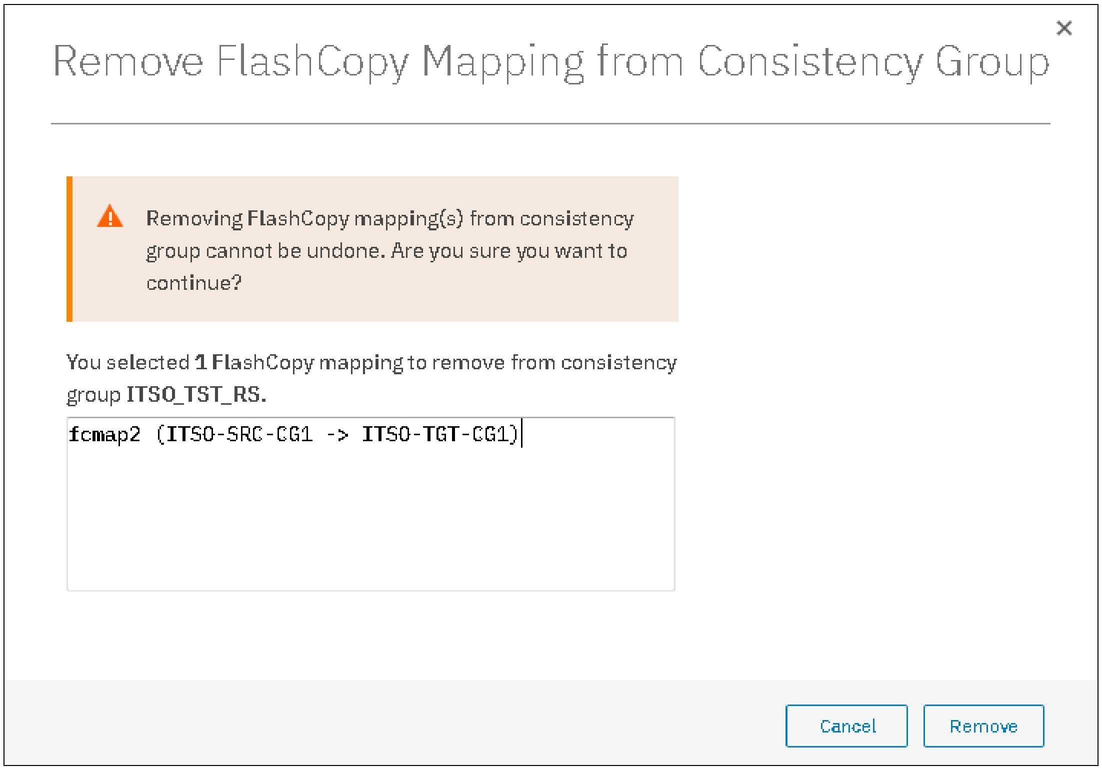


Figure 11-48 Confirm the selection of mappings to be removed

11.2.12 Modifying a FlashCopy mapping

To modify a FlashCopy mapping, complete the following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window.

   Right-click the FlashCopy mapping that you want to edit and select **Edit Properties**, as shown in Figure 11-49.

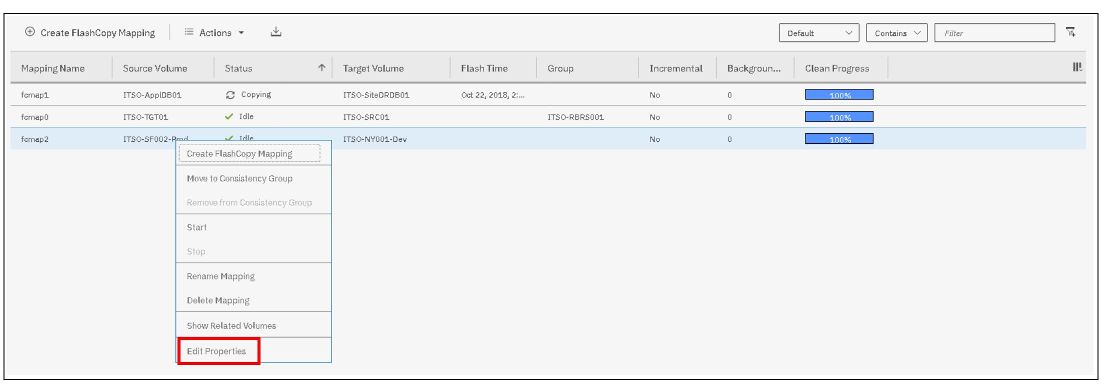


Figure 11-49 Editing a FlashCopy mapping properties

> **Note:** It is not possible to select multiple FlashCopy mappings to edit all of their properties at the same time.

2. 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 11-50.

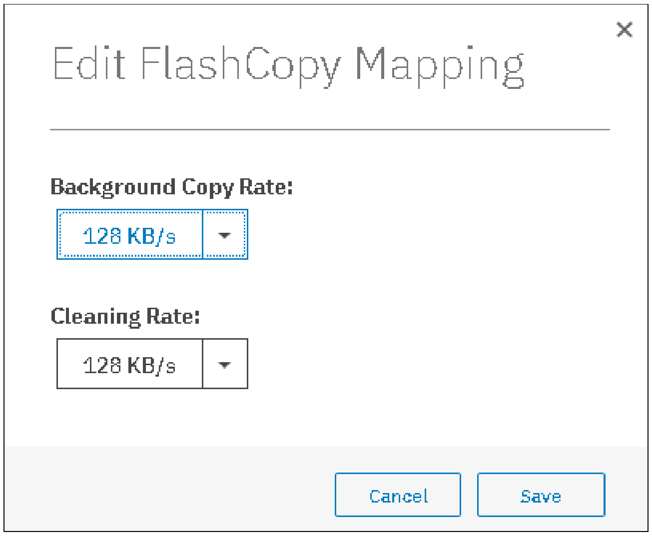


Figure 11-50 Editing copy and cleaning rates of a FlashCopy mapping

For more information about the “Background Copy Rate” and the “Cleaning Rate”, see Table 11-1 on page 448 or see Table 11-5 on page 457.

3. Click **Save** to confirm your changes.

11.2.13 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 11-51.

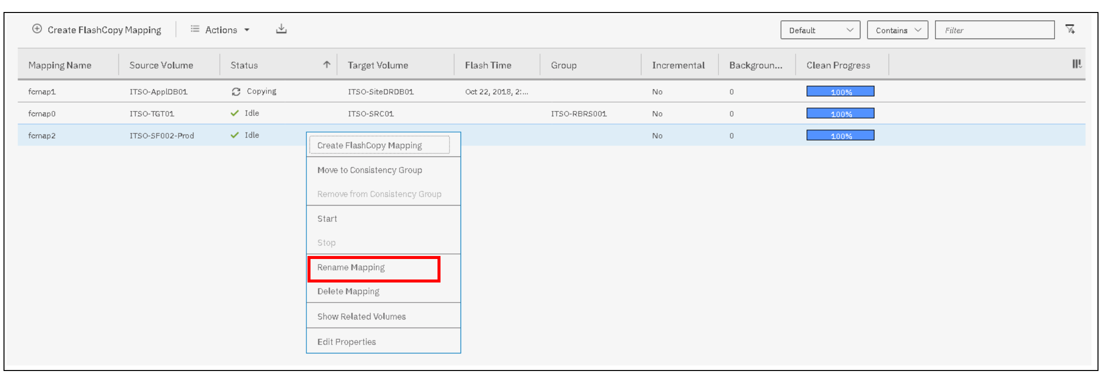


Figure 11-51 Renaming FlashCopy mappings

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 11-52.

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.

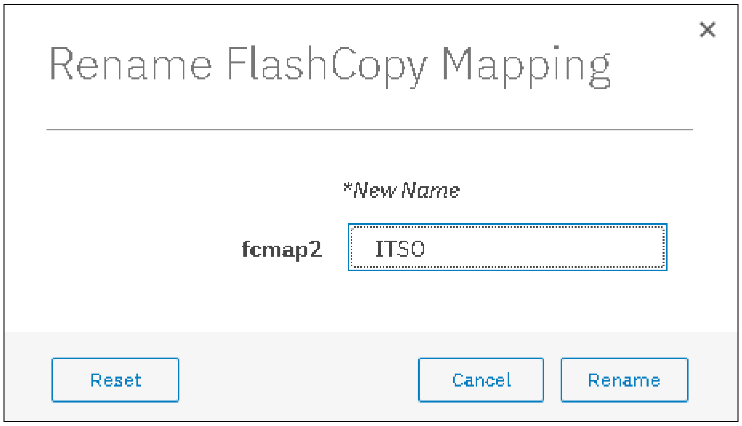


Figure 11-52 Renaming the selected FlashCopy mappings

Renaming a Consistency Group

To rename a Consistency Group, complete the following steps:

1. Open the Consistency Groups panel.
2. Right-click the consistency group you want to rename and select **Rename**, as shown in Figure 11-53.

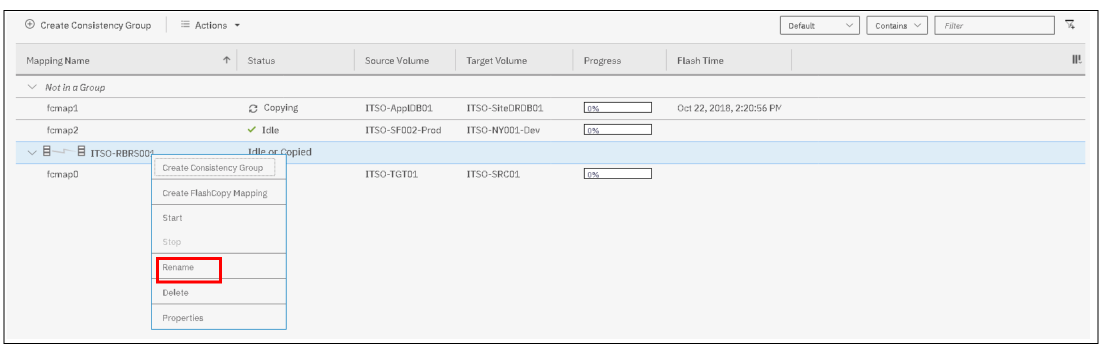


Figure 11-53 Renaming a consistency group

3. Enter the new name that you want to assign to the Consistency Group and click **Rename**, as shown in Figure 11-54.

> **Note:** It is not possible to select multiple consistency groups to edit their names all at the same time.

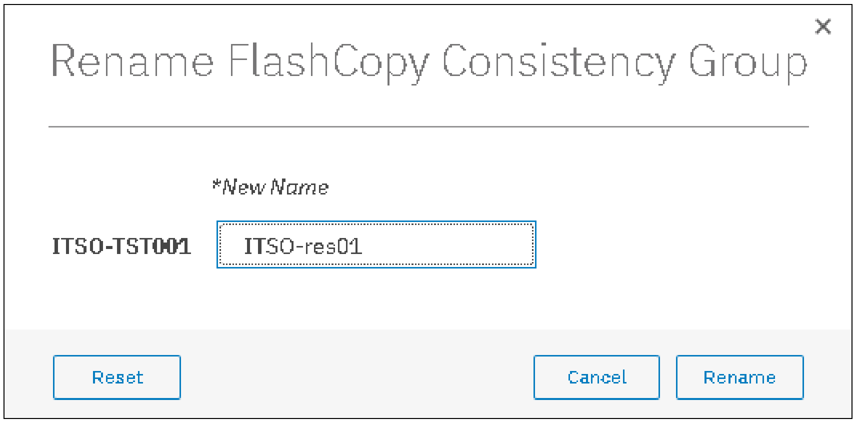


Figure 11-54 Renaming the selected consistency group

> **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.

11.2.14 Deleting FlashCopy mappings

To delete one or multiple FlashCopy mappings, complete them following steps:

1. Open the FlashCopy Consistency Groups, or FlashCopy Mappings window. Right-click the FlashCopy mappings that you want to delete and select **Delete Mapping**, as shown in Figure 11-55.

.

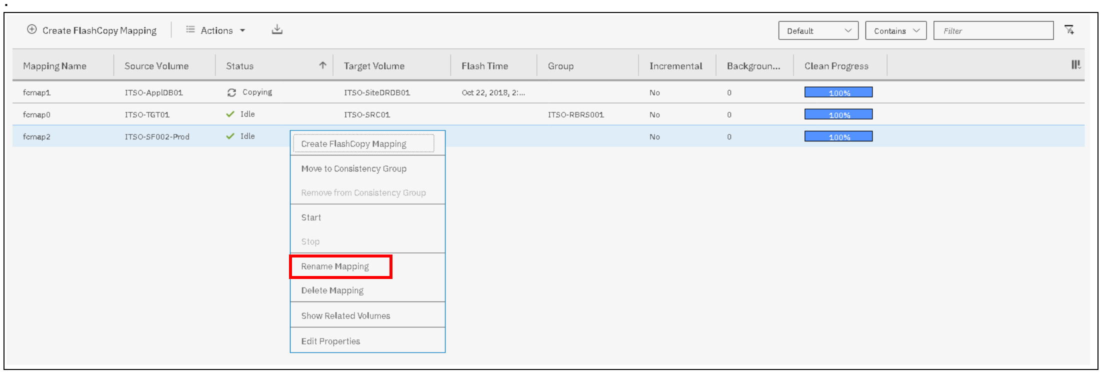


Figure 11-55 Deleting FlashCopy mappings

2. The Delete FlashCopy Mapping window opens, as shown in Figure 11-56. 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.

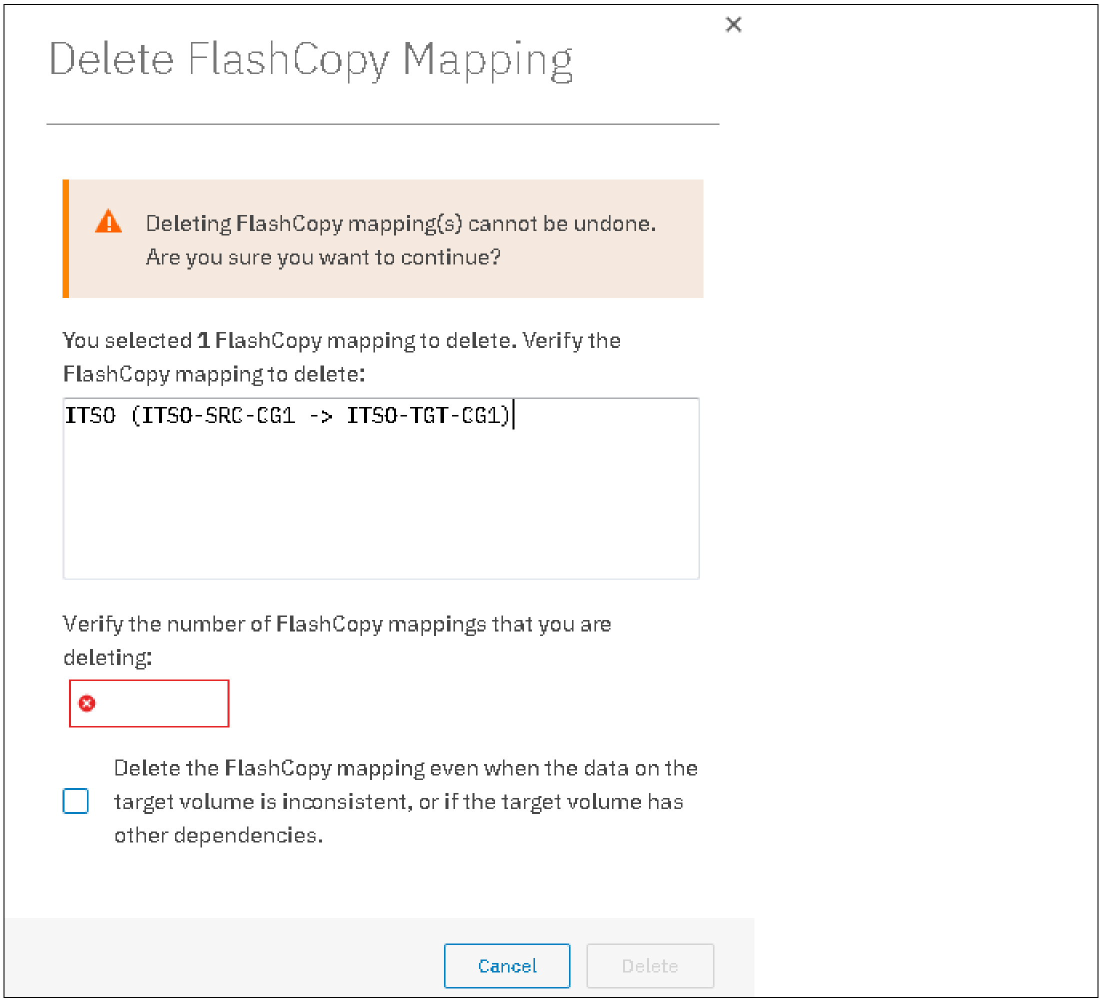


Figure 11-56 Confirming the selection of FlashCopy mappings to be deleted

3. If you still have target volumes that are inconsistent with the source volumes and you want to delete these FlashCopy mappings, click 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**.

11.2.15 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 11-57.

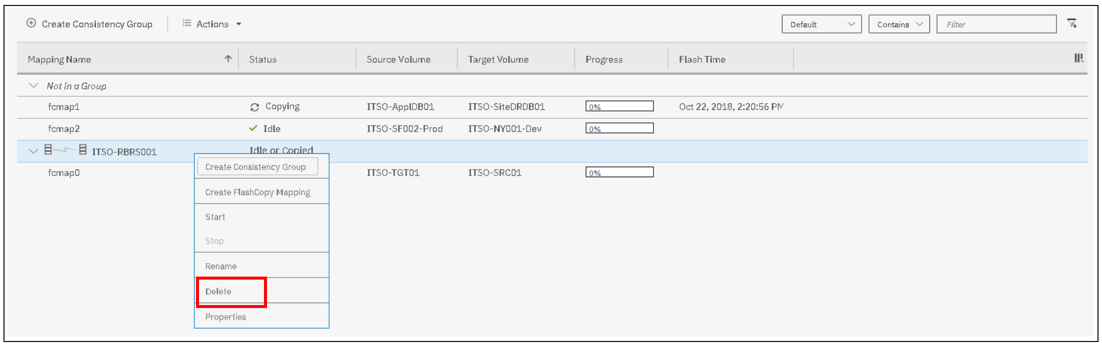


Figure 11-57 Deleting a consistency group

3. A warning message is displayed, as shown in Figure 11-58. Click **Yes**.

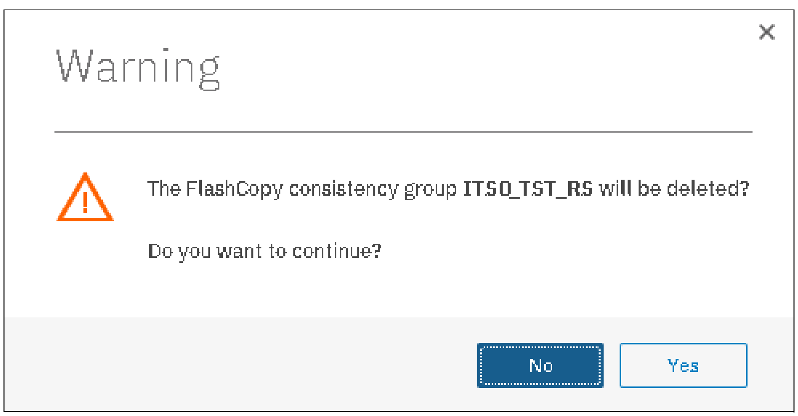


Figure 11-58 Confirming the consistency group deletion

11.2.16 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 only be started together by using the consistency group `start` command.

The use of the `start` command defines the “point-in-time”, which 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 11-59.

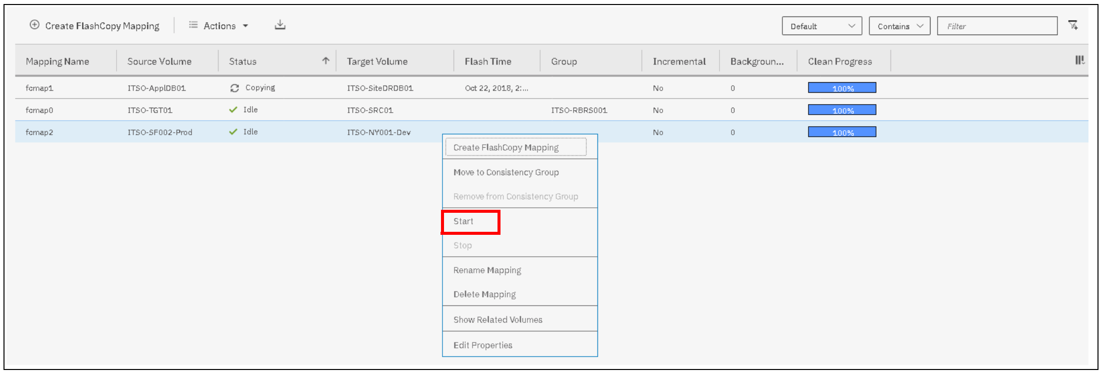


Figure 11-59 Starting FlashCopy mappings

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 11-60.

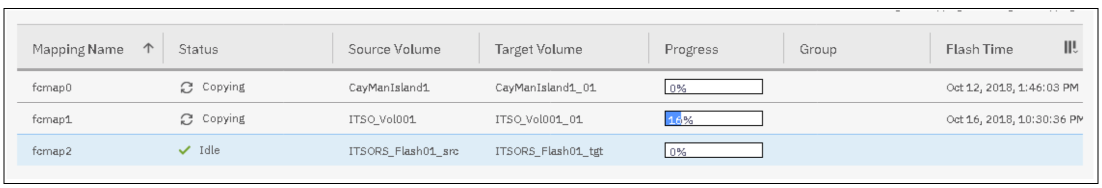


Figure 11-60 FlashCopy mappings status and progress examples

FlashCopy Snapshots are dependent 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 a long time to complete, depending on the size of the source volume and on the copyrate value. 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 11.1.10, “Starting FlashCopy mappings and Consistency Groups” on page 450.

11.2.17 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 only be stopped together 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 11-61.

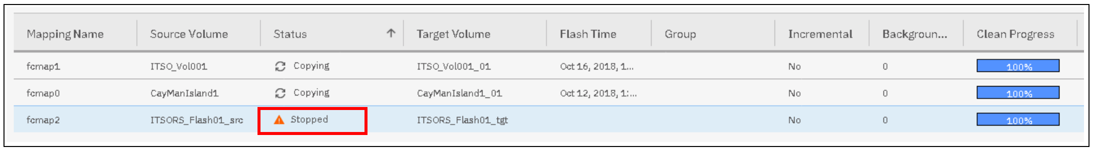


Figure 11-61   Showing target volumes state and FlashCopy mappings status

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 start and select **Stop**, as shown in Figure 11-62.

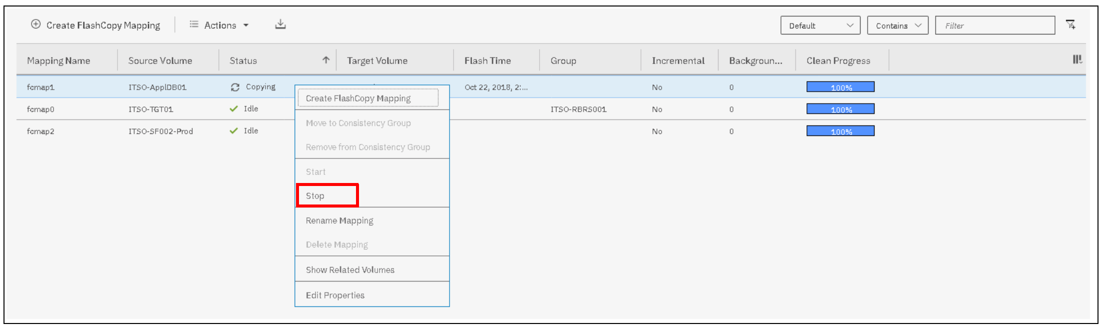


Figure 11-62   Stopping FlashCopy mappings

Note: FlashCopy mappings can be in a `stopping` state for a long 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 456.

11.2.18 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 you do not have enough bitmap space 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. One 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

Review Table 11-9 to calculate the memory requirements and confirm that your system is able to accommodate the total installation size.

Table 11-9 Memory allocation for FlashCopy services

Minimum allocated bitmap space	**Default allocated bitmap space**	**Maximum allocated bitmap space**	**Minimum[1] 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
[1]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, 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 consumes 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 Global Mirror, two internal FlashCopy mappings are created for each change volume.

You can modify the resource allocation for each I/O group of an IBM SAN Volume Controller system by opening the **Settings** → **System** window and clicking the **Resources** menu, as shown in Figure 11-63.

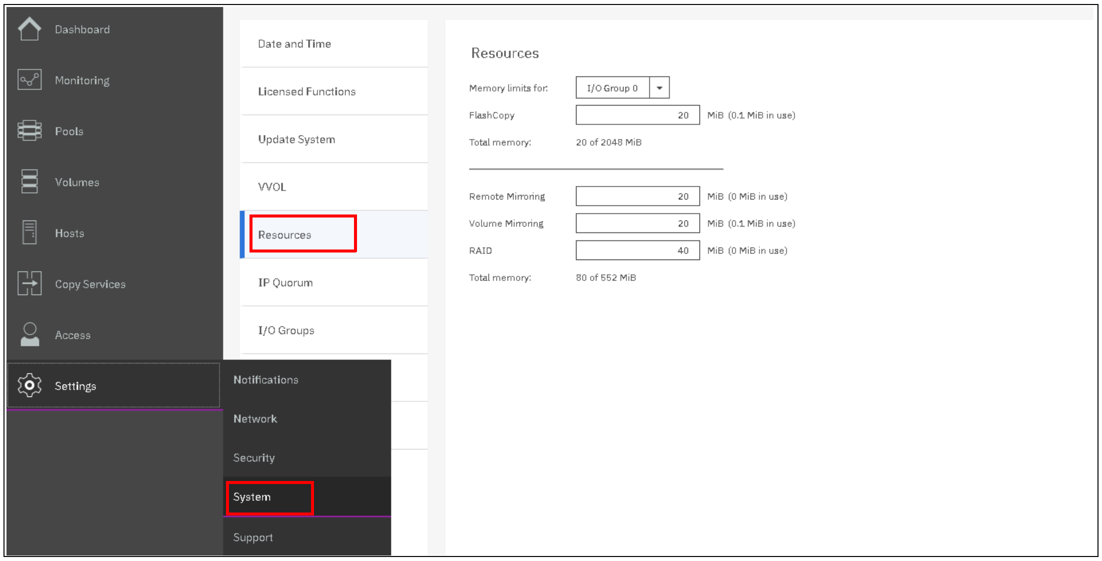


Figure 11-63 Modifying resources allocation per I/O group

11.3 Transparent Cloud Tiering

Introduced in V7.8, Transparent Cloud Tiering is a function of IBM Spectrum Virtualize that uses IBM FlashCopy mechanisms to produce a *point-in-time* snapshot of the data. Transparent Cloud Tiering 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. Transparent Cloud Tiering 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.

Transparent Cloud Tiering 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. Transparent Cloud Tiering operates below the host operating system and its cache. Therefore, the copy is not apparent to the host.

IBM Spectrum Virtualize has built-in software algorithms that allow the Transparent Cloud Tiering function to securely interact, for example with Information Dispersal Algorithms (IDA), which is essentially the interface to IBM Cloud Object Storage.

Object storage 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 a number of 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 REST application programming interface (API) to interact with an external cloud service provider 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 demo that is available at this web page.

11.3.1 Considerations for using Transparent Cloud Tiering

Transparent Cloud Tiering can help to address certain business needs. When considering whether to use Transparent Cloud Tiering, adopt a combination of business and technical views of the challenges and determine whether Transparent Cloud Tiering can solve both of those needs.

The use of Transparent Cloud Tiering 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 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 your amount of data that needs to be transmitted off 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, Transparent Cloud Tiering within the IBM Spectrum Virtualize can provide opportunities to manage the exponential data growth and to manipulate data at low cost.

Today, many Cloud Service Providers 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 related to rapid data growth, scalability, and manageability at attractive costs.

11.3.2 Transparent Cloud Tiering as backup solution and data migration

Transparent Cloud Tiering can also be used as backup and data migration solution. In certain conditions, it can be easily applied to eliminate the downtime that is associated with the needs to import and export data.

When Transparent Cloud Tiering is applied as your backup strategy, IBM Spectrum Virtualize uses the same FlashCopy functions to produce *point-in-time* snapshot of an entire volume or set of volumes.

To ensure the integrity of the snapshot, it might be necessary to flush the host operating system and application cache of any outstanding reads or writes before the snapshot is performed. Failing to flush the host operating system and application cache can produce inconsistent and useless data.

Many operating systems 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 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 cloud service providers or a private cloud, the IBM Spectrum Virtualize requires interaction with the correct architecture and specific properties. Conversely, cloud service providers offered attractive prices per 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.

11.3.3 Restore using Transparent Cloud Tiering

Transparent Cloud Tiering 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. Transparent Cloud Tiering 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: You should always consider the bandwidth characteristics and network capabilities when choosing to use Transparent Cloud Tiering.

The restore of individual files using Transparent Cloud Tiering is not possible. As mentioned, object storage is unlike a file or a block so 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.

11.3.4 Transparent Cloud Tiering restrictions

This section describes the following restrictions that must be considered before using Transparent Cloud Tiering:

- Because the object storage is normally accessed by using the HTTP protocol on top of a TPC/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 snapshot must be created so that a new Object Storage is constructed.
- Transparent Cloud Tiering cannot be applied to volumes that are part of traditional copy services, such as FlashCopy, Metro Mirror, Global Mirror, and HyperSwap.
- Volume containing two physical copies in two different storage pools cannot be part of Transparent Cloud Tiering.
- Cloud Tiering snapshots cannot be taken from a volume that is part of migration activity across storage pools.

- Because VVols are managed by a specific VMware application, these volumes are not candidates for Transparent Cloud Tiering.
- File system volumes, such as volumes provisioned by the IBM Storwize V7000 Unified platform, are not qualified for Transparent Cloud Tiering.

11.4 Implementing Transparent Cloud Tiering

This section describes the steps and requirements to implement Transparent Cloud Tiering using your IBM Spectrum Virtualize.

11.4.1 DNS 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** → **System** → **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 11-64.

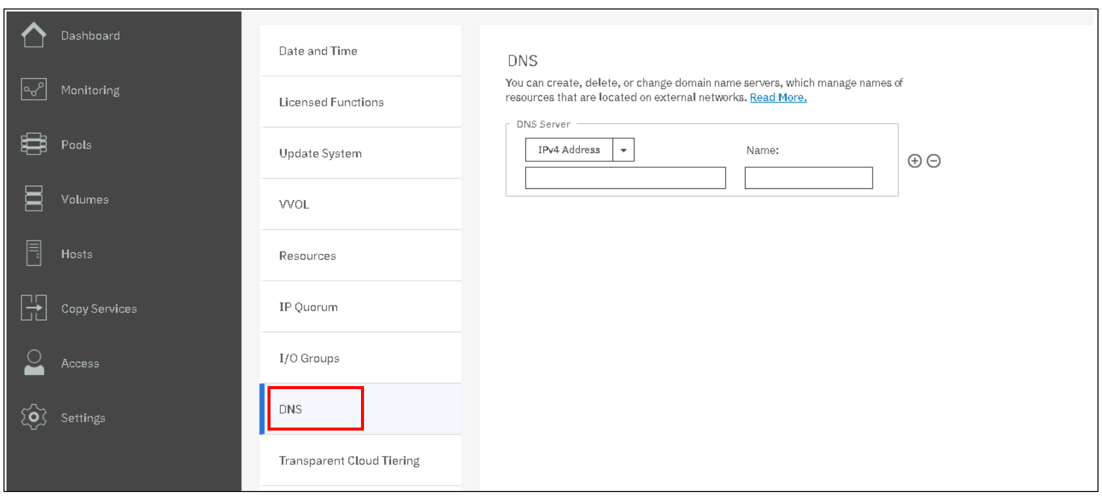


Figure 11-64 DNS settings

11.4.2 Enabling Transparent Cloud Tiering

After you complete the DNS settings, you can enable the Transparent Cloud Tiering 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 click **Enable Cloud Connection**, as shown in Figure 11-65. The Transparent Cloud Tiering wizard starts and shows the welcome warning.

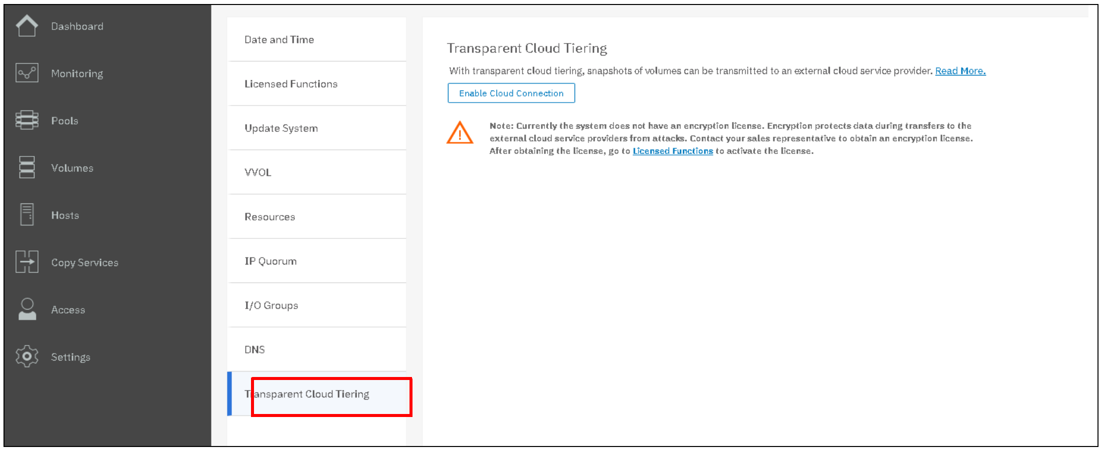


Figure 11-65 Enabling Cloud Tiering

> **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 publication, encryption is *not* enabled on our system.

2. Click **Next** to continue. You must select one of three cloud service providers:

   – IBM Cloud
   – OpenStack Swift
   – Amazon S3

Figure 11-66 shows the available options.

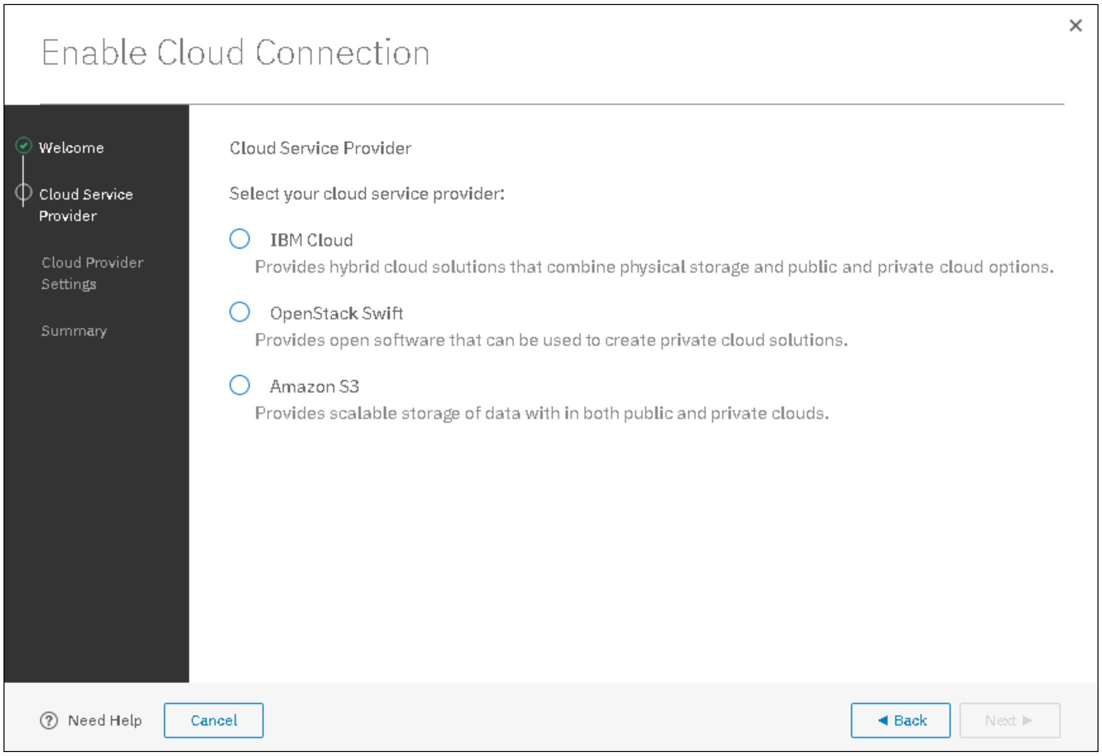


Figure 11-66 Selecting Cloud service provider

3. 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 cloud service provider. An example of an empty form for an IBM Cloud connection is shown in Figure 11-67 on page 504.

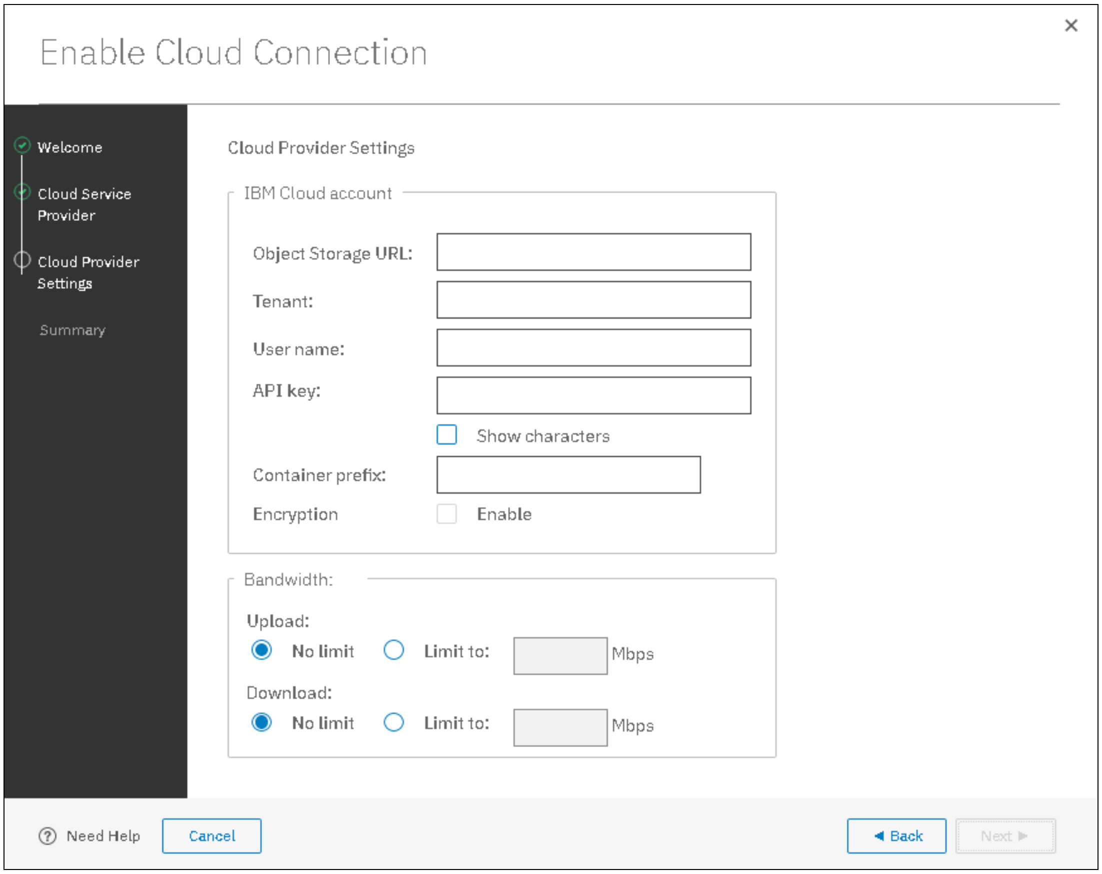


Figure 11-67 Entering Cloud Service provider information

4. Review your settings and click **Finish**, as shown in Figure 11-68.

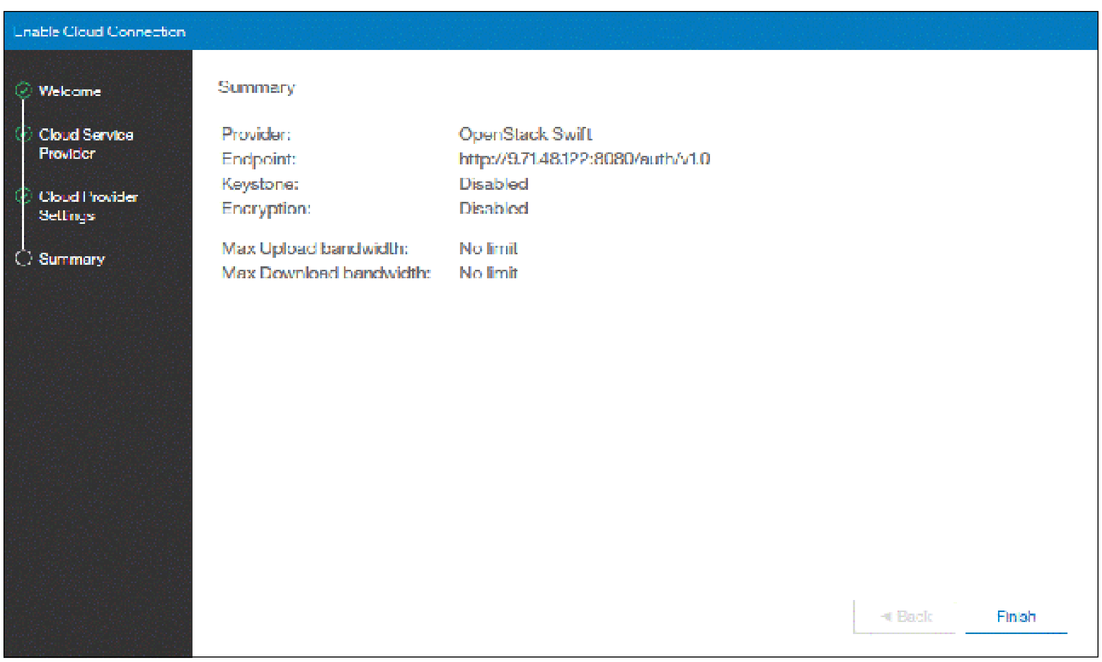


Figure 11-68 Cloud Connection summary

5. 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 functionality.

   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 11-69.

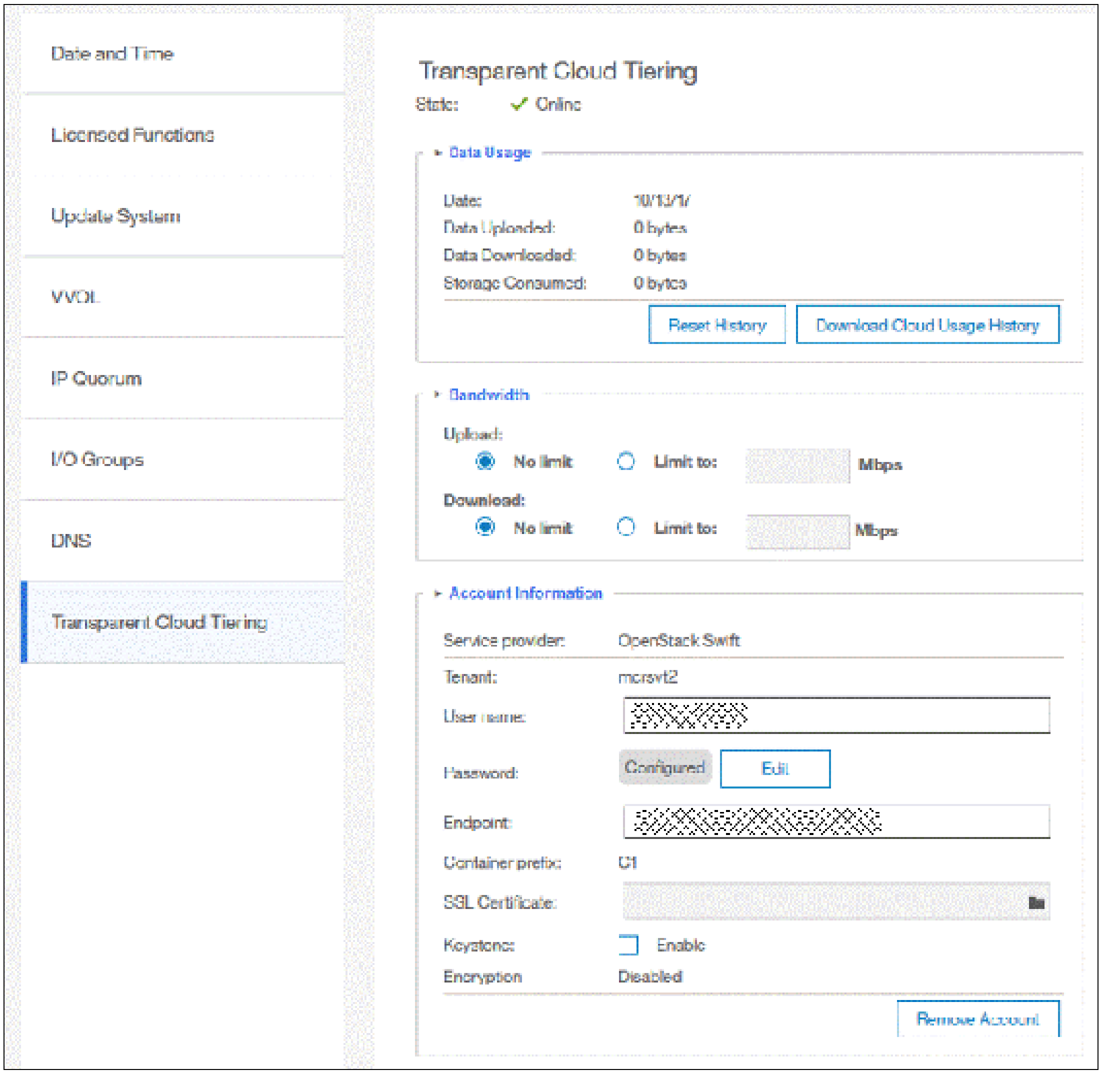


Figure 11-69 Enabled Transparent Cloud Tiering window

11.4.3 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 Transparent Cloud Tiering. As described in 11.3.4, “Transparent Cloud Tiering restrictions” on page 500, 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 only work 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 11-70.

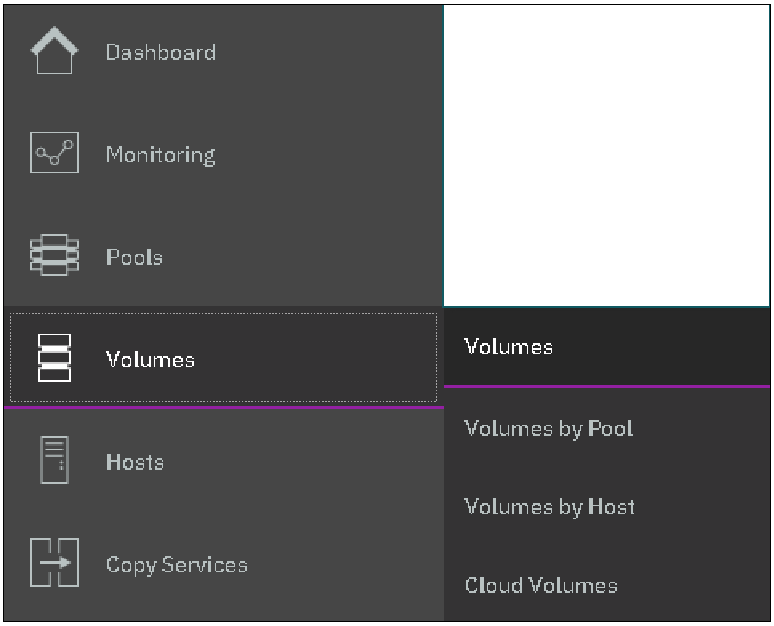


Figure 11-70 Cloud volumes menu

After Cloud Volumes is selected, 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 11-71.

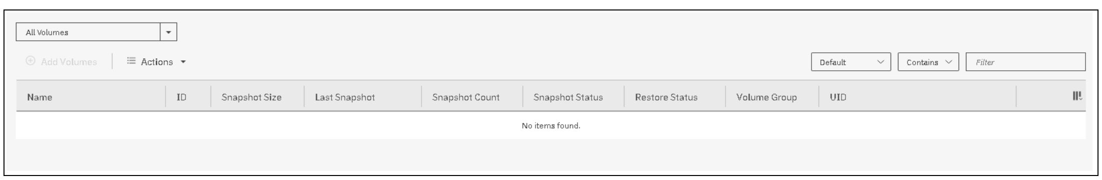


Figure 11-71 Cloud volumes window

2. Click **Add Volumes** to enable cloud-snapshot on volumes. A new window opens, as shown in Figure 11-72. Select the volumes that you want to enable Cloud Tiering for and click **Next**.

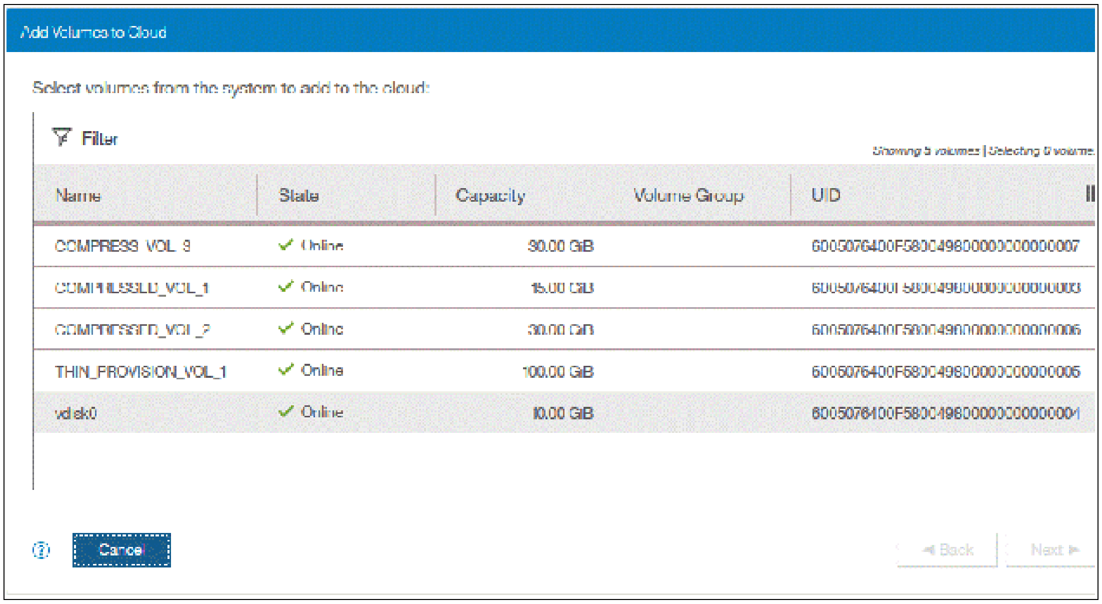


Figure 11-72 Adding volumes to Cloud Tiering

3. 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 has already been created, then 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 11-73. You can also select the second option, even if another full copy of the volume exists.

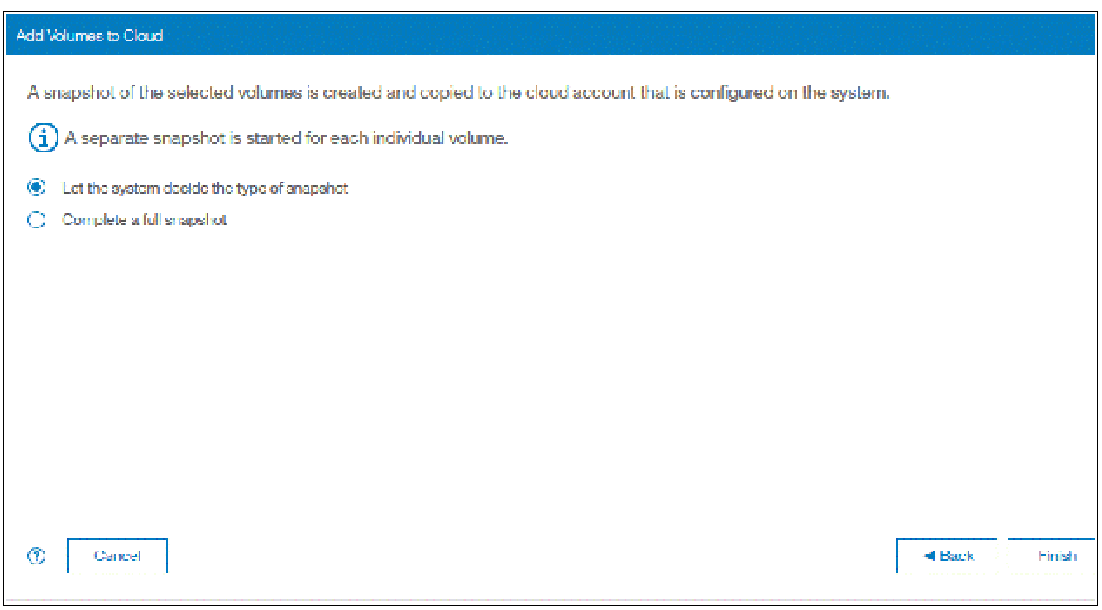


Figure 11-73 Selecting if a full copy is made or if the system decides

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
- The snapshot size
- The date and time that the last snapshot was created
- The number of snapshots that are taken for every volume
- The snapshot status
- The restore status
- The volume group for a set of volumes
- The volume UID

Figure 11-74 shows an example of a Cloud Volumes list.

Figure 11-74 Cloud Volumes list example

The Actions menu in the Cloud Volumes window allows you to create and manage snapshots. Also, you can use the menu to cancel, disable, and restore snapshots to volumes as shown in Figure 11-75.

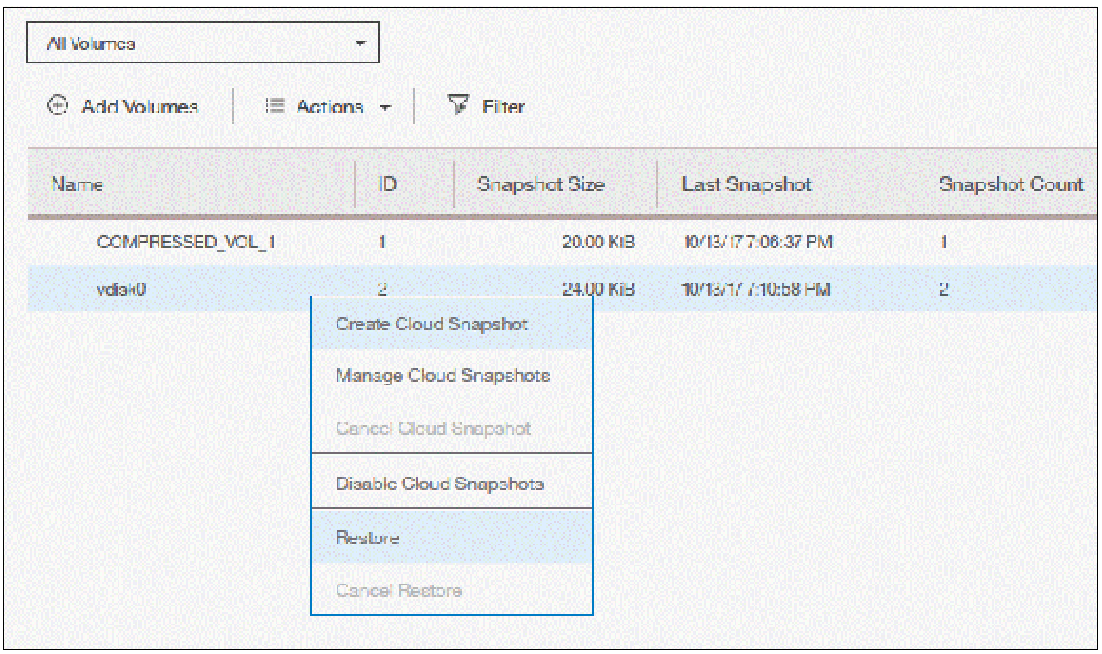


Figure 11-75 Available actions in Cloud Volumes window

11.4.4 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 deletes one or multiple versions. The list of point-in-time copies appear and provide more information about their status, type, and snapshot date, as shown in Figure 11-76 on page 509.

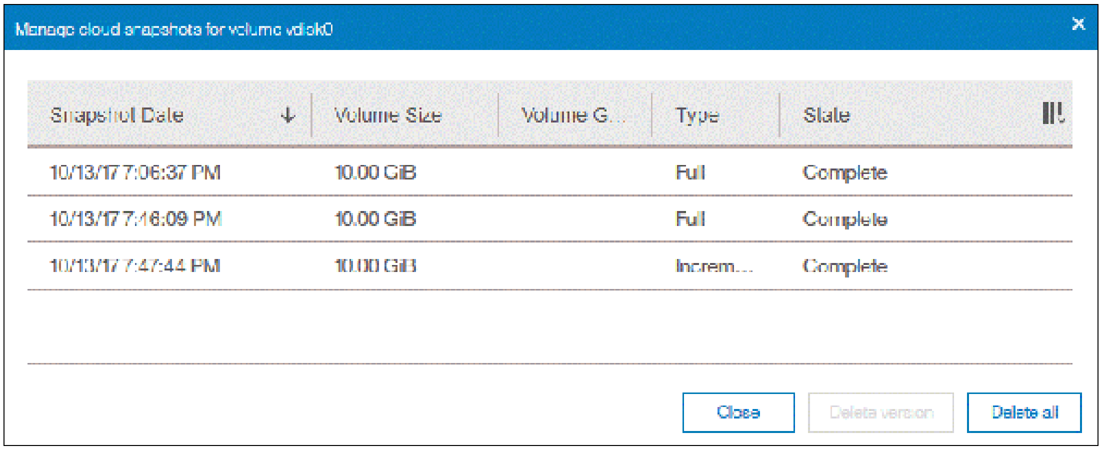


Figure 11-76 Deleting versions of a volume's snapshots

From this window, an administrator can delete old snapshots (old point-in-time 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.

11.4.5 Restoring cloud snapshots

This option allows IBM Spectrum Virtualize to restore snapshots from the cloud to the selected volumes or to create new 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 11-77.

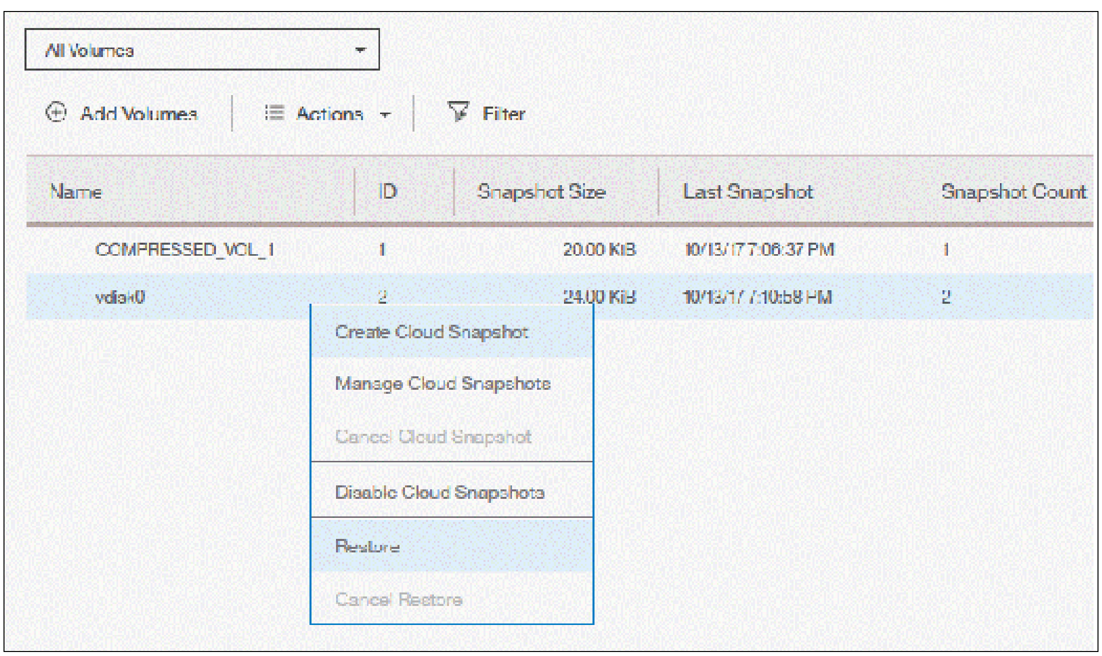


Figure 11-77 Selecting a volume to restore a snapshot from

3. A list of available snapshots is displayed. The snapshots date (point-in-time), their type (full or incremental), their state, and their size are shown, as in Figure 11-78. Select the version that you want to restore and click **Next**.

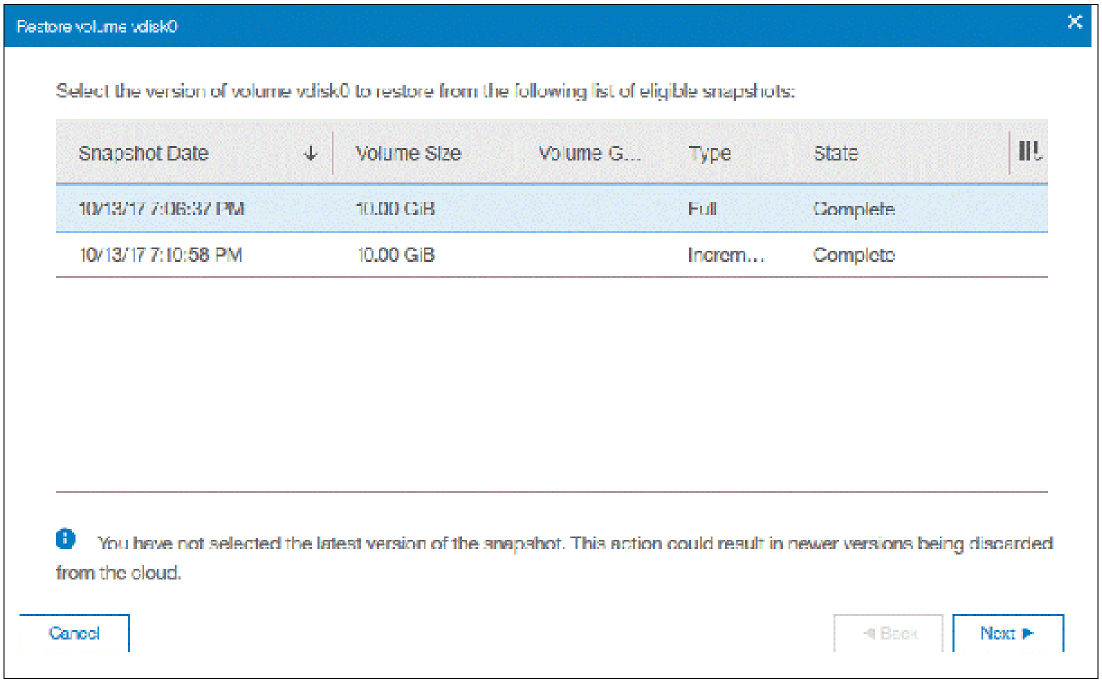


Figure 11-78 Selecting a snapshot version to restore

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 11-79. Make a selection and click **Next**.

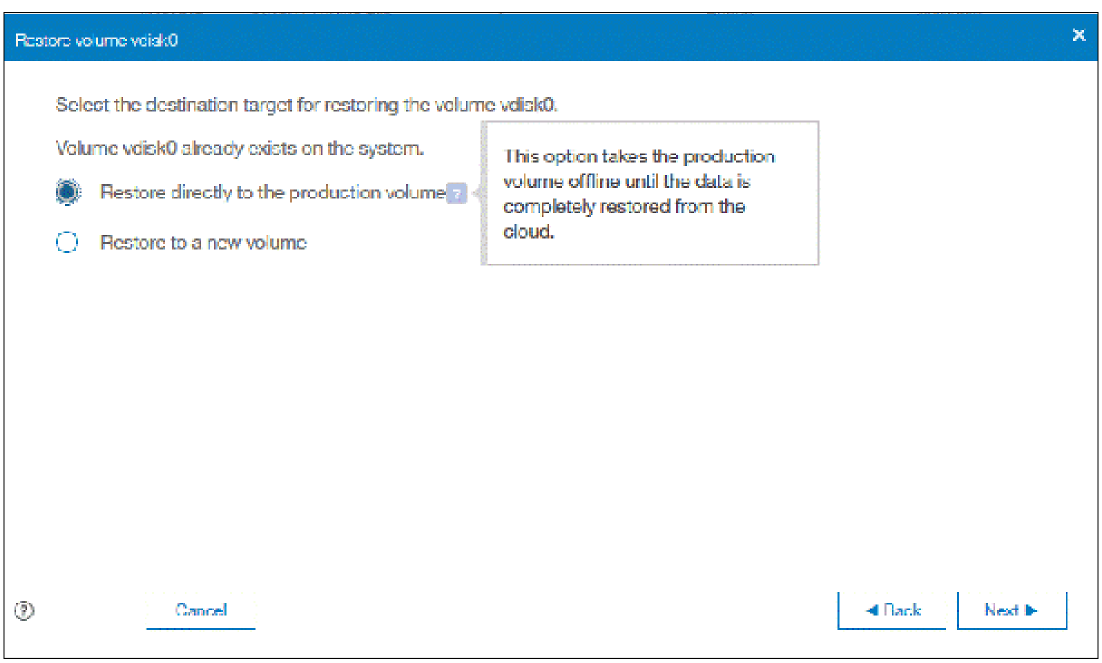


Figure 11-79 Restoring a snapshot on an existing volume or on a new volume

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 point-in-time 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 new volume to be created with the snapshot data, as shown in Figure 11-80:

   – Name
   – Storage Pool
   – Capacity Savings (None, Compressed or Thin-provisioned)
   – I/O group

   You are not prompted to enter the volume size because the new volume’s size is identical to the snapshot copy volume.

   Enter the settings for the new volume and click **Next**.

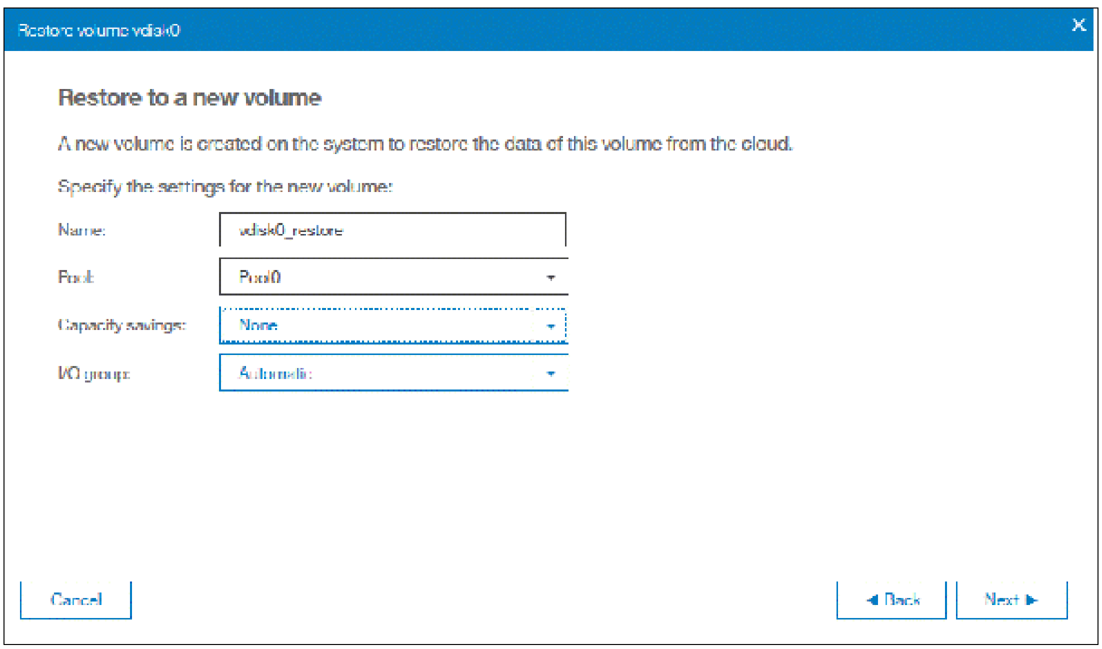


Figure 11-80 Restoring a snapshot to a new volume

6. A Summary window is displayed so you can double-check the restoration settings, as shown in Figure 11-81 on page 512. 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.

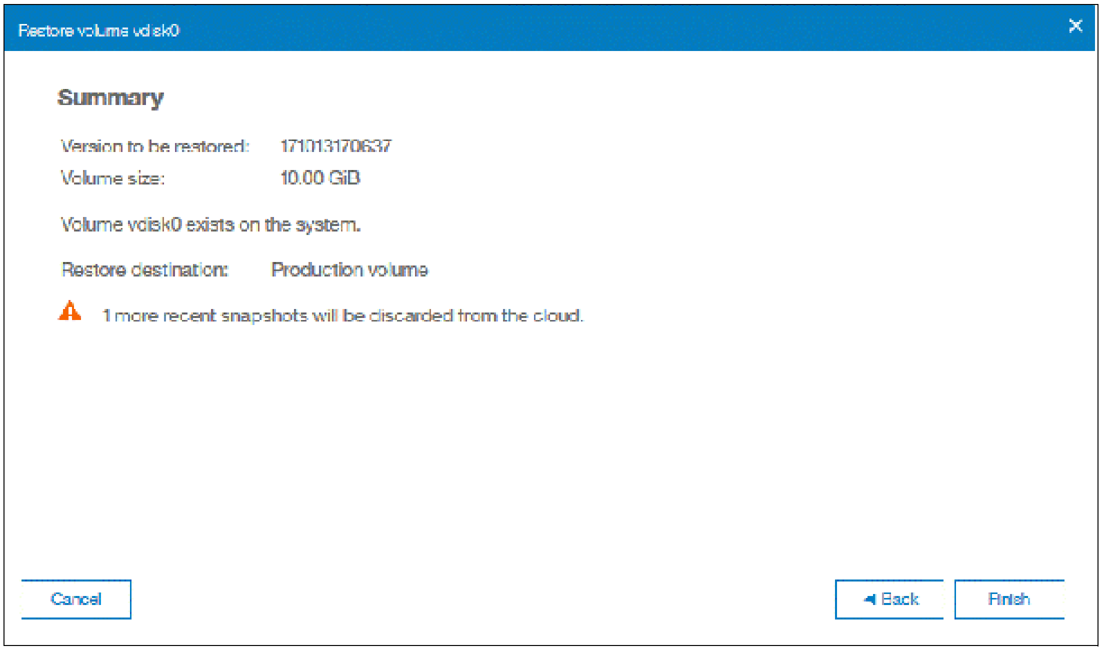


Figure 11-81 Restoring a snapshot summary

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 totally 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.

11.5 Volume mirroring and migration options

Volume mirroring is a simple 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 functionality, 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 time-out value is `Latency` or `Redundancy`.

The `Latency` tuning option, which is set with **`chvdisk -mirrorwritepriority latency`**, 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. Use 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 are using the Storwize V7000 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 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, you must take a brief outage to install the host drivers for your Storwize V7000 if you do not already have them installed.

With Volume mirroring, you can move data to different MDisks within the same storage pool or move data between different storage pools. Using 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 end up having one volume 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 V7.3 and the introduction of the new cache architecture, mirrored volume performance was significantly improved. Now, 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 now both copies use its independent cache and performs its own read prefetch. Destaging of the cache can now be done independently for each copy, so one copy does not affect performance of a second copy.

Also, because the IBM Storwize 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 11, “Advanced Copy Services” on page 435.

11.6 Remote Copy services

This section describes the Remote Copy services, which are a synchronous remote copy called Metro Mirror (MM), asynchronous remote copy called Global Mirror (GM), and Global Mirror with Change Volumes. Remote Copy in an IBM Spectrum Virtualize system is similar to Remote Copy 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 remote copy is enabled in your network (regardless of the disk subsystems that are used) if those disk subsystems are supported by the system.

The general application of remote copy 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 Storwize V7000 or IBM SAN Volume Controller 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.

11.6.1 IBM Storwize family and SAN Volume Controller system layers

An IBM Storwize family system can be in one of the two layers: the *replication* layer or the *storage* layer. The system layer affects how the system interacts with IBM SAN Volume Controller and other external Storwize family systems. The IBM SAN Volume Controller is always set to replication layer. This parameter cannot be changed.

In the storage layer, a Storwize family 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.
- The system cannot use a storage-layer system as external storage.

In the replication layer, an SAN Volume Controller or a Storwize system has the following characteristics and requirements:

- The system can perform MM and GM replication with other replication-layer systems.
- The system cannot provide external storage for a replication-layer system.
- The system can use a storage-layer system as external storage.

A Storwize family system is in the storage layer by default, but the layer can be changed. For example, you might want to change a Storwize V7000 to a replication layer if you want to virtualize other Storwize systems.

> **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 family system.
> - No system partnerships can be defined.
> - No Storwize family system can be visible on the SAN fabric.
>
> The layer can be changed during normal host I/O.

In your IBM Storwize system, use the **lssystem** command to check the current system layer, as shown in Example 11-2.

Example 11-2 Output from the lssystem command showing the system layer

```
IBM_Storwize:ITSO-V7k:superuser>lssystem
id 0000010029C056C2
name ITSO-V7k
location local
...
layer storage
rc_buffer_size 48
...
IBM_Storwize:ITSO-V7k:superuser>
```

> **Note:** Consider the following rules for creating remote partnerships between the IBM Storwize Family systems and IBM SAN Volume Controller:
>
> - An IBM SAN Volume Controller is always in the replication layer.
> - By default, the IBM Storwize 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 V6.4, an Storwize V7000 or IBM SAN Volume Controller system in the replication layer can virtualize an IBM Storwize system in the storage layer.

11.6.2 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 DR solutions.

> **Note:** For more information about restrictions and limitations of native IP replication, see 11.8.2, “IP partnership limitations” on page 550.

Figure 11-82 shows an example of a multiple system mirroring configuration.

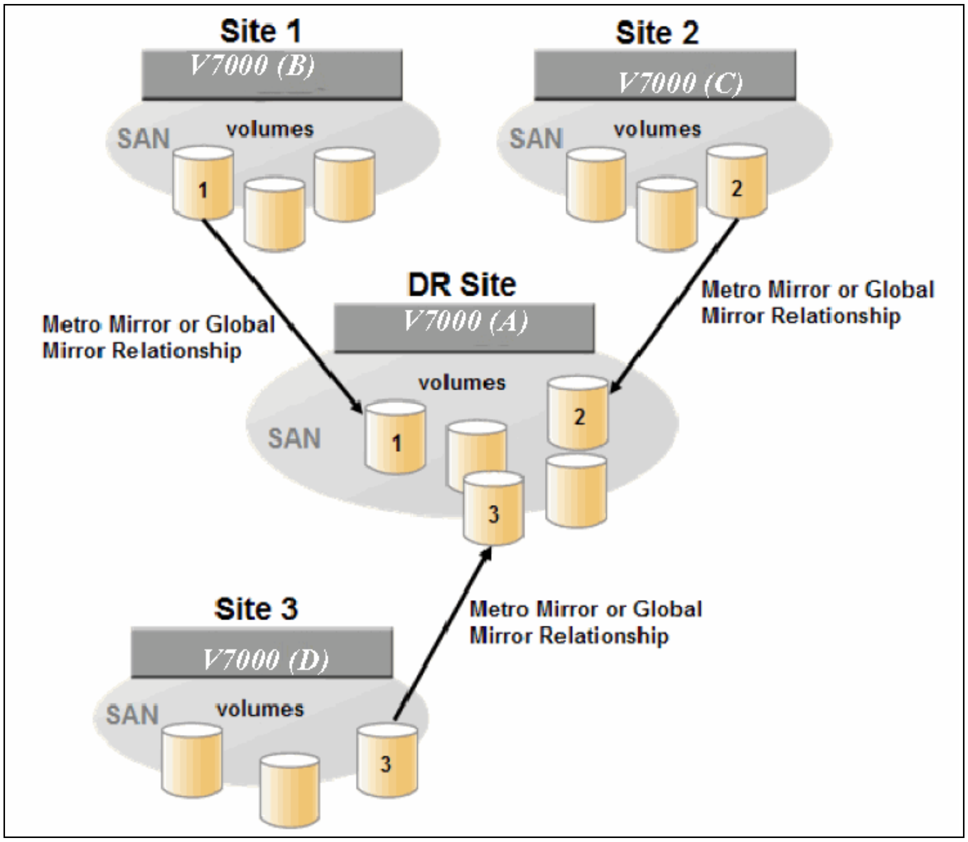


Figure 11-82 Multiple system mirroring configuration example

Supported multiple system mirroring topologies

Multiple system mirroring supports various partnership topologies, as shown in the example in Figure 11-83. This example is a star topology (A → B, A → C, and A → D).

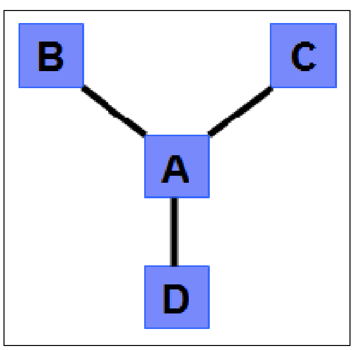


Figure 11-83 Star topology

Figure 11-83 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 → C relationship (or the B → C relationship).
4. Synchronize to system C, and ensure that A → C is established:
   – A → B, A → C, A → D, B → C, B → D, and C → D
   – A → B, A → C, and B → C

Figure 11-84 shows an example of a triangle topology (A → B, A → C, and B → C).

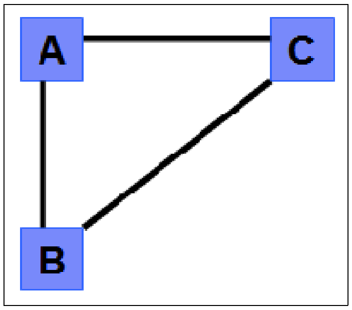


Figure 11-84 Triangle topology

Figure 11-85 shows an example of a fully connected topology (A → B, A → C, A → D, B → D, and C → D).

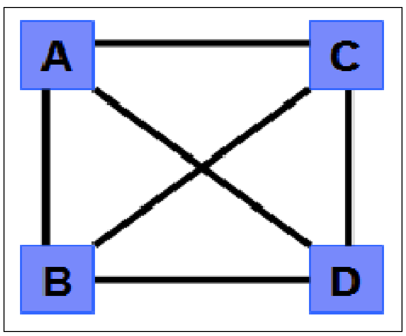


Figure 11-85 Fully connected topology

Figure 11-85 is 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 → B, A → C, and B → C.

Figure 11-86 shows a daisy-chain topology.

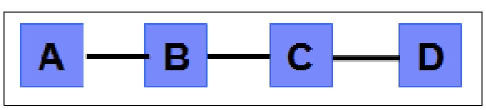


Figure 11-86 Daisy-chain topology

Although systems can have up to three partnerships, volumes can be part of only one remote copy relationship, for example A → B.

System partnership intermix: All of the preceding topologies are valid for the intermix of the IBM SAN Volume Controller with the Storwize V7000 if the Storwize V7000 is set to the replication layer and running IBM Spectrum Virtualize code 6.3.0 or later.

11.6.3 Importance of write ordering

Many applications that use block storage have a requirement 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 large numbers of 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 Metro Mirror and Global Mirror implementation operates in a manner that is designed to always keep a consistent image at the secondary site. The Global Mirror implementation uses complex algorithms that identify sets of data and number those sets of data in sequence. The data is then applied at the secondary site in the defined sequence.

Operating in this manner ensures that if the relationship is in a `Consistent_Synchronized` state, the Global Mirror target data is at least crash consistent, and supports quick recovery through application crash recovery facilities.

For more information about dependent writes, see 11.1.13, "FlashCopy and image mode Volumes" on page 459.

Remote Copy Consistency Groups

A Remote Copy 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 a Remote Copy 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 a Remote Copy 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 11-87 shows the concept of Remote Copy Consistency Groups.

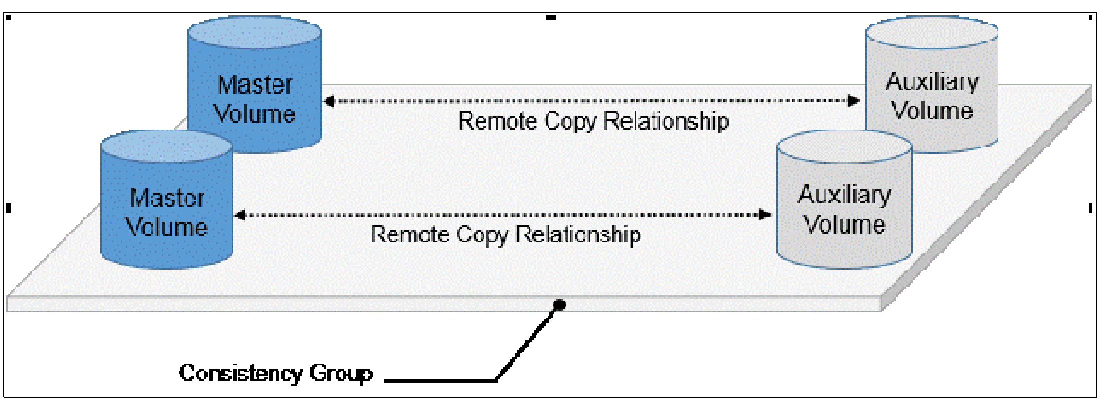


Figure 11-87 Remote Copy Consistency Group

Certain uses of MM/GM require the manipulation of more than one relationship. Remote Copy 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 Metro Mirror or only Global Mirror.

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 undesired 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 model. All of the relationships in a non-empty consistency group have the same state as the consistency group.

11.6.4 Remote copy intercluster communication

In the traditional Fibre Channel, 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 11.8.6, “States of IP partnership” on page 554.

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 Simple Network Management Protocol (SNMP) traps to the enterprise monitoring system to alert on events that indicate an interruption in internode communication occurred.

Intercluster links

All Storwize V7000 nodes maintain a database of other devices that are visible on the fabric. This database is updated as devices appear and disappear.

Devices that advertise themselves as Storwize V7000 nodes or IBM SAN Volume Controller are categorized according to the system to which they belong. Nodes that belong to the same system establish communication channels between themselves and begin to 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 a remote copy 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: Use **`chsystem`** with **`-partnerfcportmask`** to dedicate several FC ports only to system-to-system traffic to ensure that remote copy is not affected by other traffic, such as host-to-node traffic or node-to-node traffic within the same system.

11.6.5 Metro Mirror overview

Metro Mirror establishes a synchronous relationship between two volumes of equal size. The volumes in a Metro Mirror relationship are referred to as the master (primary) volume and the auxiliary (secondary) volume. Traditional FC Metro Mirror 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 Metro Mirror function always maintains a fully synchronized copy of the source data at the target site.

Metro Mirror 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 Metro Mirror 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 Metro Mirror.

Intracluster Metro Mirror

Intracluster Metro Mirror 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, there must be sufficient bitmap space within the I/O Group for both sets of volumes and licensing on the system.

Important: Performing Metro Mirror across I/O Groups within a system is not supported.

Intercluster Metro Mirror

Intercluster Metro Mirror 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 Metro Mirror partnership.

By using standard single-mode connections, the supported distance between two systems in a Metro Mirror 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 Metro Mirror purposes, the inter-switch link (ISL) hop count between a local node and a remote node cannot exceed seven.

11.6.6 Synchronous remote copy

Metro Mirror is a fully synchronous remote copy technique that ensures that writes are committed at both 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, Metro Mirror 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 11-88 on page 522 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.

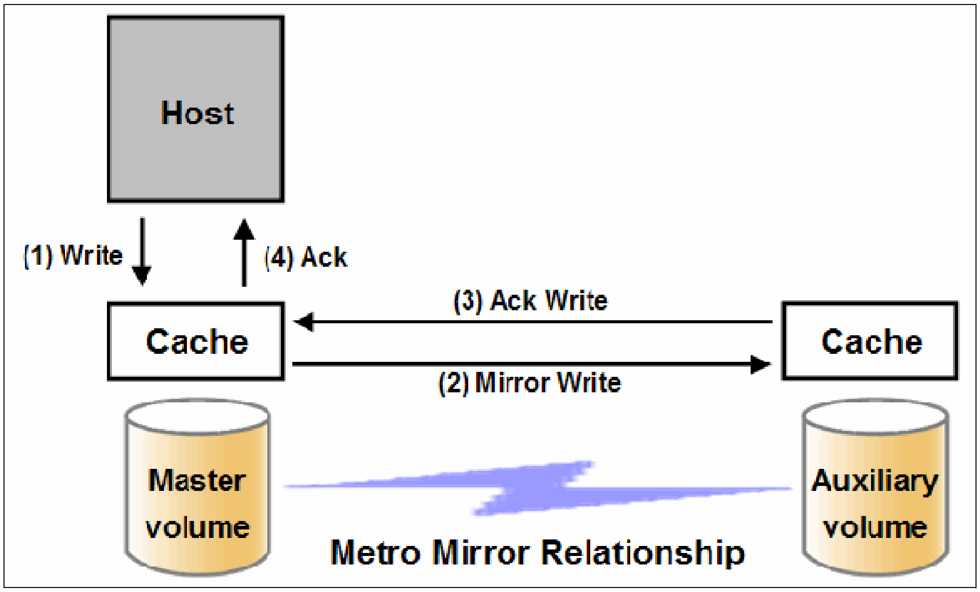


Figure 11-88 Write on volume in Metro Mirror relationship

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 Fibre Channel Metro Mirror has distance limitations that are based on your performance requirements. IBM Spectrum Virtualize does not support more than 300 km (186.4 miles).

11.6.7 Metro Mirror features

The IBM Spectrum Virtualize Metro Mirror function supports the following features:

- Synchronous remote copy of volumes that are dispersed over metropolitan distances.
- The Metro Mirror relationships between volume pairs, with each volume in a pair that is managed by a Storwize V7000 system or IBM SAN Volume Controller system (requires V6.3.0 or later).
- Supports intracluster Metro Mirror where both volumes belong to the same system (and I/O Group).
- IBM Spectrum Virtualize supports intercluster Metro Mirror where each volume belongs to a separate system. You can configure a specific system for partnership with another system. All intercluster Metro Mirror processing occurs between two IBM Spectrum Virtualize systems that are configured in a partnership.
- Intercluster and intracluster Metro Mirror can be used concurrently.
- IBM Spectrum Virtualize does not require that a control network or fabric is installed to manage Metro Mirror. For intercluster Metro Mirror, 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 Metro Mirror I/O.
- IBM Spectrum Virtualize implements a configuration model that maintains the Metro Mirror 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.

11.6.8 Metro Mirror attributes

The Metro Mirror function in IBM Spectrum Virtualize possesses the following attributes:

- A partnership is created between two IBM Spectrum Virtualize systems operating in the replication layer (for intercluster Metro Mirror).
- A Metro Mirror relationship is created between two volumes of the same size.
- To manage multiple Metro Mirror relationships as one entity, relationships can be made part of a Metro Mirror Consistency Group, which ensures data consistency across multiple Metro Mirror relationships and provides ease of management.
- When a Metro Mirror 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 Metro Mirror 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.

11.6.9 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 a Metro Mirror 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 remote copy is always in read-only mode.

While the Metro Mirror relationship is active, the auxiliary volume is not accessible for host application write I/O at any time. The IBM Storwize V7000 allows read-only access to the auxiliary volume when it contains a consistent image. IBM Storwize allows boot time operating system discovery to complete without an error, so that any hosts at the secondary site can be ready to start the applications with minimum delay, if required.

For example, many operating systems must read LBA zero to configure a logical unit. Although read access is allowed at the auxiliary in practice, the data on the auxiliary volumes cannot be read by a host because most operating systems 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 Metro Mirror 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 Metro Mirror 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. Metro Mirror 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 through failover management software. The IBM Storwize V7000 provides SNMP traps and programming (or scripting) for the CLI to enable this automation.

11.6.10 Global Mirror overview

This section describes the Global Mirror copy service, which is an asynchronous remote copy service. This service provides and maintains a consistent mirrored copy of a source volume to a target volume.

Global Mirror function establishes a Global Mirror relationship between two volumes of equal size. The volumes in a Global Mirror relationship are referred to as the *master* (source) volume and the *auxiliary* (target) volume, which is the same as Metro Mirror. Consistency groups can be used to maintain data integrity for dependent writes, which is similar to FlashCopy Consistency Groups.

Global Mirror 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.

Global Mirror has the following characteristics:

- Near-zero RPO
- Asynchronous
- Production application performance that is affected by I/O sequencing preparation time

Intracluster Global Mirror

Although Global Mirror is available for intracluster, it has no functional value for production use. Intracluster Metro Mirror provides the same capability with less processor use. However, leaving this functionality in place simplifies testing and supports client experimentation and testing (for example, to validate server failover on a single test system). As with Intracluster Metro Mirror, 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 Global Mirror 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 Global Mirror relationship.

Limit: When a local fabric and a remote fabric are connected for Global Mirror purposes, the ISL hop count between a local node and a remote node must not exceed seven hops.

11.6.11 Asynchronous remote copy

Global Mirror is an asynchronous remote copy technique. In asynchronous remote copy, 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 remote copy over distances that exceed the limitations of synchronous remote copy.

The Global Mirror function provides the same function as Metro Mirror remote copy, 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 11-89 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.

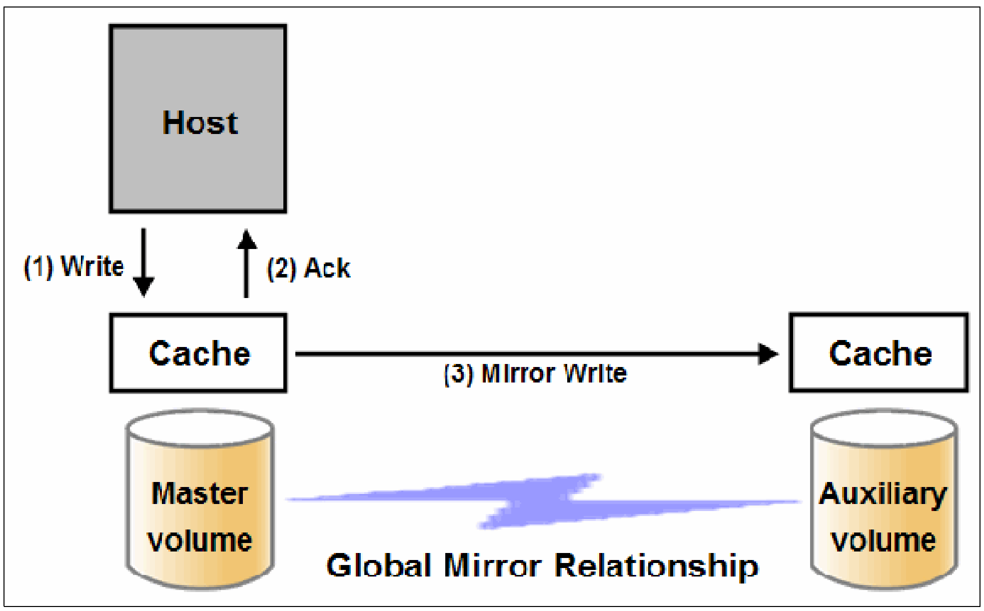


Figure 11-89 Global Mirror write sequence

The Global Mirror 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, Global Mirror 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.

Global Mirror 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 Global Mirror 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 Global Mirror secondary write I/O processing

In a failover scenario where the secondary site must become the master source of data, certain 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.

Global Mirror is supported over FC, FC over IP (FCIP), FC over Ethernet (FCoE), and native IP connections. The maximum distance cannot exceed 80 ms round trip, which is about 4000 km (2485.48 miles) between mirrored systems. However, starting with IBM Spectrum Virtualize V7.4, this distance was significantly increased for certain IBM Storwize Gen2 and IBM SAN Volume Controller configurations. Figure 11-90 shows the current supported distances for Global Mirror remote copy.

Software Version	Hardware type	Partnership type		
		FC	1Gb IP	10Gb IP
7.3 and earlier	All	80ms		
7.4, 7.5 and 7.6	* 2145-CG8 with 2nd HBA	250ms		
	* 2145-DH8			
	* 2076-524		80ms	10ms
7.7 and 7.8	* 2145-CG8 with 2nd HBA	250ms		
	* 2145-DH8			
	* 2076-524, 624 and AF6			

Figure 11-90 Supported Global Mirror distances

11.6.12 Global Mirror features

IBM Spectrum Virtualize Global Mirror supports the following features:

- Asynchronous remote copy of volumes that are dispersed over metropolitan-scale distances.
- IBM Spectrum Virtualize implements the Global Mirror relationship between a volume pair, with each volume in the pair being managed by an IBM Spectrum Virtualize system.
- IBM Spectrum Virtualize supports intracluster Global Mirror 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 11.8.2, “IP partnership limitations” on page 550.
- Intercluster and intracluster Global Mirror 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 Global Mirror. For intercluster Global Mirror, 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 Global Mirror I/O.
- IBM Spectrum Virtualize implements a configuration model that maintains the Global Mirror 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 Global Mirror 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 Global Mirror algorithm requires that only a single write is active on a volume. I/Os that overlap an active IO 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, which 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 Global Mirror 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 Global Mirror algorithm. There is still a need for master writes to 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 11-91 shows a colliding write sequence example.

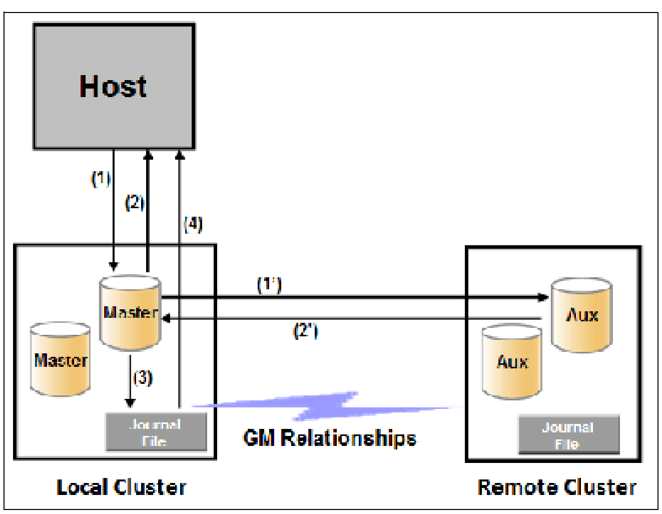


Figure 11-91 Colliding writes example

The following numbers correspond to the numbers that are shown in Figure 11-91 on page 527:

- (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

Global Mirror 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, also it provides the capability to test an application before the full deployment. The feature can be enabled separately for each of the intracluster or intercluster Global Mirrors.

By using the **`chsystem`** command, the delay setting can be setup and with the **`lssystem`** command the delay can be checked. 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, make sure that the delay simulator was properly disabled.

11.6.13 Using Change Volumes with Global Mirror

Global Mirror 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 happen 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

Global Mirror has functionality that is designed to address the following conditions, which might negatively affect certain Global Mirror 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 Global Mirror relationships. Change volumes use the FlashCopy functionality, but they cannot be manipulated as FlashCopy volumes because they are for a special purpose only. Change volumes replicate point-in-time images on a cycling period. The default is 300 seconds.

Your change rate needs to include only the condition of the data at the point-in-time 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.

Global Mirror with Change Volumes has the following characteristics:

- Larger RPO
- Point-in-time copies
- Asynchronous
- Possible system performance resource requirements because point-in-time copies are created locally

Figure 11-92 shows a simple Global Mirror relationship without Change Volumes.

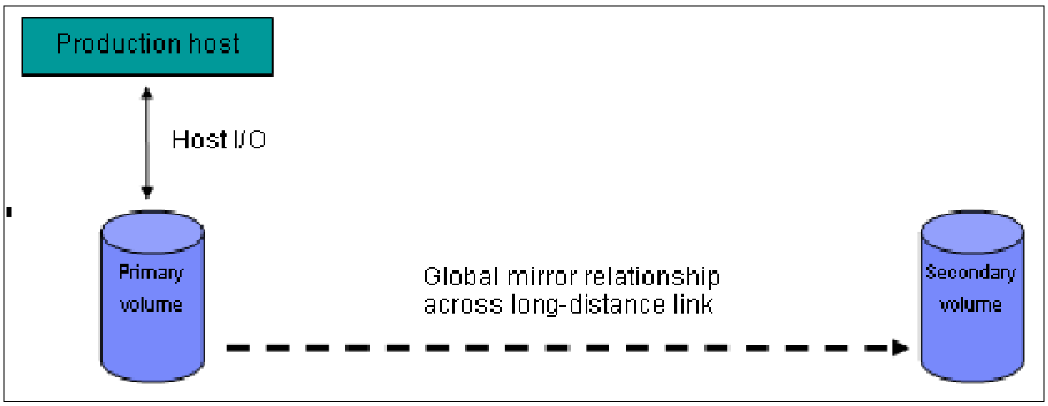


Figure 11-92 Global Mirror without Change Volumes

With Change Volumes, this environment looks as it is shown in Figure 11-93.

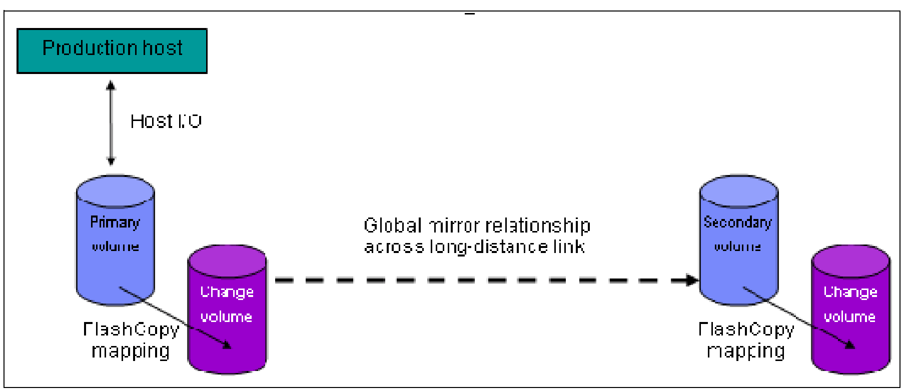


Figure 11-93 Global Mirror with Change Volumes

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 to one day). The primary change volume is then replicated to the secondary Global Mirror 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 11.1, “IBM FlashCopy” on page 436.

You can adjust the cycling period by using 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 Global Mirror with Change Volumes. Global Mirror with Change Volumes 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 Global Mirror Volumes with Change Volumes, make sure that you remember to select the change volume on the auxiliary (target) site. Failure to do so leaves you exposed during a resynchronization operation.

11.6.14 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 preferred 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 a remote copy relationship without affecting the host I/O to a particular volume.

The remote copy relationship type does not matter. The remote copy relationship type can be Metro Mirror, Global Mirror, or Global Mirror with Change Volumes. You can change the preferred node both to the source and target volumes that are participating in the remote copy relationship.

11.6.15 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 could 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.

Global Mirror 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 25MBps 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 remote copy relationships, increasing this value might affect overall system or intercluster link performance. The background copy rate can be changed to 1 - 1000 MBps.

11.6.16 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.

11.6.17 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 issuing the necessary commands. However, in certain environments, the available bandwidth can make this method unsuitable.

Use the following command sequence for a single relationship:

- Run **`mkrcrelationship`** without specifying the **`-sync`** option.
- Run **`startrcrelationship`** 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:

- Run **`mkrcrelationship`** with the **`-sync`** flag.
- Run **`startrcrelationship`** 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, possibly causing a data loss or data integrity exposure for hosts accessing data on the auxiliary volume.

11.6.18 Practical use of Global Mirror

The practical use of Global Mirror is similar to the Metro Mirror as described in 11.6.9, “Practical use of Metro Mirror” on page 523. The main difference between the two remote copy modes is that Global Mirror and Global Mirror with Change Volumes are mostly used on much larger distances than Metro Mirror. Weak link quality or insufficient bandwidth between the primary and secondary sites can also be a reason to prefer asynchronous Global Mirror over synchronous Metro Mirror. Otherwise, the use cases for MM/GM are the same.

11.6.19 IBM Spectrum Virtualize HyperSwap topology

The IBM HyperSwap topology is based on IBM Spectrum Virtualize Remote Copy 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, then 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-break 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 Storwize V7000, Spectrum Virtualize, HyperSwap, and VMware Implementation*, SG24-8317.

11.6.20 Consistency Protection for Global Mirror and Metro Mirror

Metro Mirror, Global Mirror, Global Mirror with change volumes, and HyperSwap Copy Services functions create remote copy 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 a remote copy 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 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 take place on either 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.

11.6.21 Valid combinations of FlashCopy, Metro Mirror, and Global Mirror

Table 11-10 lists the combinations of FlashCopy and MM/GM functions that are valid for a single volume.

Table 11-10 Valid combination for a single volume

FlashCopy	Metro Mirror or Global Mirror source	Metro Mirror or Global Mirror target
FlashCopy Source	Supported	Supported
FlashCopy Target	Supported	Not supported

11.6.22 Remote Copy configuration limits

Table 11-11 lists the MM/GM configuration limits.

Table 11-11 Metro Mirror configuration limits

Parameter	Value
Number of Metro Mirror or Global Mirror Consistency Groups per system	256
Number of Metro Mirror or Global Mirror relationships per system	8192
Number of Metro Mirror or Global Mirror relationships per Consistency Group	8192

Parameter	Value
Total Volume size per I/O Group	There is a per I/O Group limit of 1024 terabytes (TB) on the quantity of master and auxiliary volume address spaces that can participate in Metro Mirror and Global Mirror 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.

11.6.23 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 11-94, the MM/GM relationship diagram shows an overview of the status that can apply to a MM/GM relationship in a connected state.

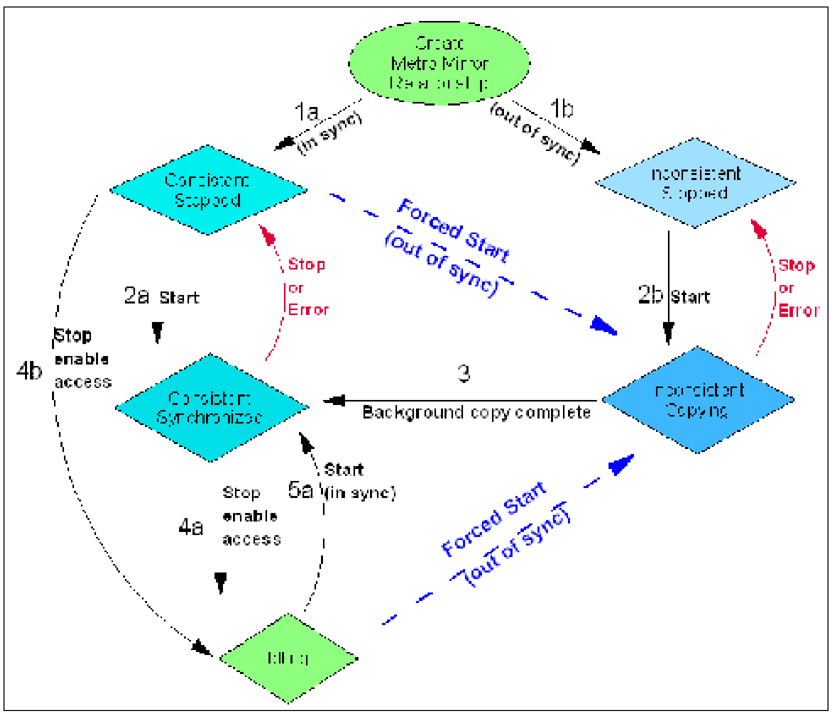


Figure 11-94 Metro Mirror or Global Mirror mapping state diagram

When the MM/GM relationship is created, you can specify whether the auxiliary volume is already 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 11-94:

- 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, issue the **`svctask stoprcrelationship`** 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 536.

> **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 spans 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, considering 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 Storwize V7000 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 11.1.13, “FlashCopy and image mode Volumes” on page 459.

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 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 point-in-time 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 11.6.20, "Consistency Protection for Global Mirror and Metro Mirror" on page 532.

Detailed states

The following sections describe the states that are portrayed to the user for Consistency Groups or relationships. They also describe 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 use 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 enter 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 issue 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. Either 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 there is a loss of consistency.

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 (either 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 Disaster Recovery 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.

11.7 Remote Copy commands

This section presents commands that need to be issued to create and operate remote copy services.

11.7.1 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 disaster recovery.
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.

For more information about MM/GM commands, see *IBM System Storage SAN Volume Controller and IBM Storwize V7000 Command-Line Interface User's Guide,* GC27-2287.

The command set for MM/GM contains the following broad groups:

- Commands to create, delete, and manipulate relationships and Consistency Groups
- Commands to cause state changes

If a configuration command affects more than one system, MM/GM coordinates configuration activity between the systems. Certain configuration commands can be performed only when the systems are connected, and fail with no effect when they are disconnected.

Other configuration commands are permitted even though 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 either of the following methods:

- CLI can be used for scripting and automation.
- GUI can be used for one-off tasks.

11.7.2 Listing available system partners

Use 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.

11.7.3 Changing the system parameters

When you want to change system parameters specific to any remote copy or Global Mirror 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 remote copy relationship can synchronize. The default value for the relationship bandwidth limit is 25 MBps (megabytes per second), 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 Global Mirror 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 Global Mirror link tolerance timer starts counting down. This threshold value determines the additional effect that Global Mirror operations can add to the response times of the Global Mirror 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 is the system-wide setting, and applies to MM/GM relationships.

Use 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 using the **lssystem** ***<system_name>*** command.

Focus on the **gmlinktolerance** 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 begin to queue at the primary system. This queue results in an extended response time to application hosts. In this situation, the **gmlinktolerance** 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 **gmlinktolerance** feature by setting the **gmlinktolerance** value to `0` (zero). However, the **gmlinktolerance** feature cannot protect applications from extended response times if it is disabled. It might be appropriate to disable the **gmlinktolerance** 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 **gmlinktolerance** 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 **gmlinktolerance** feature might detect the high latency and stop the GM relationships.

  Disabling the **gmlinktolerance** 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.

11.7.4 System partnership

To create a partnership, use the following command, depending on the connection type:

- Use 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.
- Use 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 issue this command on both systems. This step is a prerequisite for creating MM/GM relationships between volumes on the IBM Spectrum Virtualize systems.

When creating the partnership, you must specify the Link Bandwidth and can specify the Background Copy Rate:

- The Link Bandwidth, expressed in Mbps (megabits per second), 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, then 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, use the **chpartnership -backgroundcopyrate** ***<percentage_of_link_bandwidth>*** command to specify the percentage of whole link capacity to be used by the background copy process.

11.7.5 Creating a Metro Mirror/Global Mirror consistency group

Use the **mkrcconsistgrp** 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 using the **chrelationship** command.

11.7.6 Creating a Metro Mirror/Global Mirror relationship

Use 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, a Metro Mirror relationship is created rather than a Global Mirror 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 an existing 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 an existing Consistency Group, or it can be a stand-alone MM/GM relationship.

The lsrcrelationshipcandidate command

Use 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.

11.7.7 Changing Metro Mirror/Global Mirror relationship

Use 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.

11.7.8 Changing Metro Mirror/Global Mirror consistency group

Use the **chrcconsistgrp** command to change the name of an MM/GM Consistency Group.

11.7.9 Starting Metro Mirror/Global Mirror relationship

Use the **startrcrelationship** command to start the copy process of an MM/GM relationship.

When the command is issued, 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 already a part of a consistency group.

You can issue 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) takes place. Therefore, this data is unusable for Disaster Recovery 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.

11.7.10 Stopping Metro Mirror/Global Mirror relationship

Use the **stoprcrelationship** 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 issue a **startrcrelationship** 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 **stoprcrelationship** command to enable write access to the auxiliary volume.

11.7.11 Starting Metro Mirror/Global Mirror consistency group

Use the **startrcconsistgrp** 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.

11.7.12 Stopping Metro Mirror/Global Mirror consistency group

Use the **stoprcconsistgrp** 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 issue 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.

11.7.13 Deleting Metro Mirror/Global Mirror relationship

Use the **rmrcrelationship** 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 themselves.

If the relationship is disconnected at the time that the command is issued, the relationship is deleted only on 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 issue the **rmrcrelationship** 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.

11.7.14 Deleting Metro Mirror/Global Mirror consistency group

Use the **rmrcconsistgrp** 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 only on 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 issue the **`rmrcconsistgrp`** 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.

11.7.15 Reversing Metro Mirror/Global Mirror relationship

Use 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.

11.7.16 Reversing Metro Mirror/Global Mirror consistency group

Use the **`switchrcconsistgrp`** 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:** Remember that 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.

11.8 Native IP replication

IBM Spectrum Virtualize can implement Remote Copy services by using FC connections or IP connections. This chapter describes the IBM Spectrum Virtualize IP replication technology and implementation.

> **Demonstration:** The IBM Client Demonstration Center shows how data is replicated using Global Mirror with Change Volumes (cycling mode set to `multiple`). This configuration perfectly fits the new IP replication functionality because it is well designed for links with high latency, low bandwidth, or both. For more information, see this web page (log in required).

11.8.1 Native IP replication technology

Remote Mirroring over IP communication is supported on the IBM SAN Volume Controller and Storwize Family systems by using Ethernet communication links. The 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 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 disaster recovery 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 11-95.

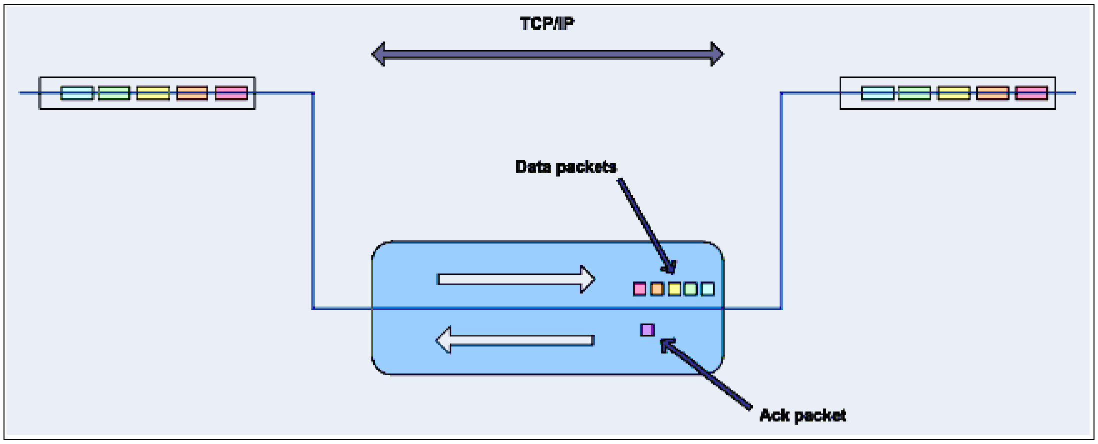


Figure 11-95 Typical Ethernet network data flow

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 (VC) that share IP links and addresses. The artificial intelligence 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 VC's ACK, it sends more packets across other VCs. If packets are lost from any VC, data is automatically retransmitted, as shown in Figure 11-96.

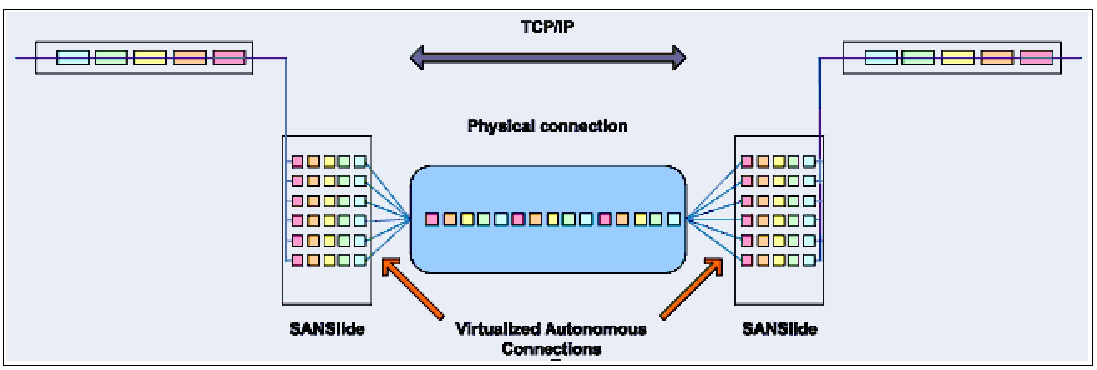


Figure 11-96 Optimized network data flow by using Bridgeworks SANSlide technology

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:

- Metro Mirror

  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.

- Global Mirror and GM with Change Volumes

  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 GM with Change Volumes. GM with Change Volumes is the preferred method for use with native IP replication.

> **Note:** For IP partnerships, generally use the Global Mirror with change Volumes method of copying (asynchronous copy of changed grains only). This method can have performance benefits. Also, Global Mirror and Metro Mirror might be more susceptible to the loss of synchronization.

11.8.2 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 remote copy 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 remote copy 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 need to 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 is 80 milliseconds (ms) for 1 Gbps links.
- The maximum supported round-trip time 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 remote copy 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 LAN (VLAN) tagging of the IP addresses that are configured for remote copy is supported starting with V7.4.
- Management IP and iSCSI IP on the same port can be in a different network starting with V7.4.
- An added layer of security is provided by using Challenge Handshake Authentication Protocol (CHAP) authentication.
- Transmission Control Protocol (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 Remote Copy (RC) data session per physical link can be established. It is intended that only one connection (for sending/receiving Remote Copy 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 could be connected (by Ethernet switches) to this physical link. Similarly, multiple IP addresses on remote system B could be connected (by Ethernet switches) to the same physical link. At any point in 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 or 10 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

  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 11-97shows the scaling of host I/O.

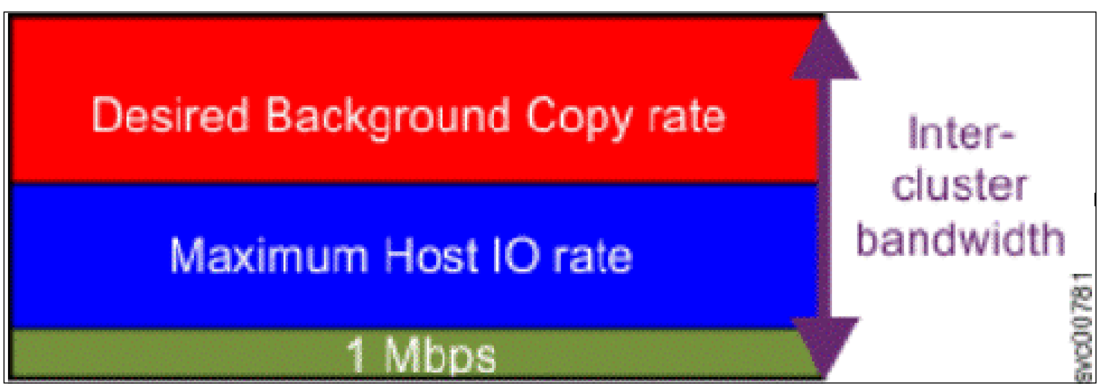


Figure 11-97 Scaling of host I/O

The equation that can describe the approximate minimum bandwidth that is required between two systems with < 5 ms round-trip time and errorless link follows:

```
Minimum intersite link bandwidth in Mbps > Required Background Copy in Mbps +
Maximum Host I/O in Mbps + 1 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/resyncs 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 GM with Change Volumes.

11.8.3 IP Partnership and data compression

When creating an IP partnership between two systems, you can specify whether you want to use the data compression or not. 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 (V7.7 or later).

To fully enable compression in an IP partnership, each system must enable compression. When compression is enabled on the local system, data sent to the remote system is compressed so it needs to be decompressed on the remote system, and vice versa.

Although IP compression uses the same RtC algorithm as for volumes, there is no need for a RtC license on any of the local or remote system.

Replicated volumes using IP partnership compression can be compressed or uncompressed on the system because there is no link between volumes compression and IP Replication compression:

- Read operation decompresses the data
- Decompressed data is transferred to the Remote Copy code
- Data is compressed before being sent over the IP link
- Remote system remote copy code decompresses the received data
- Write operation on a volume compresses the data

11.8.4 VLAN support

Starting with V7.4, VLAN tagging is supported for both iSCSI host attachment and IP replication. Hosts and remote-copy 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 the same 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 used for either 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 will 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, there are no paths 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. You can use the CLI or GUI to optionally set the VLAN ID for port IPs on both systems in the IP partnership.
- When a VLAN ID is configured for the port IP addresses that are used in remote copy 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.

11.8.5 IP partnership and terminology

The IP partnership terminology and abbreviations that are used are listed in Table 11-12.

Table 11-12 Terminology for IP partnership

IP partnership terminology	Description
Remote copy group or Remote copy port group	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 remote copy group can form remote copy connections with the partner system: ▶ 0 – Ports that are not configured for remote copy ▶ 1 – Ports that belong to remote copy port group 1 ▶ 2 – Ports that belong to remote copy port group 2 Each IP address can be shared for iSCSI host attach and remote copy functionality. Therefore, appropriate settings must be applied to each IP address.
IP partnership	Two systems that are partnered to perform remote copy over native IP links.
FC partnership	Two systems that are partnered to perform remote copy over native Fibre Channel 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 IPs 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.
linkbandwidthmbits	Aggregate bandwidth of all physical links between two sites in Mbps.

IP partnership terminology	Description
IP partnership or partnership over native IP links	These terms are used to describe the IP partnership feature.
Discovery	Process by which two IBM Spectrum Virtualize systems exchange information about their IP address configuration. For IP-based partnerships, only IP addresses configured for Remote Copy are discovered. 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).

11.8.6 States of IP partnership

The different partnership states in IP partnership are listed in Table 11-13.

Table 11-13 States of IP partnership

State	Systems connected	Support for active remote copy 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 remote copy 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 Metro Mirror and Global Mirror relationships.

The following steps must be completed to establish two systems in the IP partnerships:

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 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 **cfgportip** CLI command), and completes the following steps:
   a. Configure the IP addresses for remote copy data.
   b. Add the IP addresses in the respective remote copy 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 GM with Change Volume 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.

11.8.7 Remote copy groups

This section describes remote copy groups (or remote copy 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 remote copy port groups was introduced.

Remote copy port group ID is a numerical tag that is associated with an IP port of an IBM Spectrum Virtualize system to indicate which physical IP link it is connected to. Multiple nodes might be connected to the same physical long-distance link, and must therefore share the same remote copy port group ID.

In scenarios with two physical links between the local and remote clusters, two remote copy 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 the **cfgportip** CLI command.

> **Remember:** IP ports on both partners must have been configured with identical remote copy 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 remote copy port groups. The system supports three remote copy groups: 0, 1, and 2.

The systems' IP addresses are, by default, in remote copy port group 0. Ports in port group 0 are not considered for creating remote copy data paths between two systems. For partnerships to be established over IP links directly, IP ports must be configured in remote copy group 1 if a single inter-site link exists, or in remote copy 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 remote copy port group.

The administrator might want to use IPv6 addresses for remote copy 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 explicitly 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 SAN Volume Controller node must have only a single remote copy port group that is configured, 1 or 2. The remaining IP addresses must be in remote copy port group 0.

11.8.8 Supported configurations

Note: For explanation purposes, this section shows a node with 2 ports available, 1 and 2. This number generally increments when IBM SAN Volume Controller nodes model DH8 or SV1 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 11-98 (configuration 1).

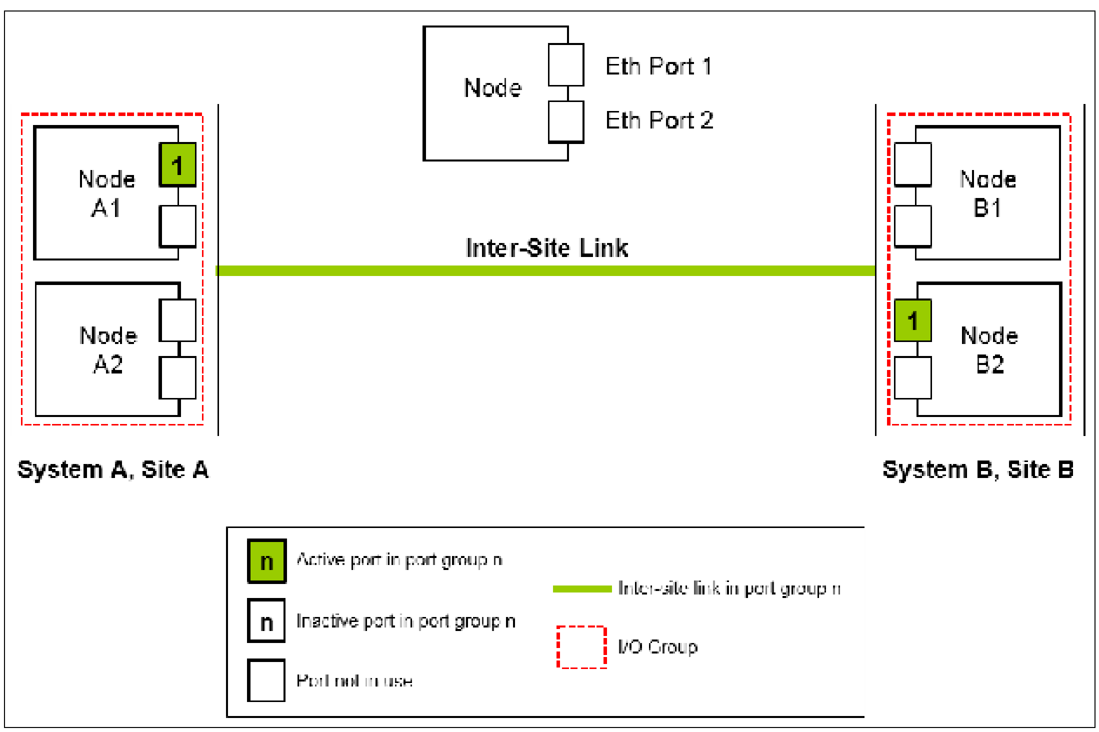


Figure 11-98 Single link with only one remote copy port group configured in each system

As shown in Figure 11-98 on page 556, there are two systems:

- System A
- System B

A single remote copy port group 1 is created on Node A1 on System A and on Node B2 on System B because there is only a single inter-site link to facilitate the IP partnership traffic. An administrator might choose to configure the remote copy port group on Node B1 on System B rather than Node B2.

At any time, only the IP addresses that are configured in remote copy 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 a remote copy 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 were to 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 11-99 (configuration 2).

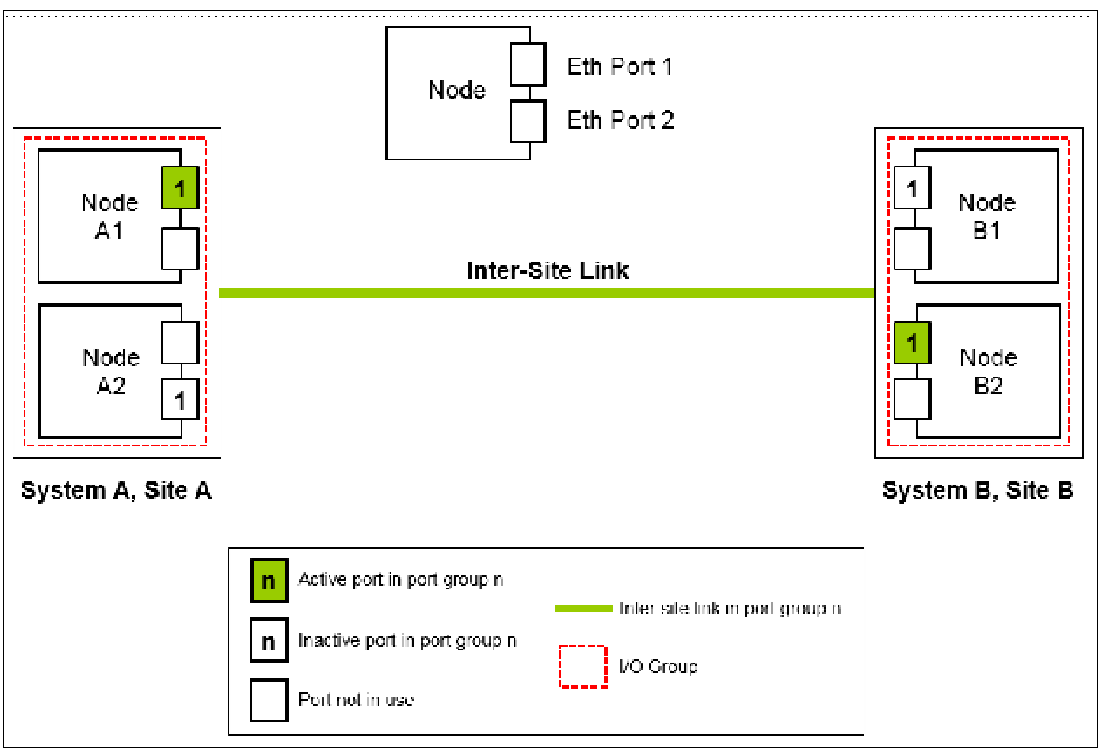


Figure 11-99 One remote copy group on each system and nodes with failover ports configured

As shown in Figure 11-99 on page 557, there are two systems:

– System A
– System B

A single remote copy port group 1 is configured on two Ethernet ports, one each on Node A1 and Node A2 on System A. Similarly, a single remote copy 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 remote copy 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 there is a node failure. 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 **`lsportip`** output (reported as `used`).

This configuration has the following characteristics:

– Each node in the I/O group has the same remote copy port group that is configured. However, only one port in that remote copy 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 11-100 on page 559 (configuration 3).

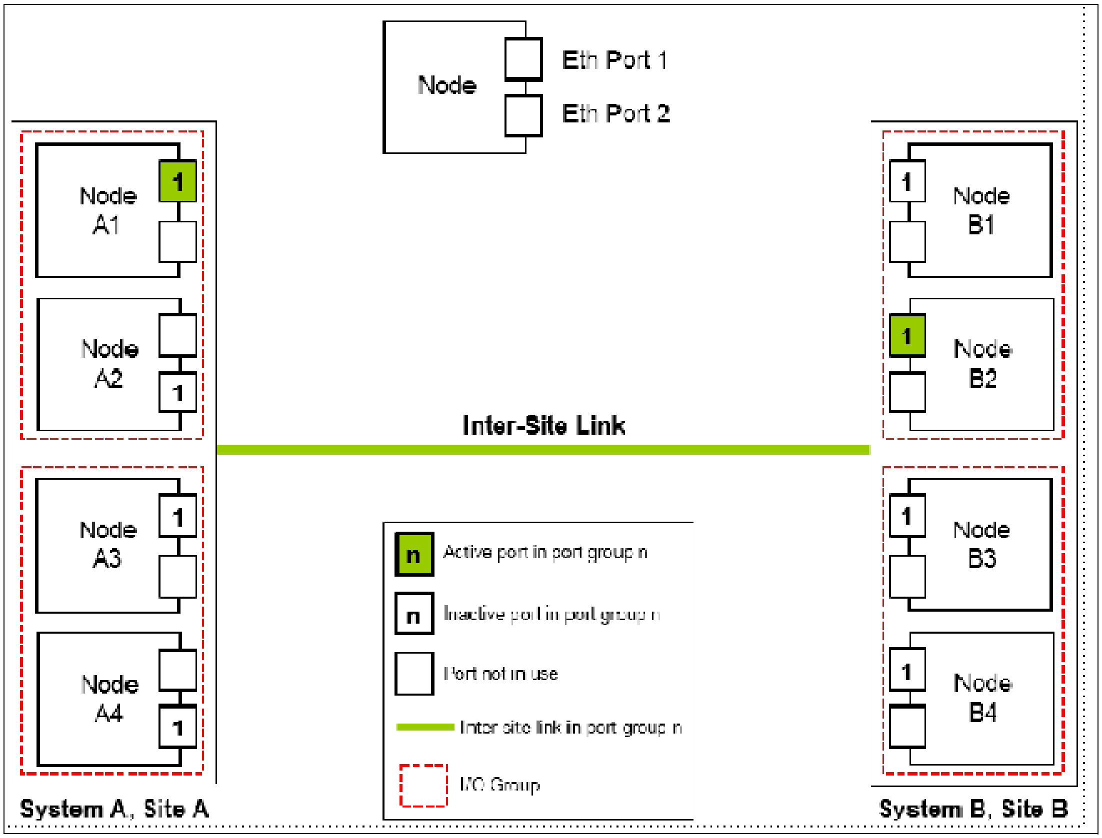


Figure 11-100 Multinode systems single inter-site link with only one remote copy port group

As shown in Figure 11-100, there are two 4-node systems:

– System A
– System B

A single remote copy 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 remote copy group 1, only one Ethernet port in each remote copy 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 remote copy 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 remote copy relationships to stop. The administrator might need to manually verify the issues in the event log and start the relationships or remote copy 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 remote copy port group that is configured in both I/O groups. However, only one port in that remote copy 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 11-101 (configuration 4).

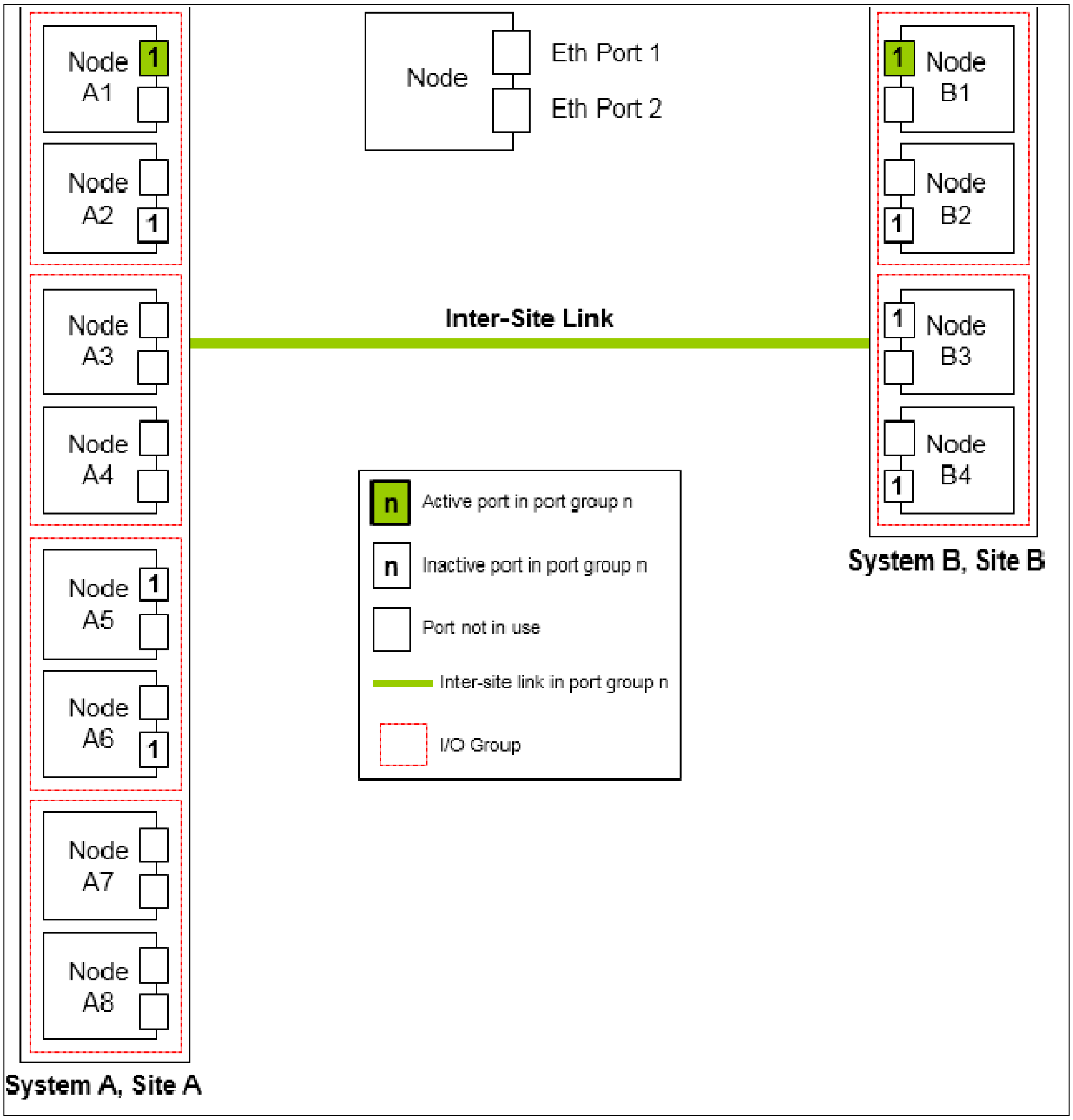


Figure 11-101 Multinode systems single inter-site link with only one remote copy port group

As shown in Figure 11-101, there is an eight-node system (System A in Site A) and a four-node system (System B in Site B). A single remote copy port group 1 is configured on nodes A1, A2, A5, and A6 on System A at Site A. Similarly, a single remote copy port group 1 is configured on nodes B1, B2, B3, and B4 on System B.

Although there are four I/O groups (eight nodes) 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 using one of the ports that is configured in remote copy 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, and this delay might cause partnerships to change to the `Not_Present` state.

This process can lead to remote copy 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 is provided in **`lsportip`** output (reported as `used`).

This configuration has the following characteristics:

- Each node has the remote copy 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 remote copy 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 11-102 (configuration 5).

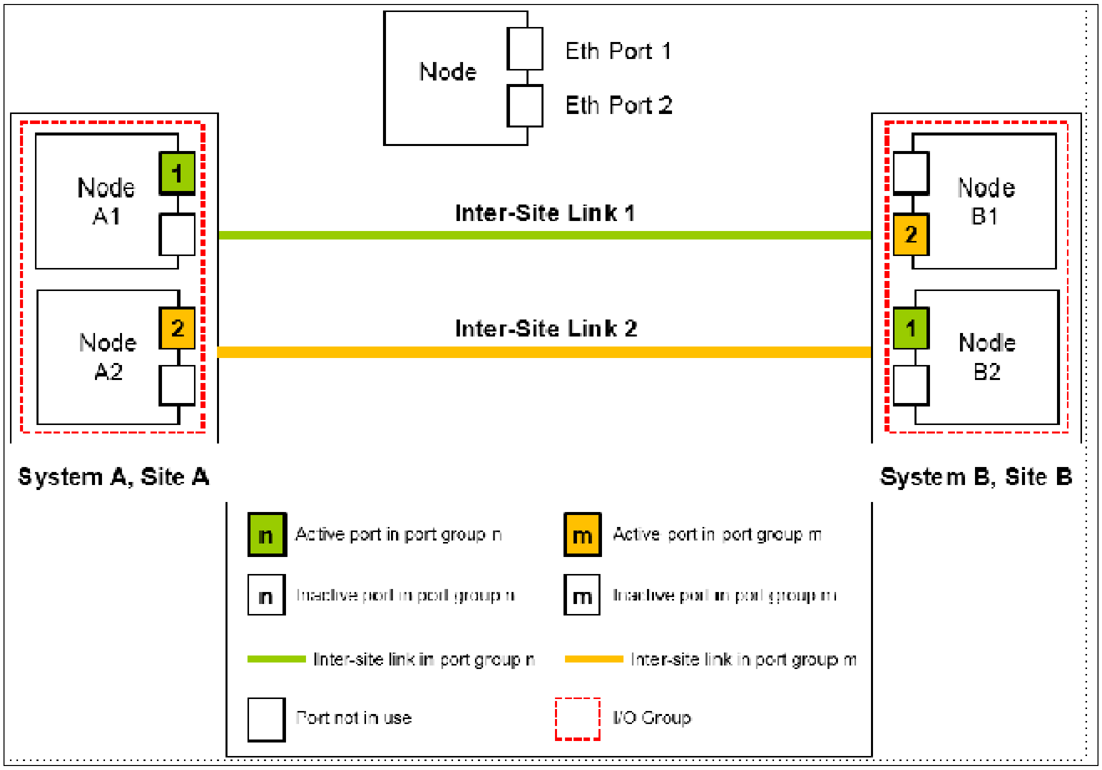


Figure 11-102 Dual links with two remote copy groups on each system configured

As shown in Figure 11-102, remote copy 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 remote copy 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 remote copy 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 remote copy port groups are configured.
- Each node has only one IP port in remote copy port group 1 or 2.
- Both the IP ports in the two remote copy 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 happens, 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 11-103 (configuration 6).

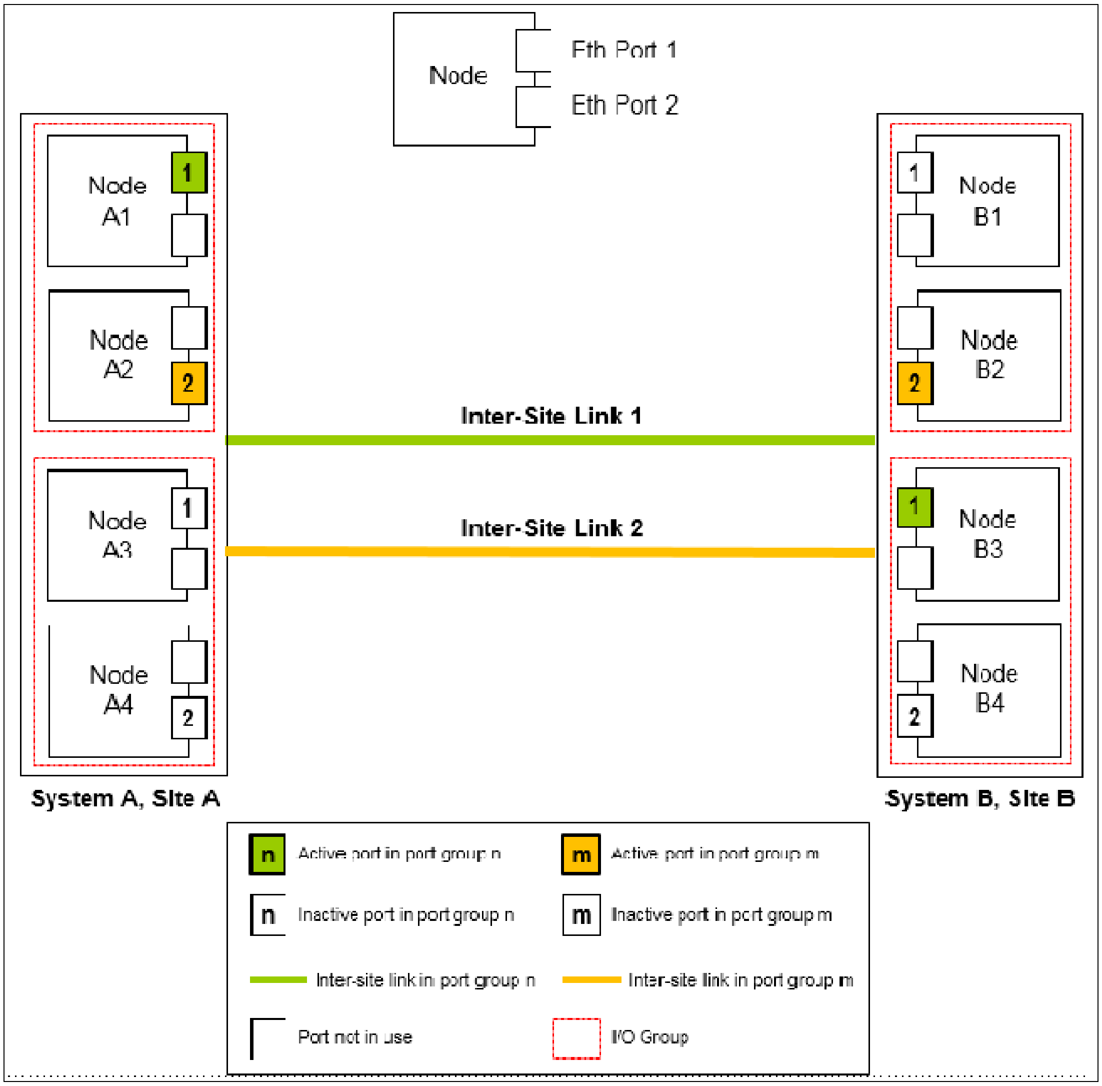


Figure 11-103 Multinode systems with dual inter-site links between the two systems

As shown in Figure 11-103, there are two 4-node systems:

- 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 remote copy port groups 1 or 2.

Although two ports are configured in remote copy port groups 1 and 2 on each system, only one IP port in each remote copy port group on each system actively participates in IP partnership. The other ports that are configured in the same remote copy port group act as standby ports in the event of failure. Which port in a configured remote copy 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, remote copy port group 2) and at the same time the failover also causes discovery in remote copy port group 1. Therefore, the IP partnership traffic continues from Node A3 on which remote copy port group 1 is configured. The details of the particular IP port that is actively participating in IP partnership process is provided in the **`lsportip`** output (reported as `used`).

This configuration has the following characteristics:

- Each node has the remote copy port group that is configured in the I/O groups 1 or 2. However, only one port per system in both remote copy port groups remains active and participates in IP partnership.
- Only a single port per system from each configured remote copy 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 remote copy port group but in a different I/O Group.
- The pathing algorithm can start discovery of available ports in the affected remote copy 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 11-104 on page 565 (configuration 7).

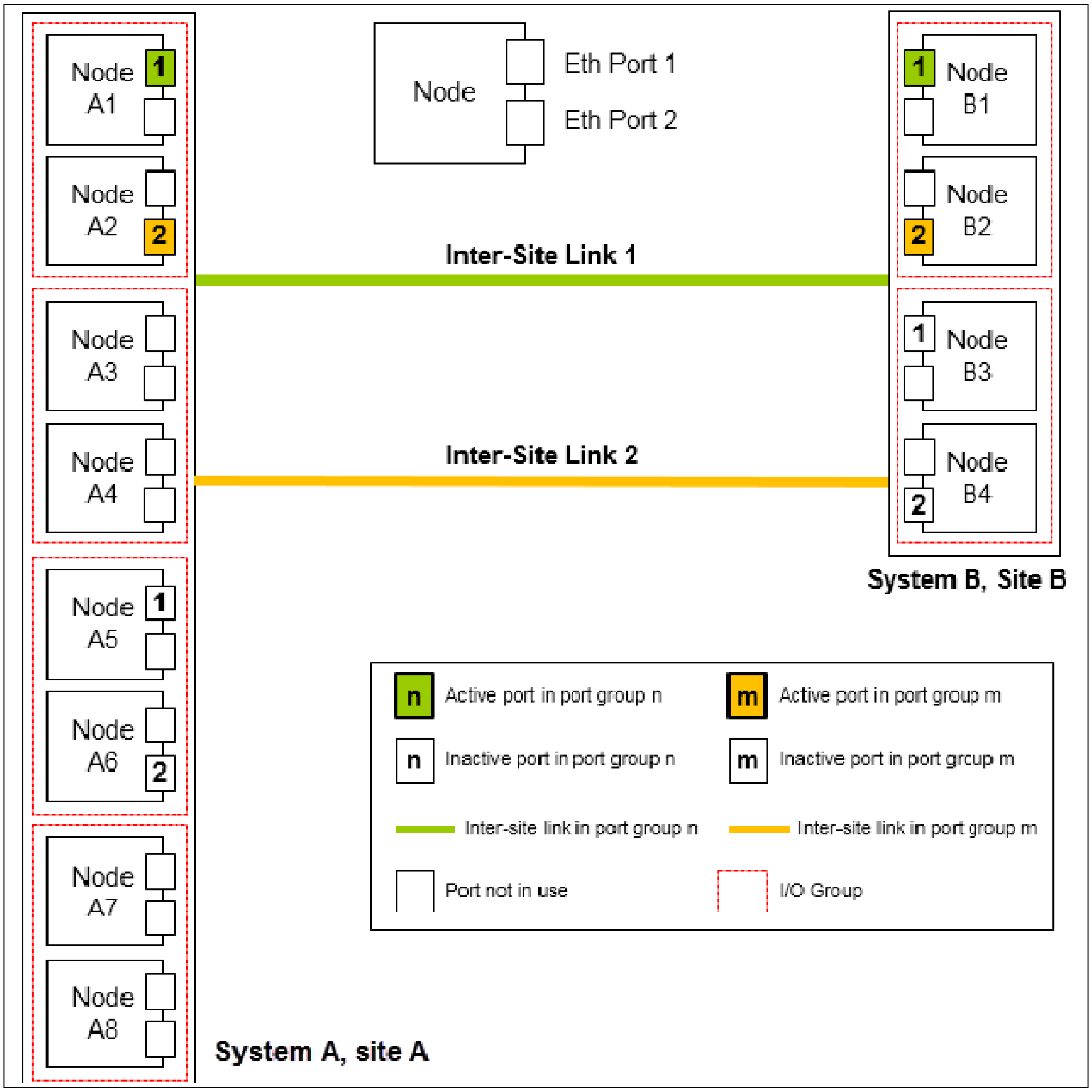


Figure 11-104 Multinode systems (two I/O groups on each system) with dual inter-site links between the two systems

As shown in Figure 11-104, there is an eight-node System A in Site A and a four-node System B in Site B. Because a maximum of two I/O groups in IP partnership is supported in a system, although there are four I/O groups (eight nodes), nodes from only two I/O groups are configured with remote copy port groups in System A. The remaining or all of the I/O groups can be configured to be remote copy partnerships over FC.

In this configuration, two links and two I/O groups are configured with remote copy 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 remote copy 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, remote copy port group 2). The failover also causes IP partnership traffic to continue from Node A5 on which remote copy port group 1 is configured. The details of the particular IP port actively participating in IP partnership process is provided in the **`lsportip`** output (reported as `used`).

This configuration has the following characteristics:

– There are two I/O Groups with nodes in those I/O groups that are configured in two remote copy port groups because there are two inter-site links for participating in IP partnership. However, only one port per system in a particular remote copy port group remains active and participates in IP partnership.

– One port per system from each remote copy 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 remote copy port group.

– The path selection algorithm starts discovery of available ports in the affected remote copy 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 remote copy partnerships with other systems.

- An example of an *unsupported* configuration for a single inter-site link is shown in Figure 11-105 (configuration 8).

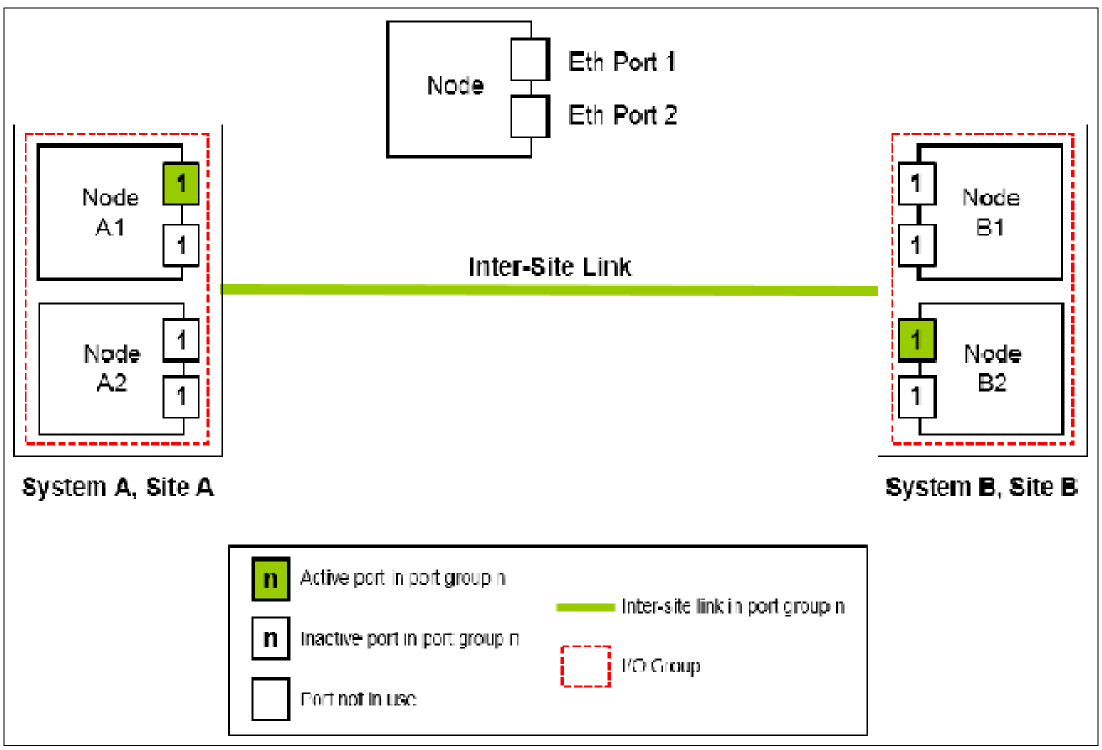


Figure 11-105 Two node systems with single inter-site link and remote copy port groups configured

As shown in Figure 11-105, this configuration is similar to Configuration 2, but differs because each node now has the same remote copy 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 remote copy group on the same node is *not supported*.

- An example of an *unsupported* configuration for a dual inter-site link is shown in Figure 11-106 (configuration 9).

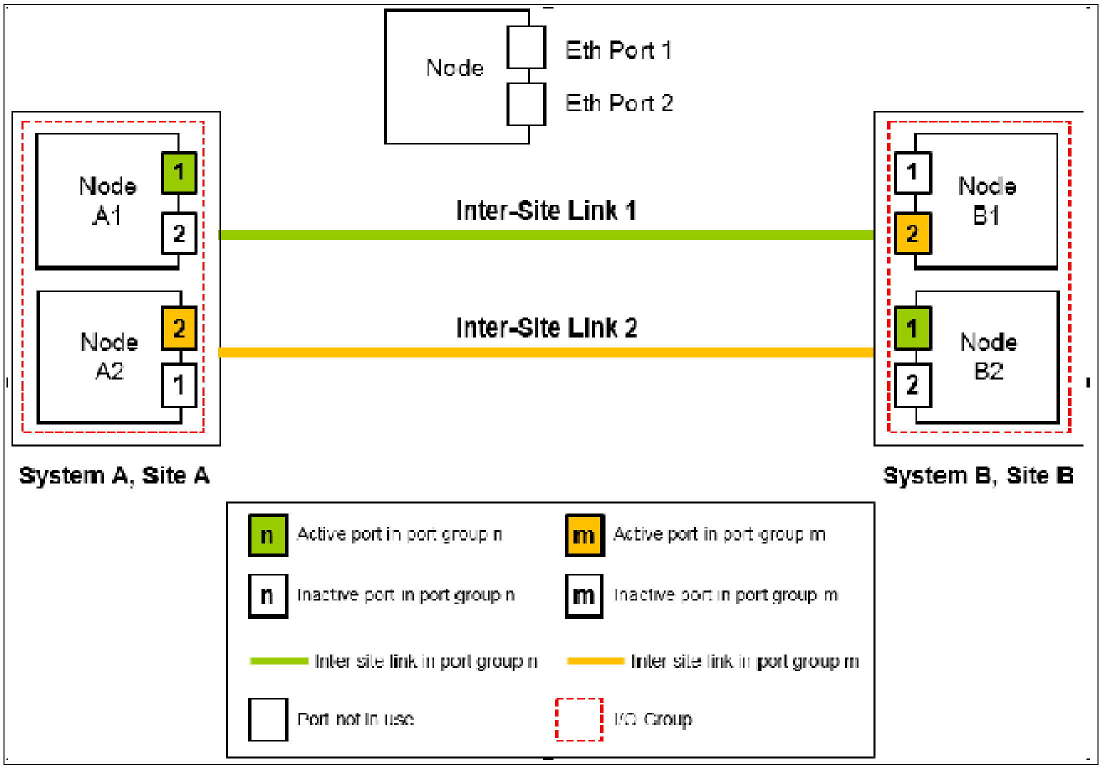


Figure 11-106 Dual Links with two Remote Copy Port Groups with failover Port Groups configured

As shown in Figure 11-106, this configuration is similar to Configuration 5, but differs because each node now also has two ports that are configured with remote copy 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 11-107 (configuration 10).

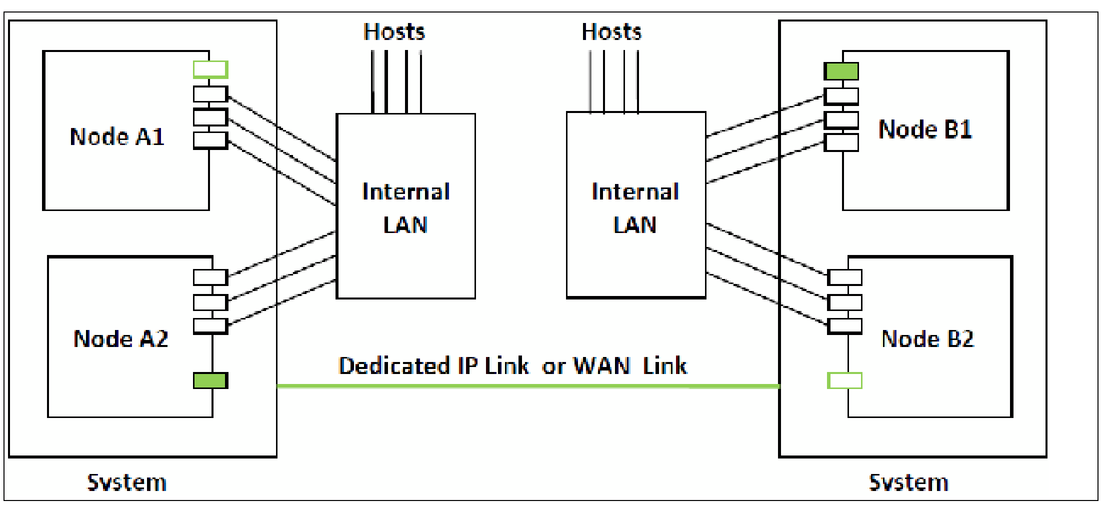


Figure 11-107 Deployment example

In this configuration, one port on each node in System A and System B is configured in remote copy group 1 to establish IP partnership and support remote copy relationships. A dedicated inter-site link is used for IP partnership traffic, and iSCSI host attach is disabled on those ports.

The following configuration process is 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 remote copy on both the systems by using the following settings:

   - Remote copy 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 on Storwize V7000).

e. Establish IP partnerships from both of the systems.

f. After the partnerships are in the `Fully_Configured` state, you can create the remote copy relationships.

- Figure 11-107 on page 567 is an example deployment for the configuration that is shown in Figure 11-101 on page 560. Ports that are shared with host access are shown in Figure 11-108 (configuration 11).

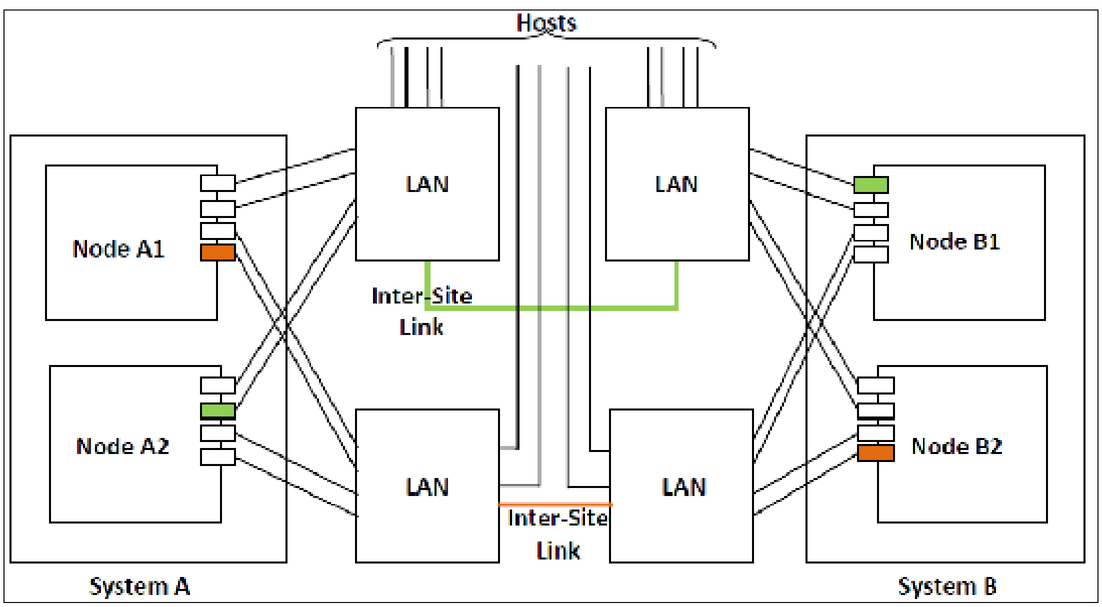


Figure 11-108 Deployment example

In this configuration, IP ports are to be shared by both iSCSI hosts and for IP partnership.

The following configuration process is 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 remote copy on System A1 by using the following settings:

- Node 1:

  - Port 1, remote copy port group 1
  - Host: Yes
  - Assign IP address

- Node 2:

  - Port 4, Remote Copy Port Group 2
  - Host: Yes
  - Assign IP address

d. Configure IP ports for remote copy on System B1 by using the following settings:

- Node 1:

  - Port 1, remote copy port group 1
  - Host: Yes
  - Assign IP address

- Node 2:

  - Port 4, remote copy port group 2
  - Host: Yes
  - Assign IP address

e. Check the MTU levels across the network as set (default MTU is 1500 on IBM SAN Volume Controller and Storwize V7000).

f. Establish IP partnerships from both systems.

g. After the partnerships are in the `Fully_Configured` state, you can create the remote copy relationships.

11.9 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 a remote copy level.

> **Note:** The **Copy Services** → **Consistency Groups** menu relates to FlashCopy consistency groups only, not Remote Copy ones.

The following panels are used to visualize and manage your remote copies:

- The Remote Copy panel

  To open the Remote Copy panel, click **Copy Services** → **Remote Copy** in the main menu as shown in Figure 11-109.

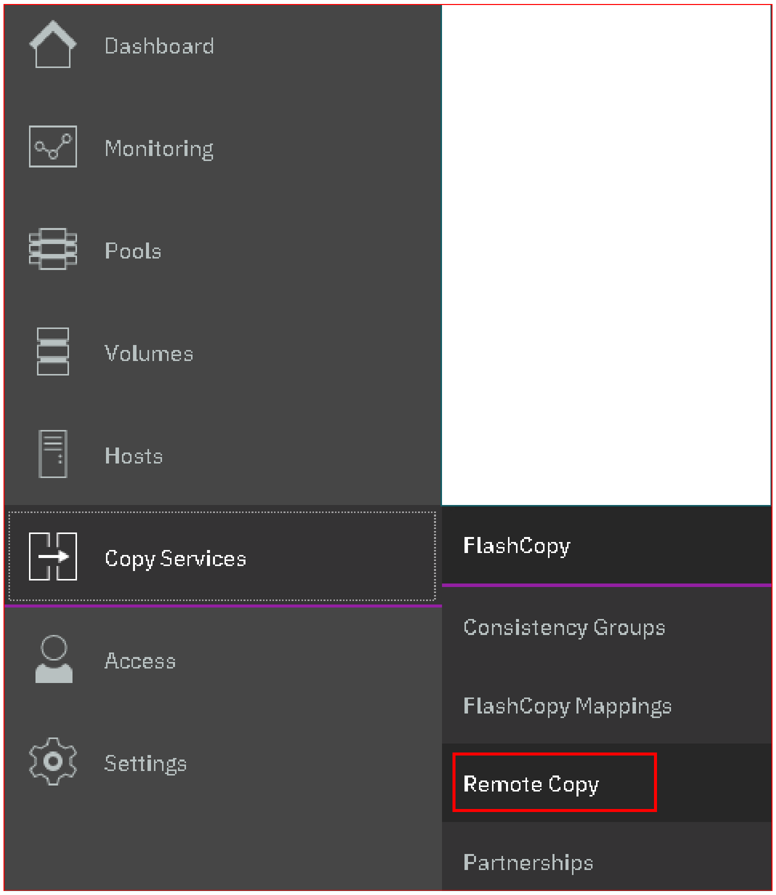


Figure 11-109 Remote Copy menu

  The Remote Copy panel opens, as shown in Figure 11-110.

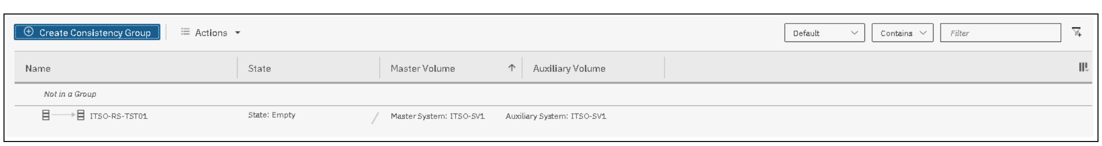


Figure 11-110 Remote Copy panel

- Partnerships panel

  To open the Partnership panel, click **Copy Services** → **Partnership** in the main menu, as shown in Figure 11-111.

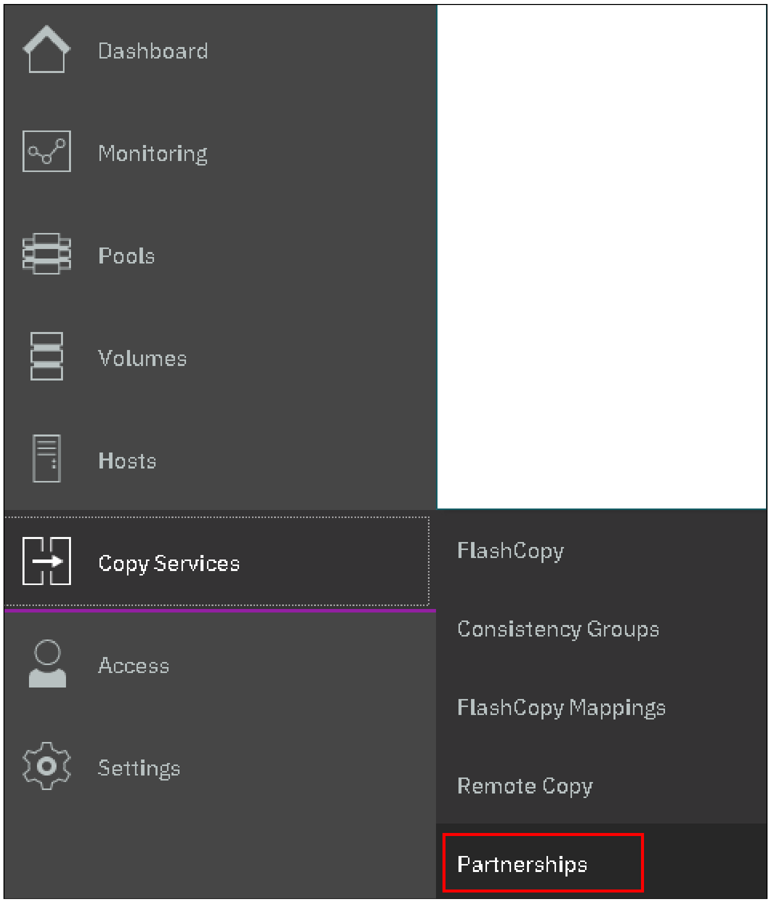


Figure 11-111 Partnership menu

The Partnership panel opens, as shown in Figure 11-112.

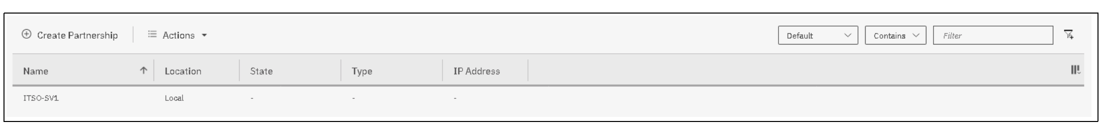


Figure 11-112 Partnership panel

11.9.1 Creating Fibre Channel partnership

Intra-cluster Metro Mirror: If you are creating an intra-cluster Metro Mirror, do not perform this next step to create the Metro Mirror partnership. Instead, see 11.9.2, “Creating remote copy relationships” on page 573.

To create an FC partnership between IBM Spectrum Virtualize systems by using the GUI, open the Partnerships panel and click **Create Partnership** to create a partnership, as shown in Figure 11-113 on page 572.

Figure 11-113 Creating a Partnership

In the Create Partnership window, enter the following information, as shown in Figure 11-114:

1. Select the partnership type (**Fibre Channel** or **IP**). If you choose IP partnership, you must provide the IP address of the partner system and the partner system's CHAP key.
2. If your partnership is based on Fibre Channel protocol, select an available partner system from the menu. If no candidate is available, the `This system does not have any candidates` error message is displayed.
3. Enter a link bandwidth in Mbps that is used by the background copy process between the systems in the partnership.
4. Enter the background copy rate.
5. Click **OK** to confirm the partnership relationship.

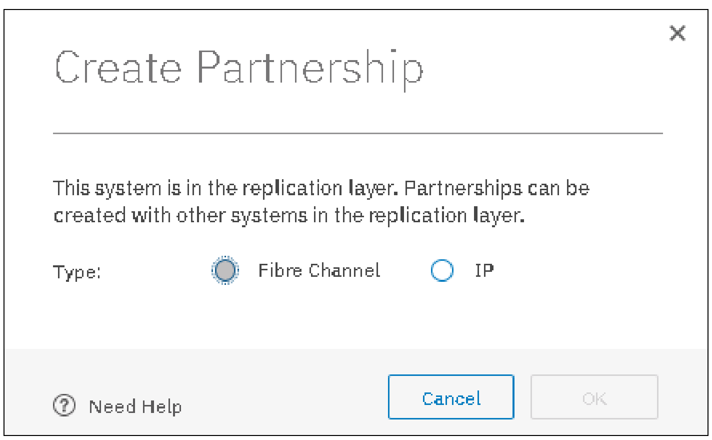


Figure 11-114 Creating a Partnership details

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 `Partially Configured`.

When both sides of the system partnership are defined, the partnership is displayed as `Fully Configured`, as shown in Figure 11-115.

Figure 11-115 Fully configured FC partnership

11.9.2 Creating remote copy relationships

This section shows how to create remote copy 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 a remote copy relationship, complete these steps:

1. Open the **Copy Services** → **Remote Copy** panel.
2. Right-click the Consistency Group that you want to create the relationship for, and select **Create Relationship**, as shown in Figure 11-116. If you want to create a stand-alone relationship (not in a consistency group), right-click the **Not in a Group** group.

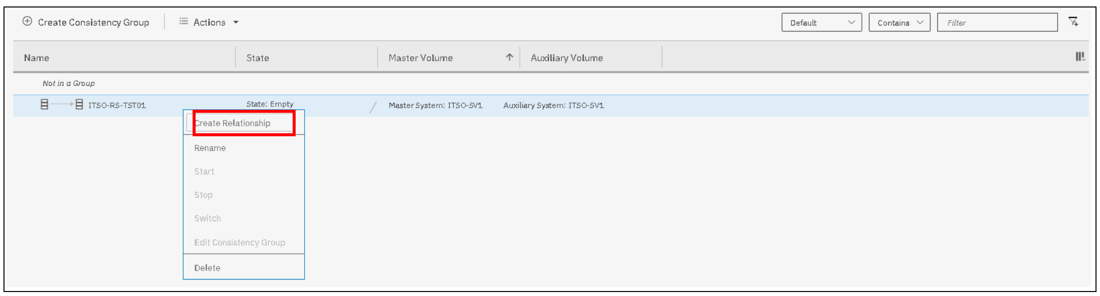


Figure 11-116 Creating remote copy relationships

3. In the Create Relationship window, select one of the following types of relationships that you want to create, as shown in Figure 11-117, and click **Next**.
   - Metro Mirror
   - Global Mirror (with or without Consistency Protection)
   - Global Mirror with Change Volumes

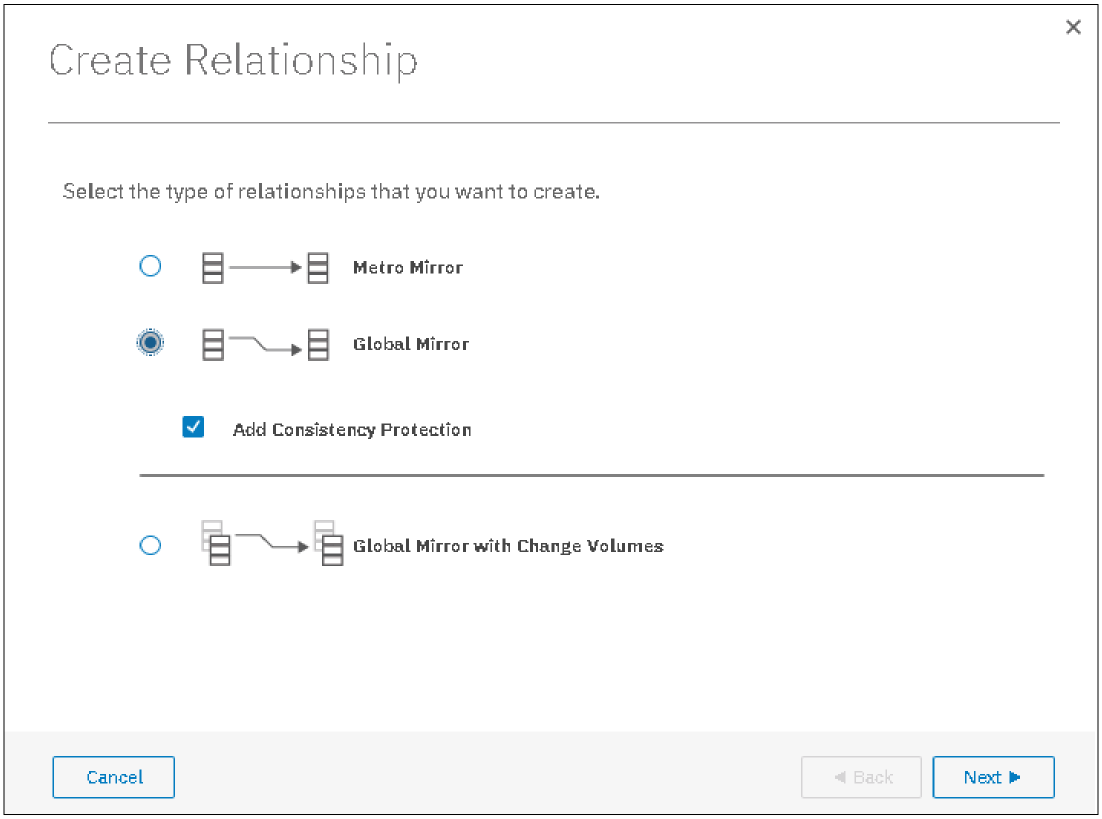


Figure 11-117 Creating a remote copy relationship

4. In the next window, select the master and auxiliary volumes, as shown in Figure 11-118, and click **ADD**.

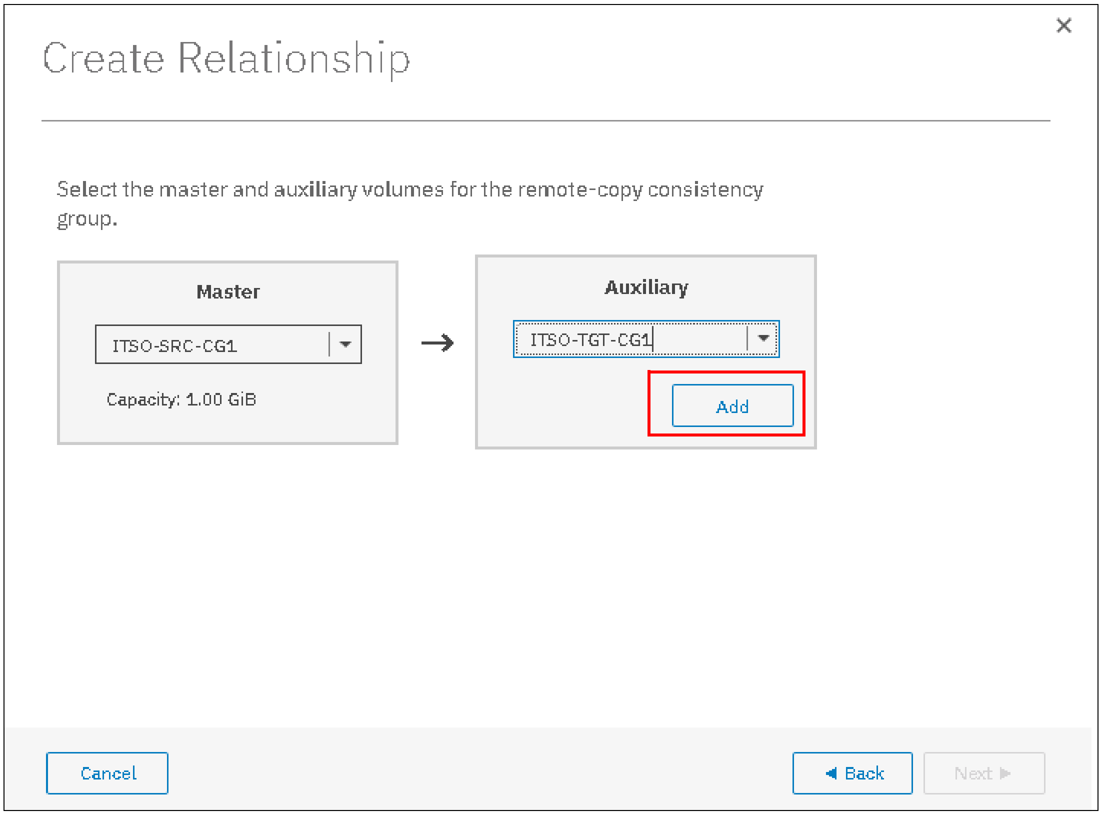


Figure 11-118 Selecting the master and auxiliary volumes

Important: The master and auxiliary volumes must be of equal size. Therefore, only the targets with the appropriate size are shown in the list for a specific source volume.

5. In the next window, you can add change volumes, as shown in Figure 11-119. Click **Finish**.

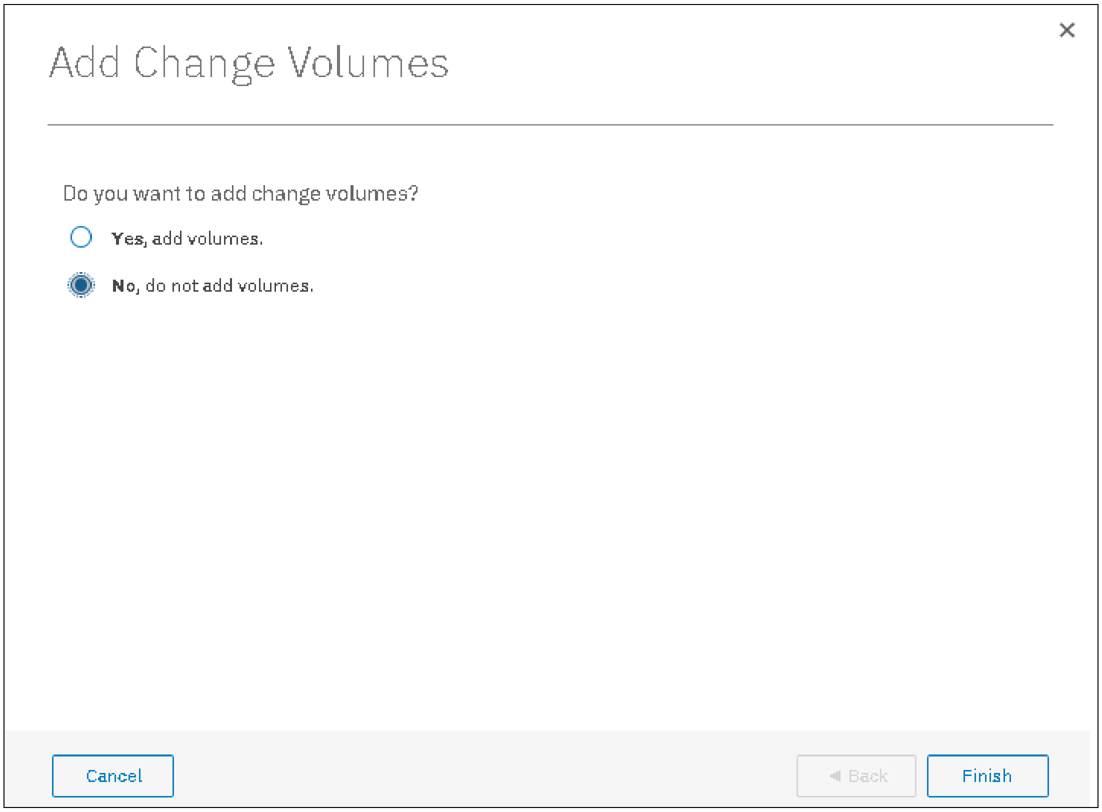


Figure 11-119 Add Change Volumes

6. In the next window, check the relationship created and if you want to add more relationships at the same Consistency Group add the new relationships, as shown in Figure 11-120. Then, click **Next**.

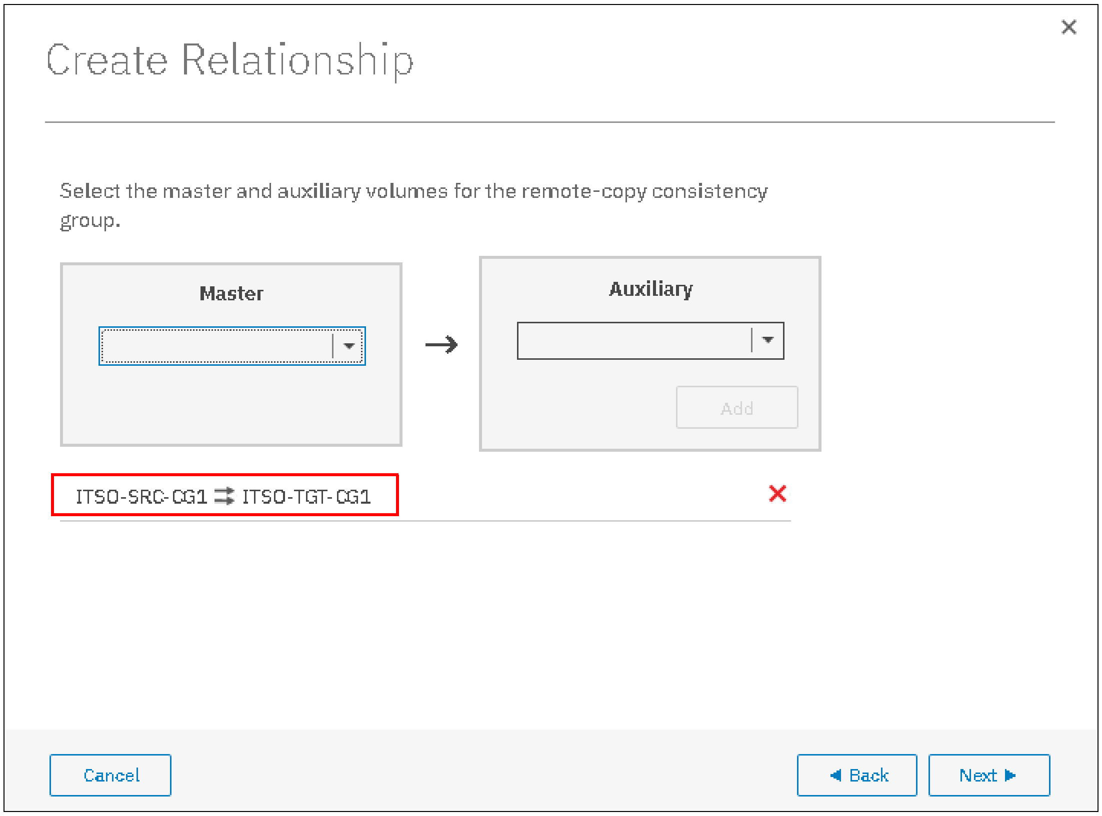


Figure 11-120 Checking and adding the relationship

7. In the next window, select if the volumes are synchronized or not, so the relationship will be created, as shown in Figure 11-121. Click **Finish**.

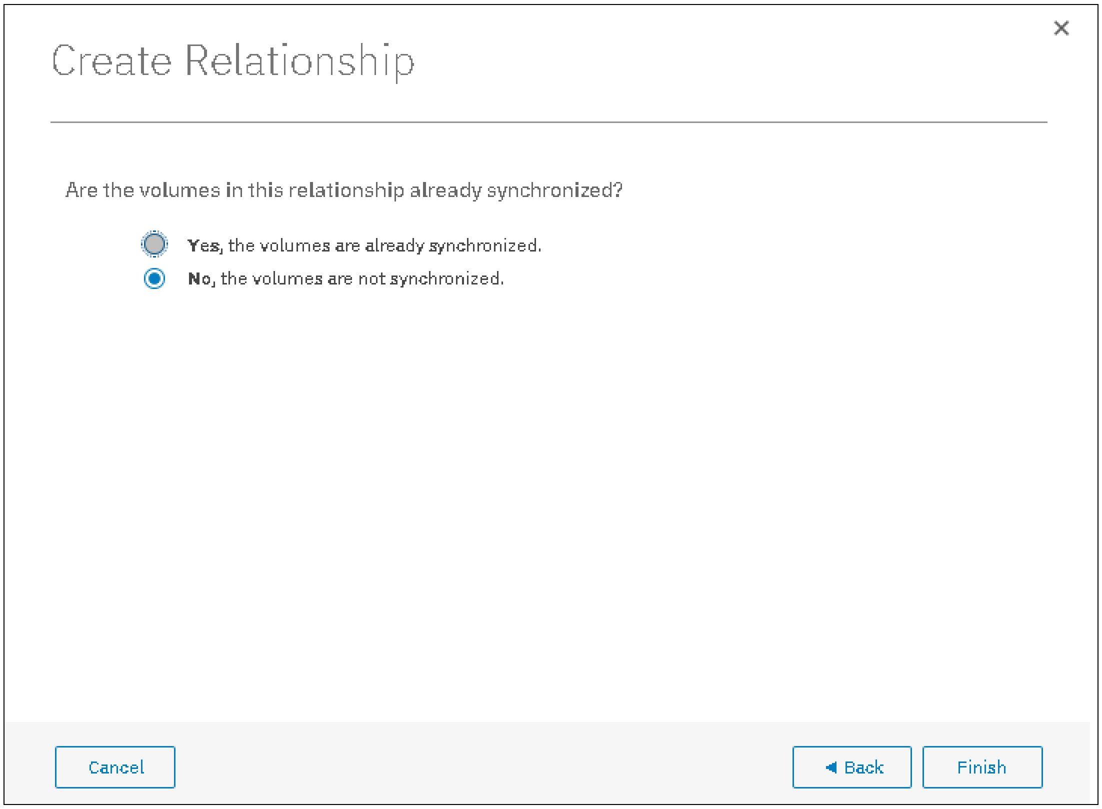


Figure 11-121 Selecting if volumes are synchronized

Note: If the volumes are not synchronized, the initial copy copies the entire source volume to the remote target volume. If you suspect volumes are different or if you have a doubt, synchronize them to ensure consistency on both sides of the relationship.

11.9.3 Creating Consistency Group

To create a Consistency Group, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel and click **Create Consistency Group**, as shown in Figure 11-122.

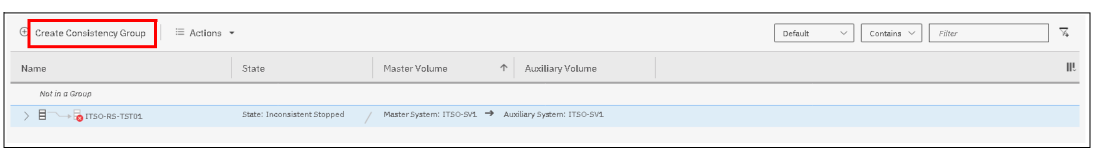


Figure 11-122 Creating a Remote Copy Consistency Group

2. Enter a name for the Consistency Group and click **Next**, as shown in Figure 11-123.

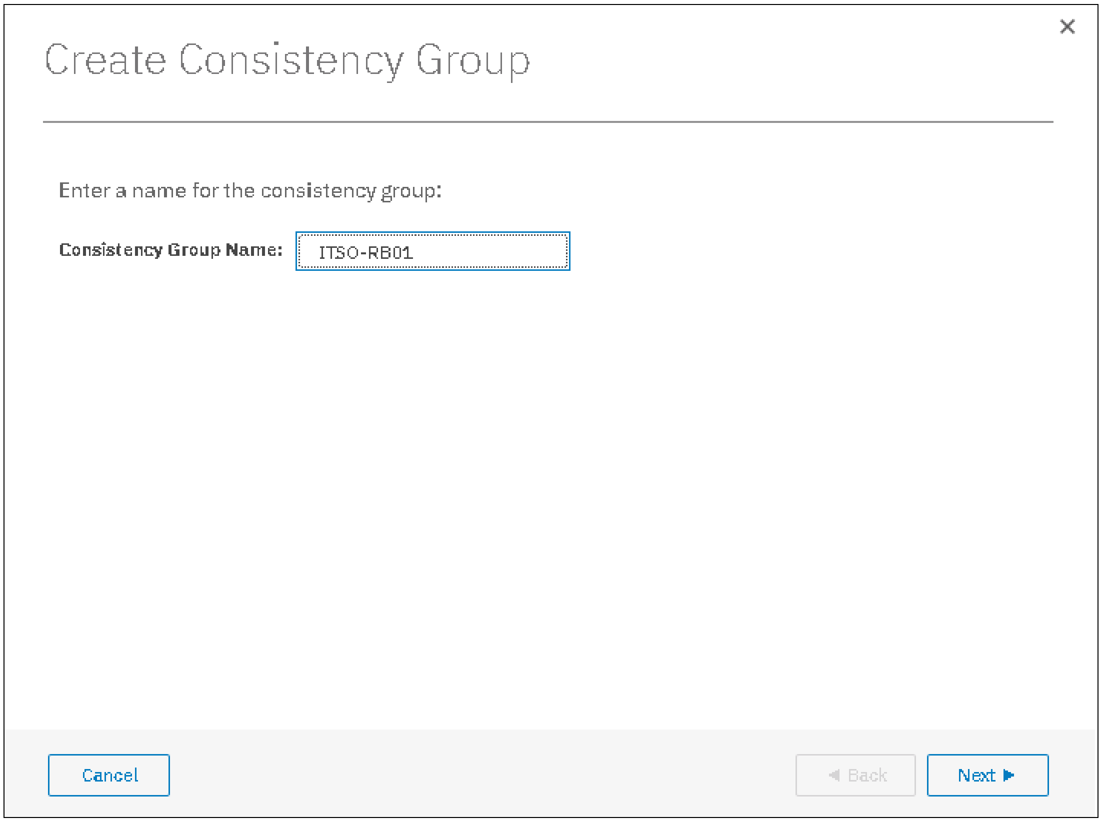


Figure 11-123 Entering a name for the new Consistency Group

3. In the next window, select the location of the auxiliary volumes in the Consistency Group, as shown in Figure 11-124, and click **Next**.
   – **On this system**, which means that the volumes are local.
   – **On another system**, which means that you select the remote system from the menu.

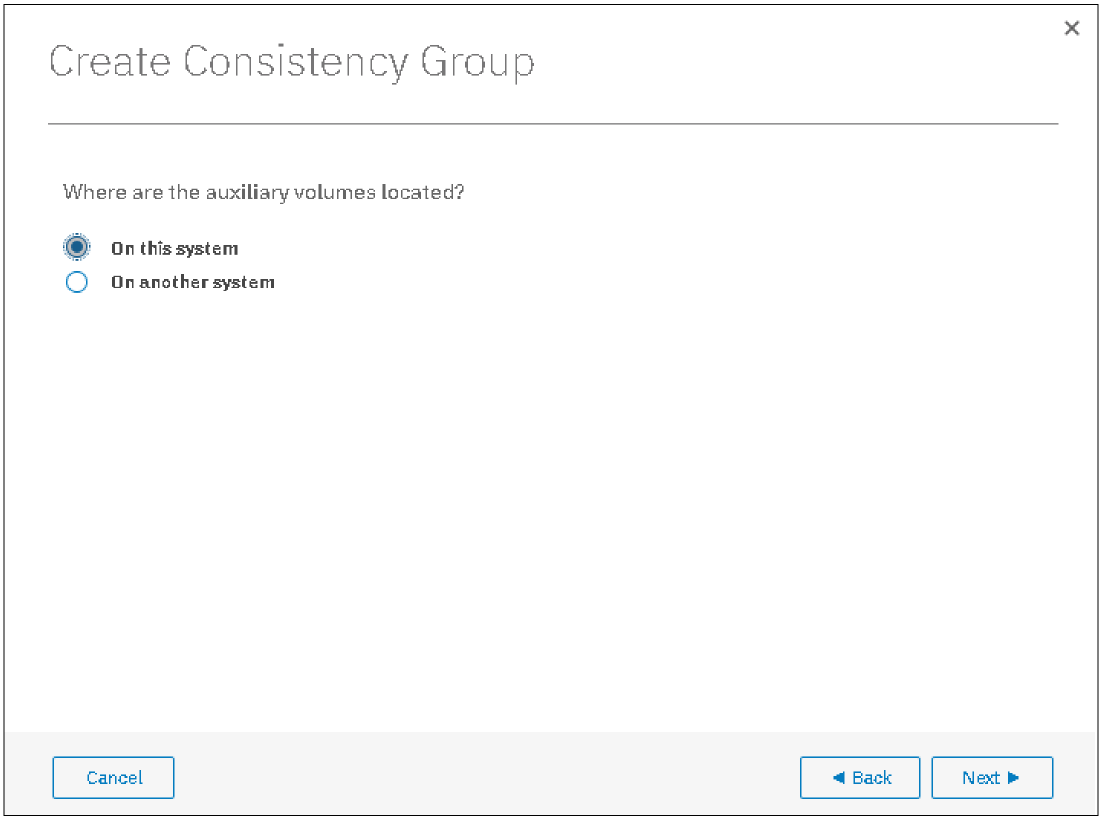


Figure 11-124 Selecting the system to create the Consistency Group with

4. Select whether you want to add relationships to this group, as shown in Figure 11-125. The following options are available:
   - If you select **No**, click **Finish** to create an empty Consistency Group that can be used later.
   - If you select Yes, click **Next** to continue the wizard and continue with the next steps.

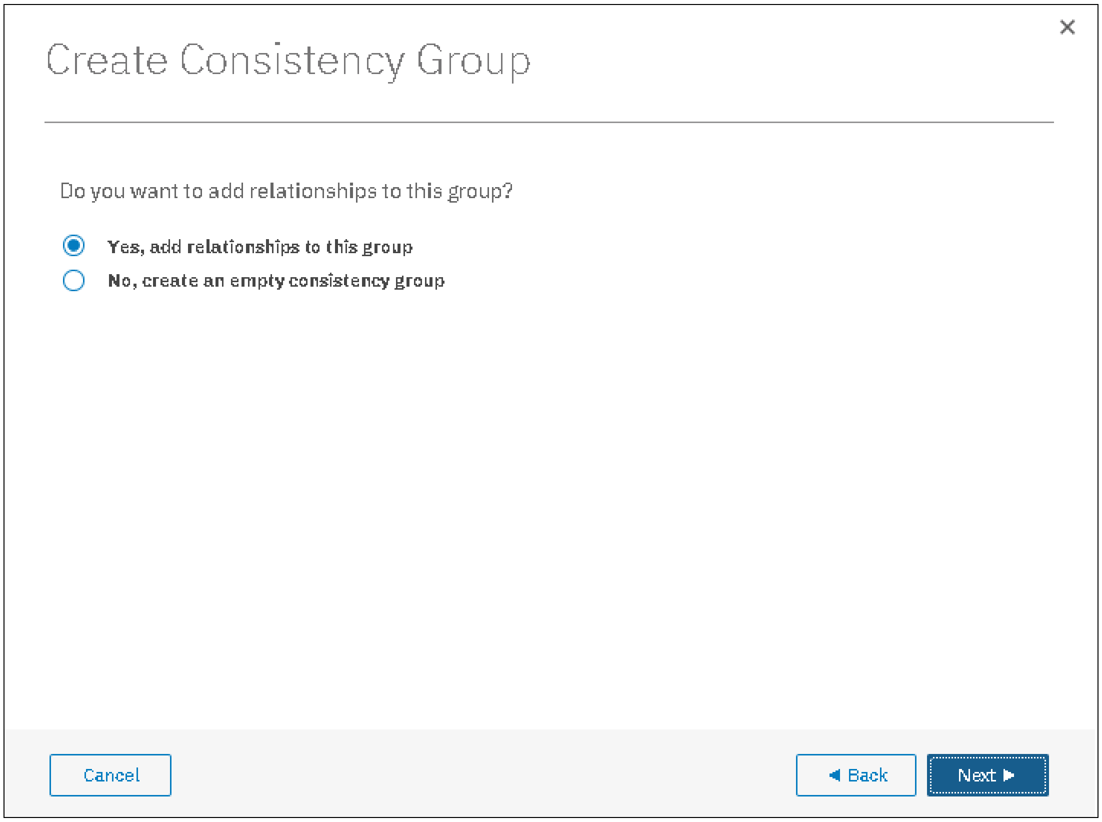


Figure 11-125 Selecting whether relationships should be added to the new Consistency Group

Select one of the following types of relationships that you want to create or add, as shown in Figure 11-126, and click **Next**:

– Metro Mirror
– Global Mirror (with or without Consistency Protection)
– Global Mirror with Change Volumes

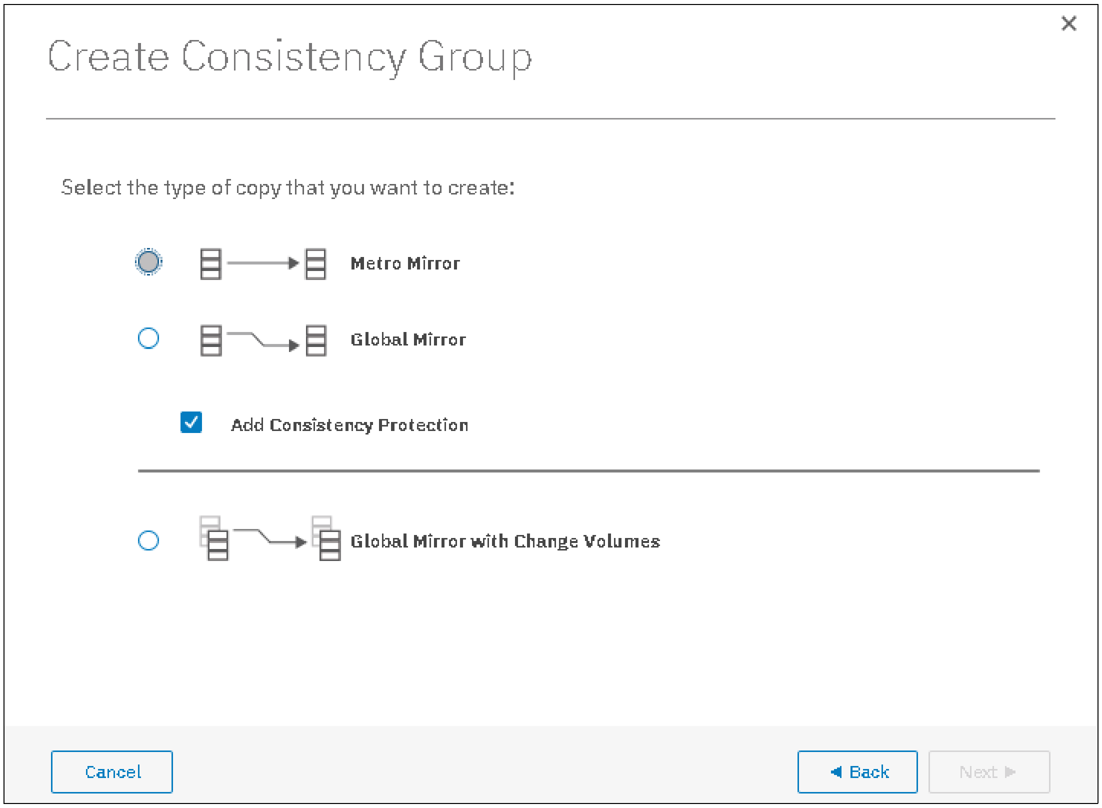


Figure 11-126 Selecting the type of emote copy relationships to create/add

5. As shown in Figure 11-127, you can optionally select existing relationships to add to the group. Click **Next**.

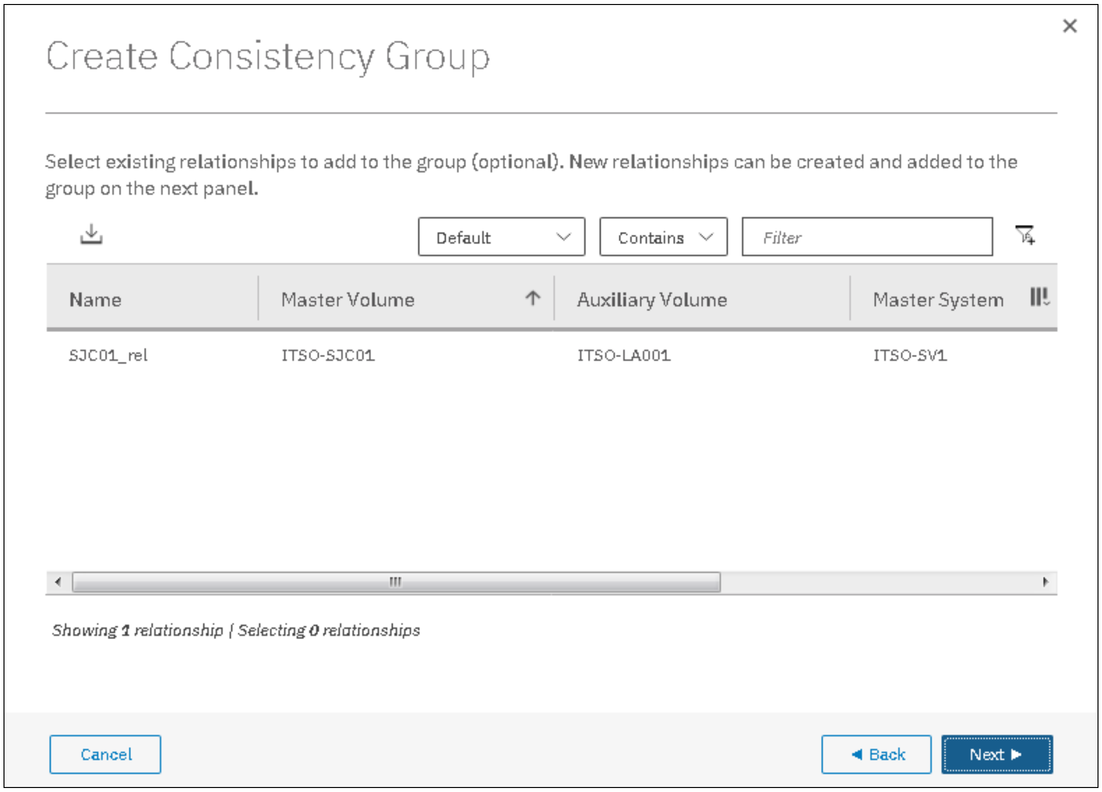


Figure 11-127 Adding Remote Copy relationships to the new Consistency Group

> **Note:** Only relationships of the type that was previously selected are listed.

6. In the next window, you can create relationships between master volumes and auxiliary volumes to be added to the Consistency Group that is being created, as shown in Figure 11-128. Click **Add** when both volumes are selected. You can add multiple relationships in this step by repeating the selection.

   When all the relationships you need are created, click **Next**.

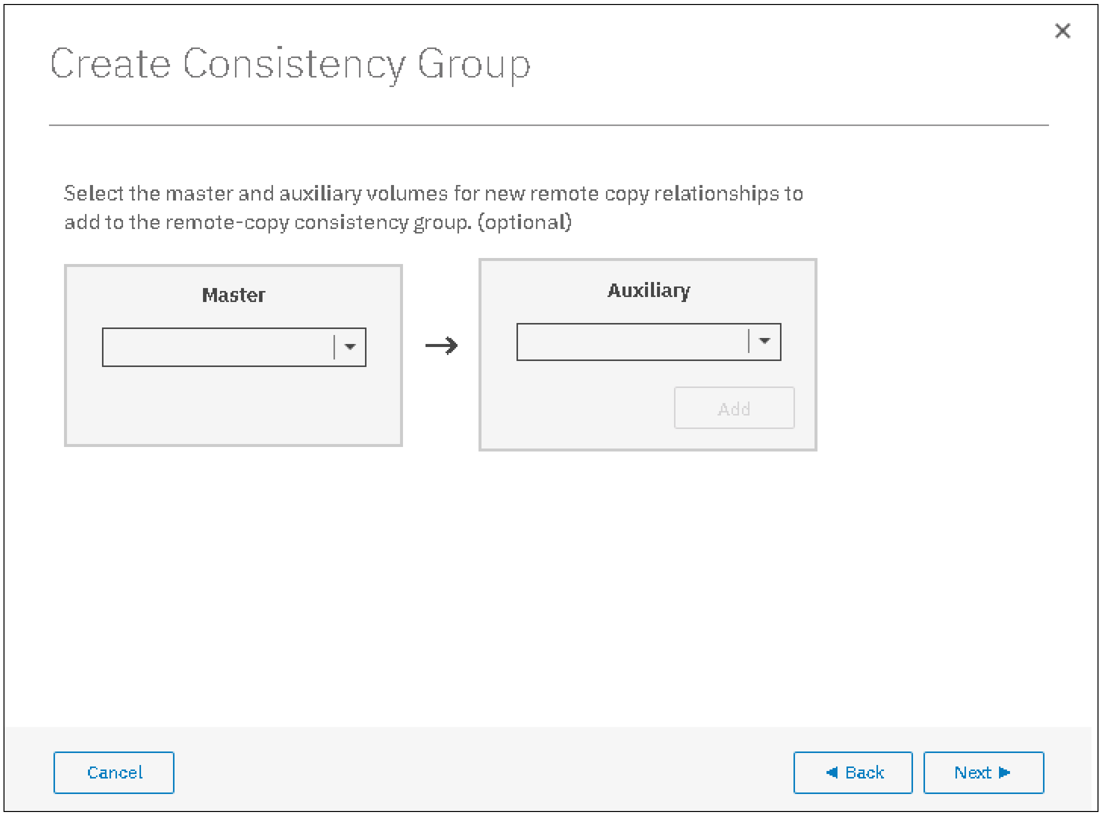


Figure 11-128 Creating new relationships for the new Consistency Group

> **Important:** The master and auxiliary volumes must be of equal size. Therefore, only the targets with the appropriate size are shown in the list for a specific source volume.

7. Specify whether the volumes in the Consistency Group are synchronized, as shown in Figure 11-120 on page 577, and click **Next**.

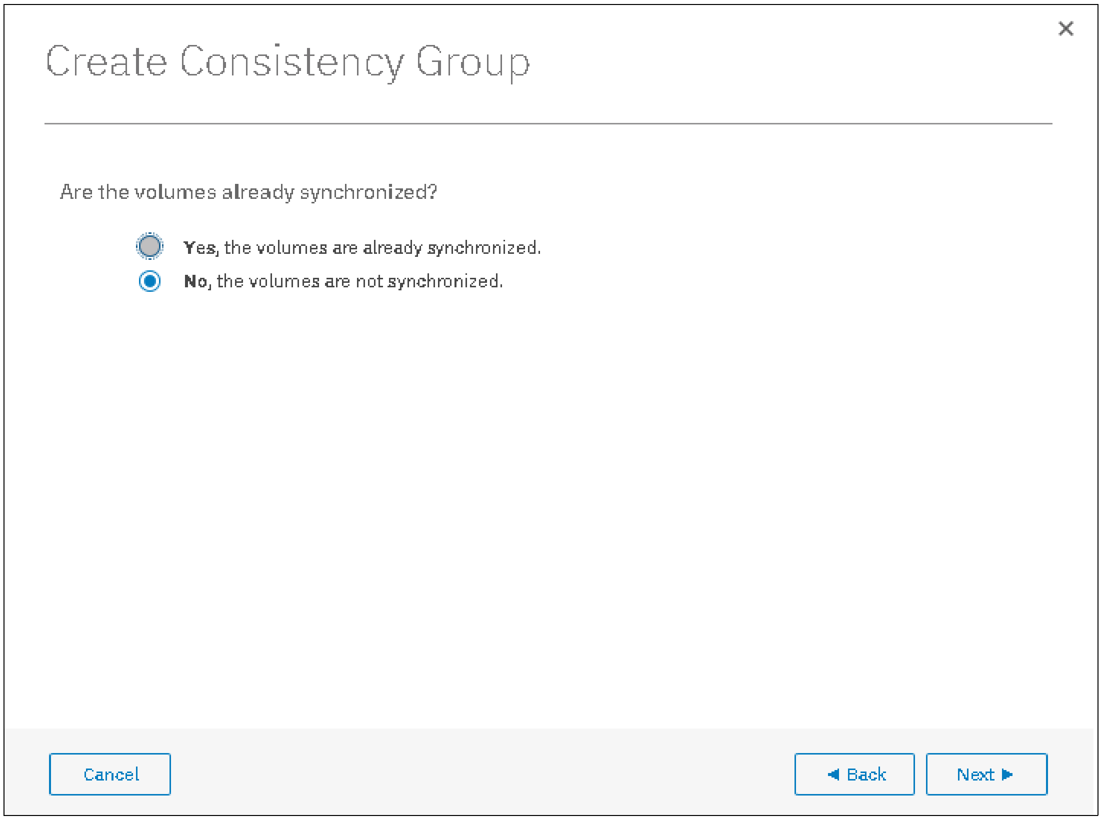


Figure 11-129 Selecting if volumes in the new Consistency Group are already synchronized or not

> **Note:** If the volumes are not synchronized, the initial copy will copy the entire source volume to the remote target volume. If you suspect volumes are different or if you have a doubt, then synchronize them to ensure consistency on both sides of the relationship.

8. In the last window, select whether you want to start the copy of the Consistency Group or not, as shown in Figure 11-130. Click **Finish**.

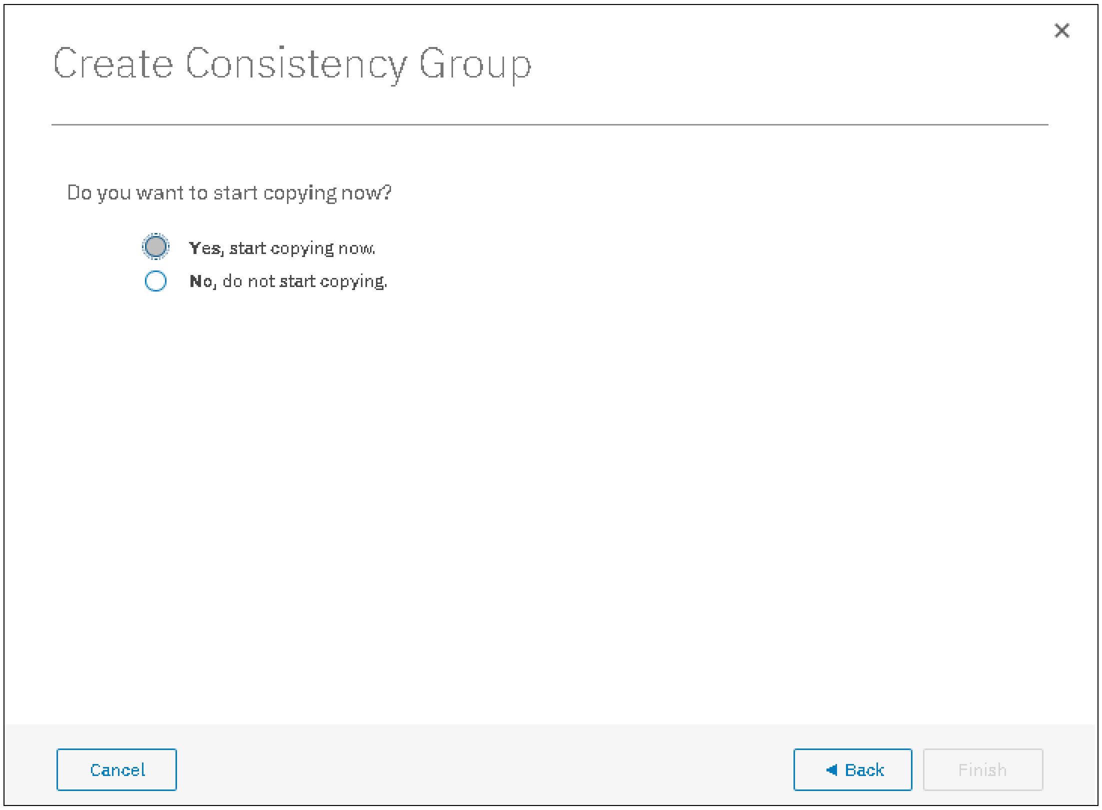


Figure 11-130 Selecting whether copy should start or not

11.9.4 Renaming remote copy relationships

To rename one or multiple remote copy relationships, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.
2. Right-click the relationships to be renamed and select **Rename**, as shown in Figure 11-131.

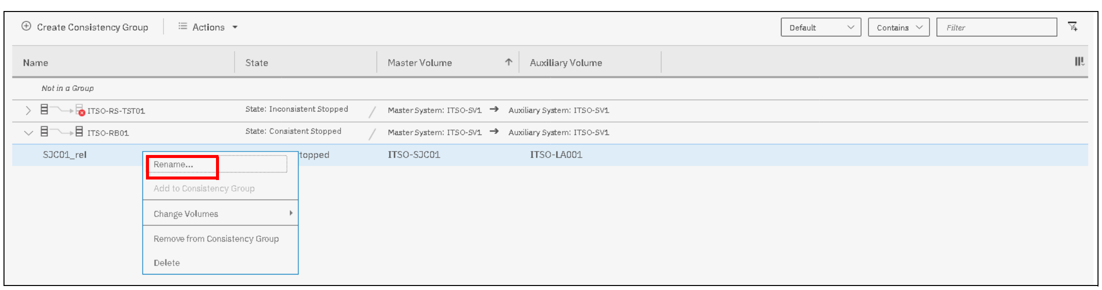


Figure 11-131 Renaming Remote Copy relationships

3. Enter the new name that you want to assign to the relationships and click **Rename**, as shown in Figure 11-132.

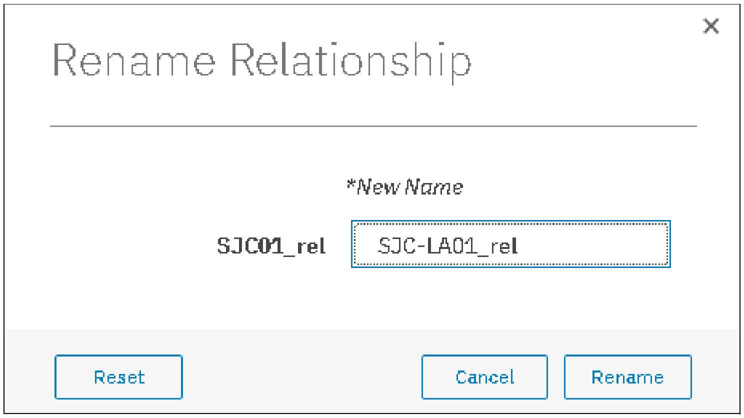


Figure 11-132 Renaming Remote Copy relationships

> **Remote copy relationship name:** You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The remote copy name can be 1 - 15 characters. No blanks are allowed.

11.9.5 Renaming a remote copy consistency group

To rename a remote copy consistency group, complete the following steps:

1. Open the **Copy Services** → → **Remote Copy** panel.
2. Right-click the consistency group to be renamed and select **Rename**, as shown in Figure 11-133.

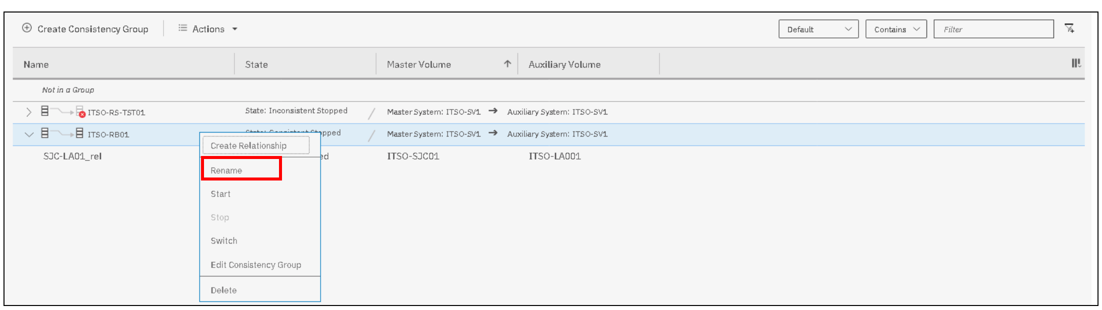


Figure 11-133 Renaming a Remote Copy Consistency Group

3. Enter the new name that you want to assign to the Consistency Group and click **Rename**, as shown in Figure 11-134.

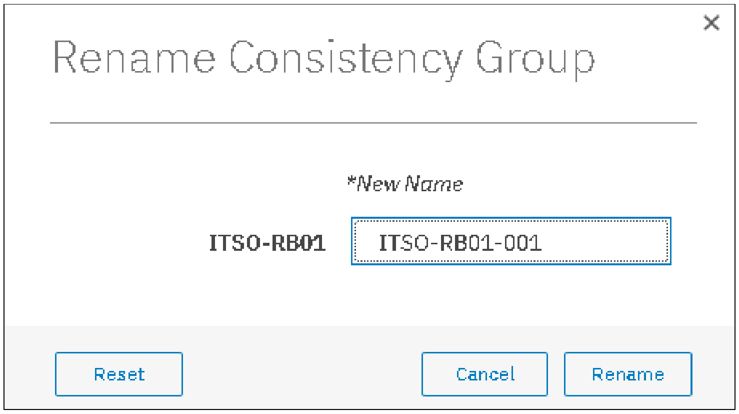


Figure 11-134 Entering new name for Consistency Group

Remote copy consistency group name: You can use the letters A - Z and a - z, the numbers 0 - 9, and the underscore (_) character. The remote copy name can be 1 - 15 characters. No blanks are allowed.

11.9.6 Moving stand-alone remote copy relationships to Consistency Group

To add one or multiple stand-alone relationships to a remote copy consistency group, complete the following steps:

1. Open the **Copy Services** → → **Remote Copy** panel.
2. Right-click the relationships to be moved and select **Add to Consistency Group**, as shown in Figure 11-135.

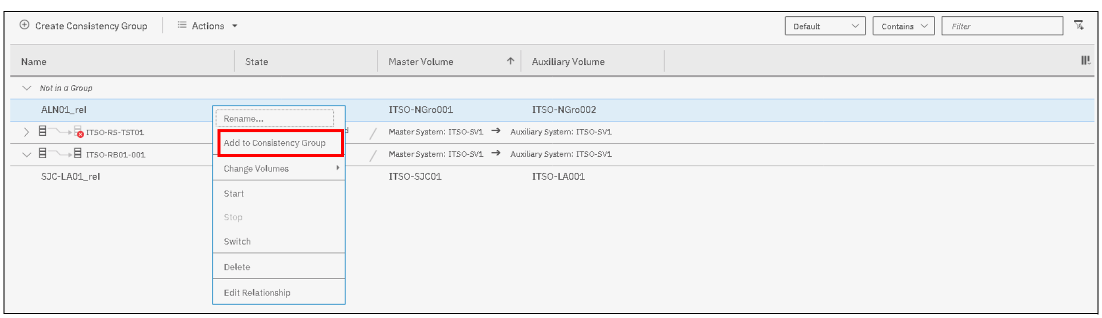


Figure 11-135 Adding relationships to a Consistency Group

3. Select the Consistency Group for this remote copy relationship by using the menu, as shown in Figure 11-136. Click **Add to Consistency Group** to confirm your changes.

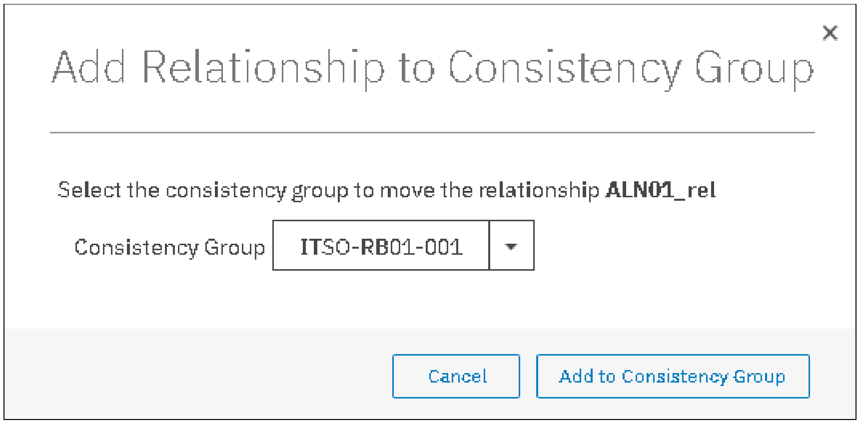


Figure 11-136 Selecting the Consistency Group to add the relationships to

11.9.7 Removing remote copy relationships from Consistency Group

To remove one or multiple relationships from a remote copy consistency group, complete the following steps:

1. Open the **Copy Services** → → **Remote Copy** panel.
2. Right-click the relationships to be removed and select **Remove from Consistency Group**, as shown in Figure 11-137.

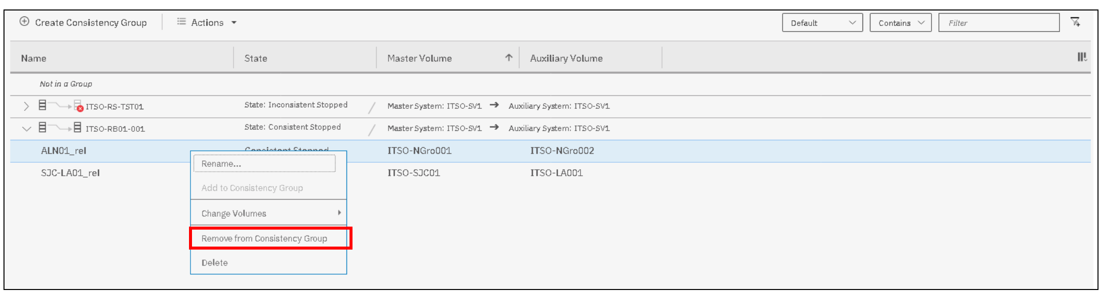


Figure 11-137 Removing relationships from a Consistency Group

3. Confirm your selection and click **Remove** as shown in Figure 11-138.

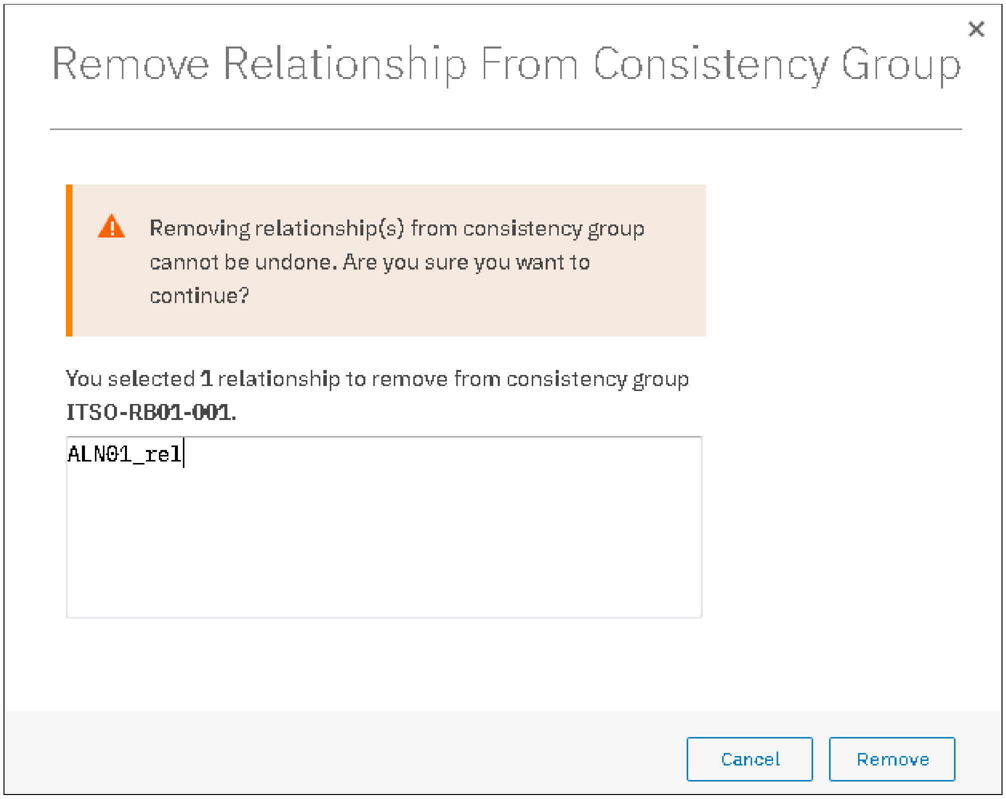


Figure 11-138 Confirm the removal of relationships from a Consistency Group

11.9.8 Starting remote copy relationships

When a remote copy relationship is created, the remote copy process can be started. Only relationships that are not members of a Consistency Group, or the only relationship in a Consistency Group, can be started. In any other case, consider starting the Consistency Group instead.

To start one or multiple relationships, complete the following steps:

1. Open the **Copy Services** → → **Remote Copy** panel.
2. Right-click the relationships to be started and select **Start**, as shown in Figure 11-139.

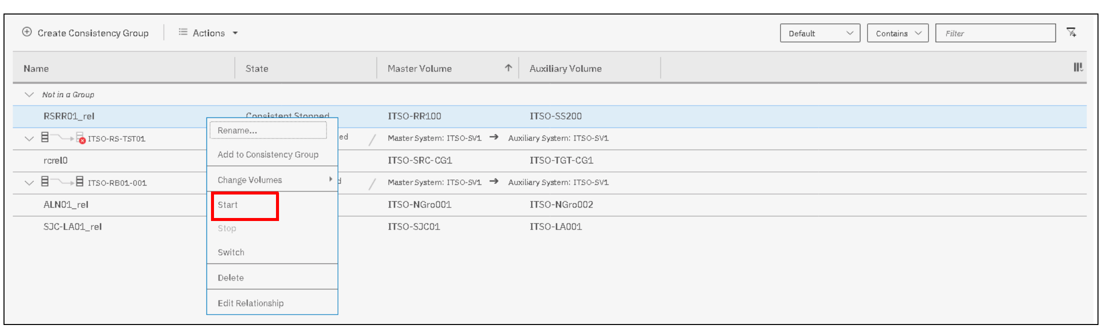


Figure 11-139 Starting remote copy relationships

11.9.9 Starting a remote copy Consistency Group

When a remote copy consistency group is created, the remote copy process can be started, for all the relationships that are part of the consistency groups.

To start a consistency group, open the **Copy Services** → **Remote Copy** panel, right-click the consistency group to be started, and select **Start**, as shown in Figure 11-140.

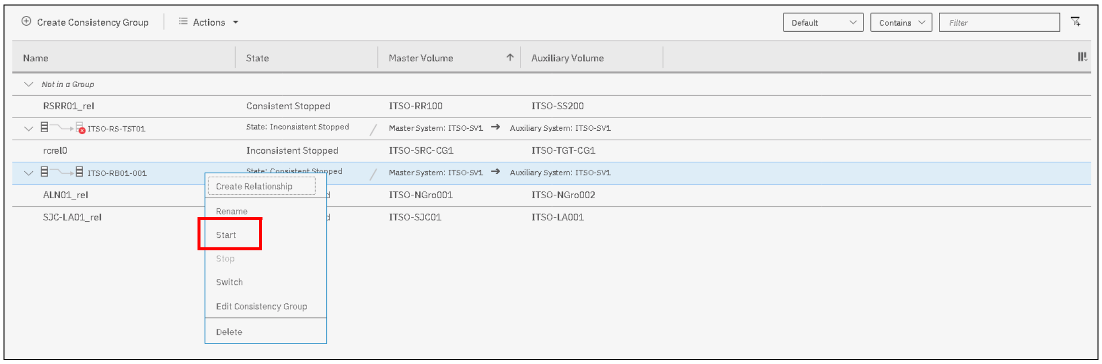


Figure 11-140   Starting a remote copy Consistency Group

11.9.10 Switching a relationship copy direction

When a remote copy relationship is in the Consistent synchronized state, the copy direction for the relationship can be changed. Only relationships that are not a member of a Consistency Group, or the only relationship in a Consistency Group, can be switched. In any other case, consider switching the Consistency Group instead.

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 inhibited 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 remote copy relationship, complete the following steps:

1. Open the **Copy Services** → → **Remote Copy** panel.
2. Right-click the relationship to be switched and select **Switch**, as shown in Figure 11-141.

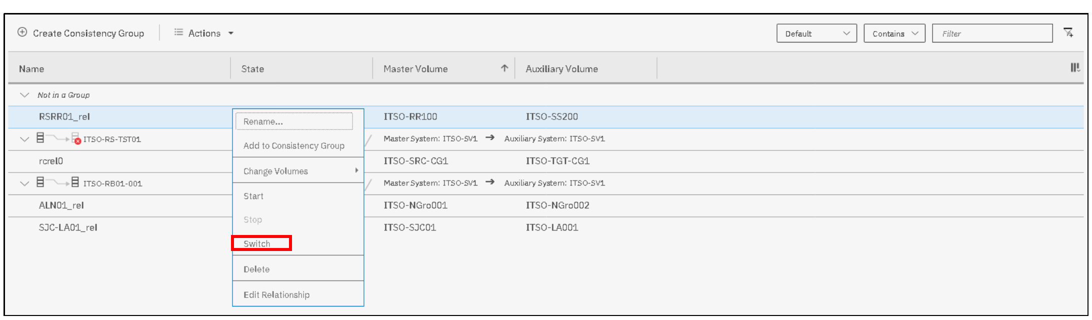


Figure 11-141   Switching remote copy relationship direction

3. Because the master-auxiliary relationship direction is reversed, write access is disabled on the new auxiliary volume (former master volume), whereas it is enabled on the new master volume (former auxiliary volume). A warning message is displayed, as shown in Figure 11-142. Click **Yes**.

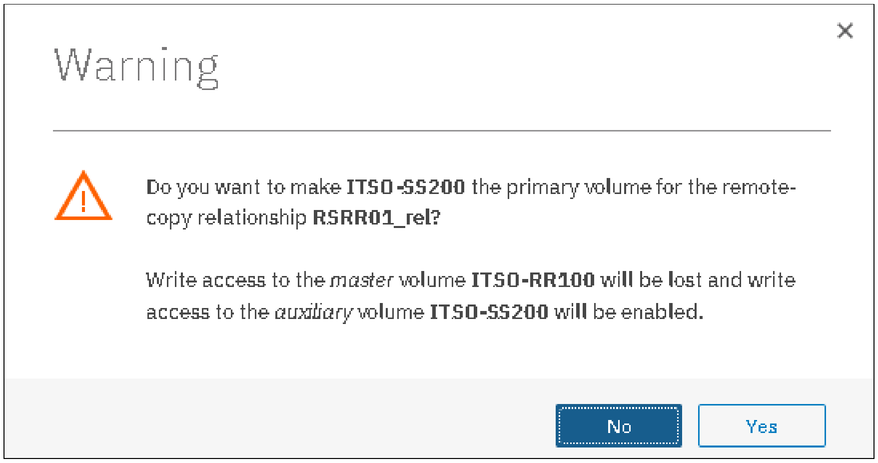


Figure 11-142 Switching master-auxiliary direction of relationships changes the write access

4. When a remote copy relationship is switched, an icon is displayed in the Remote Copy panel list, as shown in Figure 11-143.

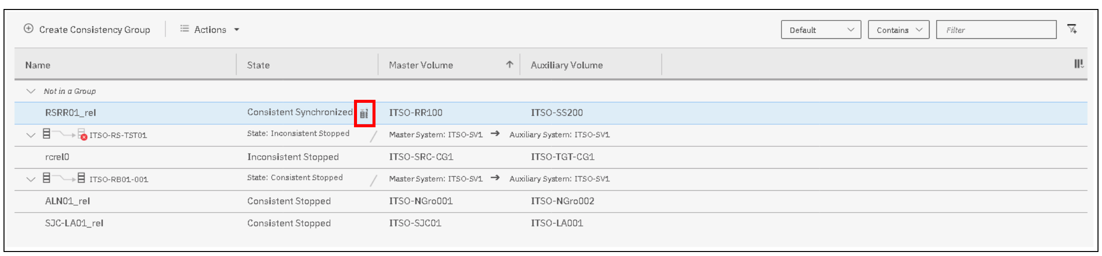


Figure 11-143 Switched Remote Copy Relationship

11.9.11 Switching a Consistency Group direction

When a remote copy 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 of the I/O is inhibited 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 remote copy consistency group, complete the following steps:

1. Open the **Copy Services** → → **Remote Copy** panel.
2. Right-click the consistency group to be switched and select **Switch**, as shown in Figure 11-144.

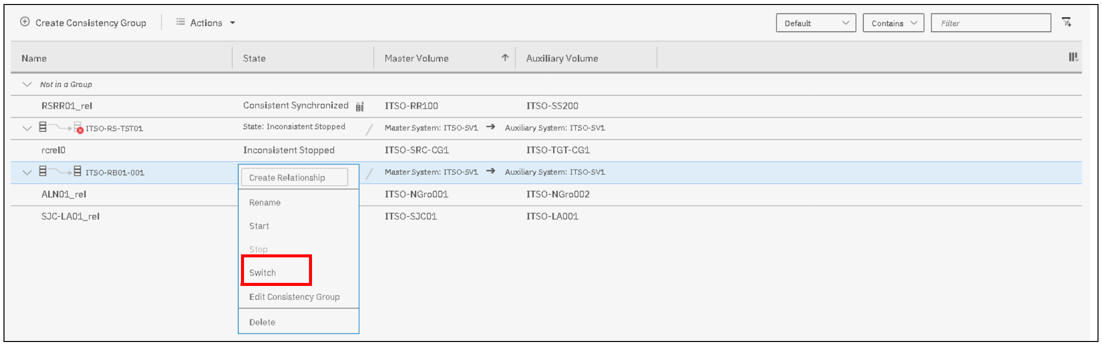


Figure 11-144 Switching a Consistency Group direction

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 11-145. Click **Yes**.

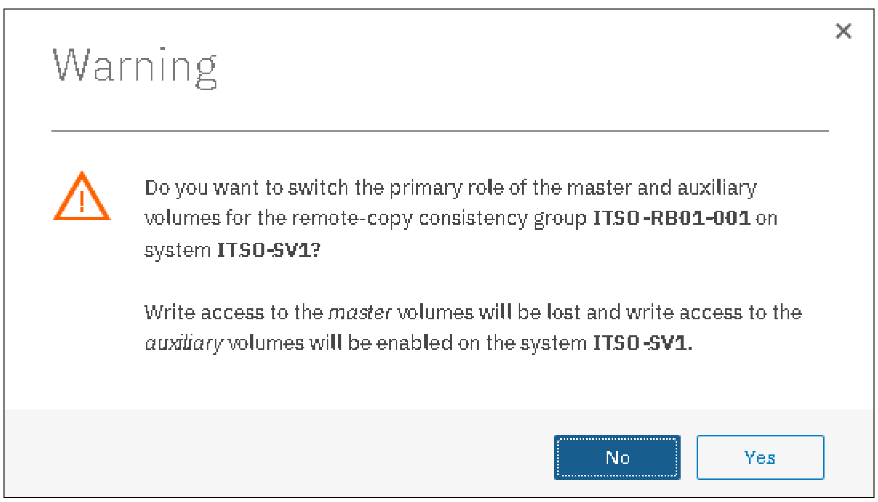


Figure 11-145 Switching direction of Consistency Groups changes the write access

11.9.12 Stopping remote copy relationships

When a remote copy relationship is created and started, the remote copy process can be stopped. Only relationships that are not members of a Consistency Group, or the only relationship in a Consistency Group, can be stopped. In any other case, consider stopping the Consistency Group instead.

To stop one or multiple relationships, complete the following steps:

1. Open the **Copy Services** → → **Remote Copy** panel.

2. Right-click the relationships to be stopped and select **Stop**, as shown in Figure 11-146.

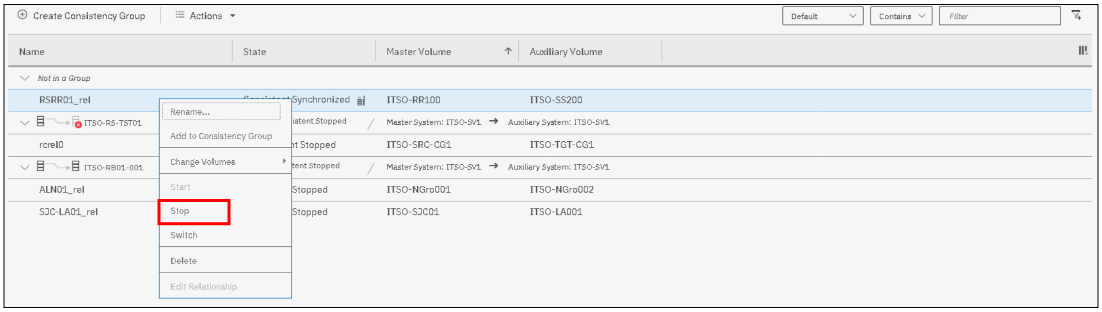


Figure 11-146 Stopping a Remote Copy relationship

3. When a remote copy relationship is stopped, access to the auxiliary volume can be changed so it can be read and written by a host. A confirmation message is displayed, as shown in Figure 11-147.

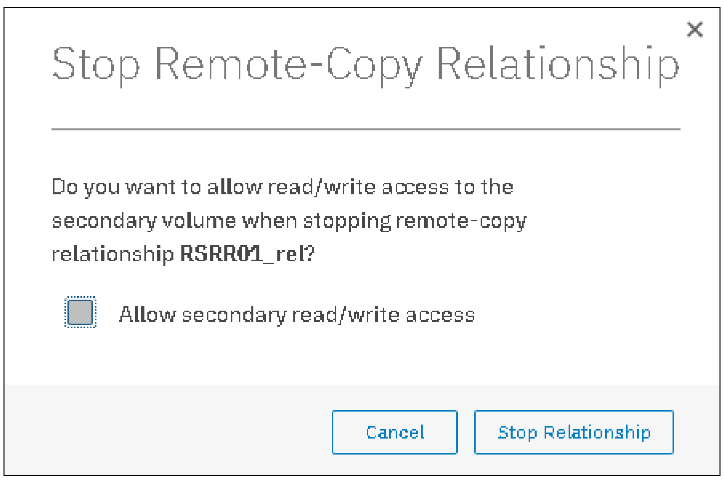


Figure 11-147 Grant access in read and write to the auxiliary volume

11.9.13 Stopping a Consistency Group

When a remote copy consistency group is created and started, the remote copy process can be stopped.

To stop a consistency group, complete the following steps:

1. Open the **Copy Services** → **Remote Copy** panel.
2. Right-click the consistency group to be stopped and select **Stop**, as shown in Figure 11-148.

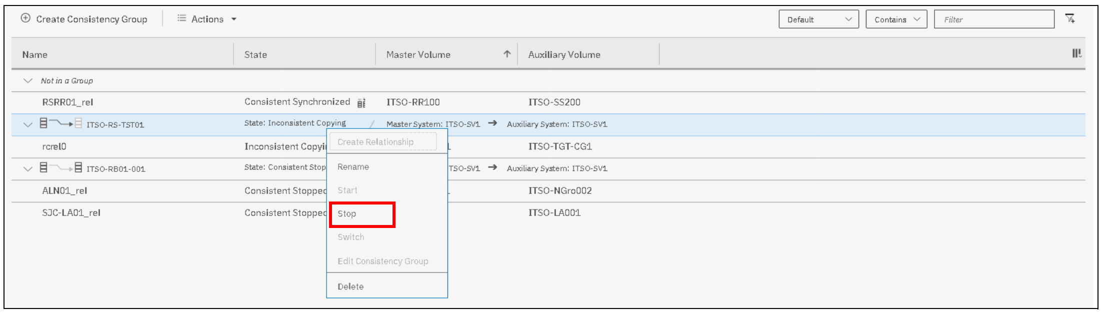


Figure 11-148 Stopping a Consistency Group

3. When a remote copy 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 is displayed, as shown in Figure 11-149.

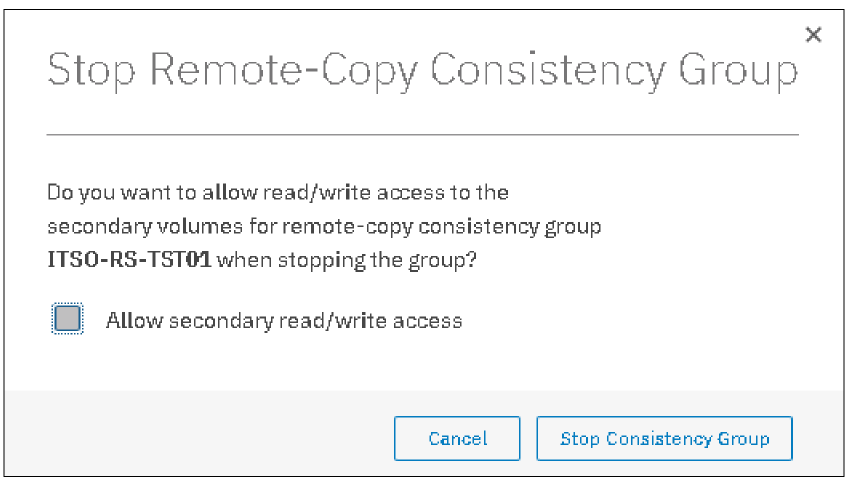


Figure 11-149 Grant access in read and write to the auxiliary volumes

11.9.14 Deleting remote copy relationships

To delete remote copy relationships, complete the following steps:

1. Open the **Copy Services** → → **Remote Copy** panel.

2. Right-click the relationships that you want to delete and select **Delete**, as shown in Figure 11-150.

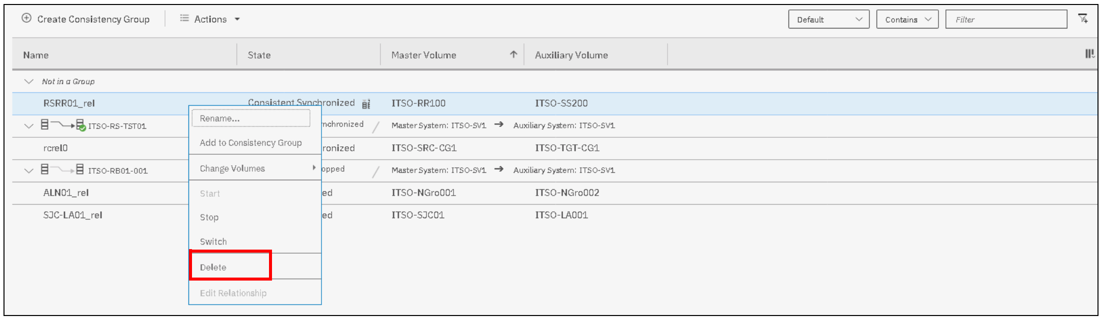


Figure 11-150 Deleting Remote Copy Relationships

3. A confirmation message is displayed, requesting the user to enter the number of relationships to be deleted, as shown in Figure 11-151.

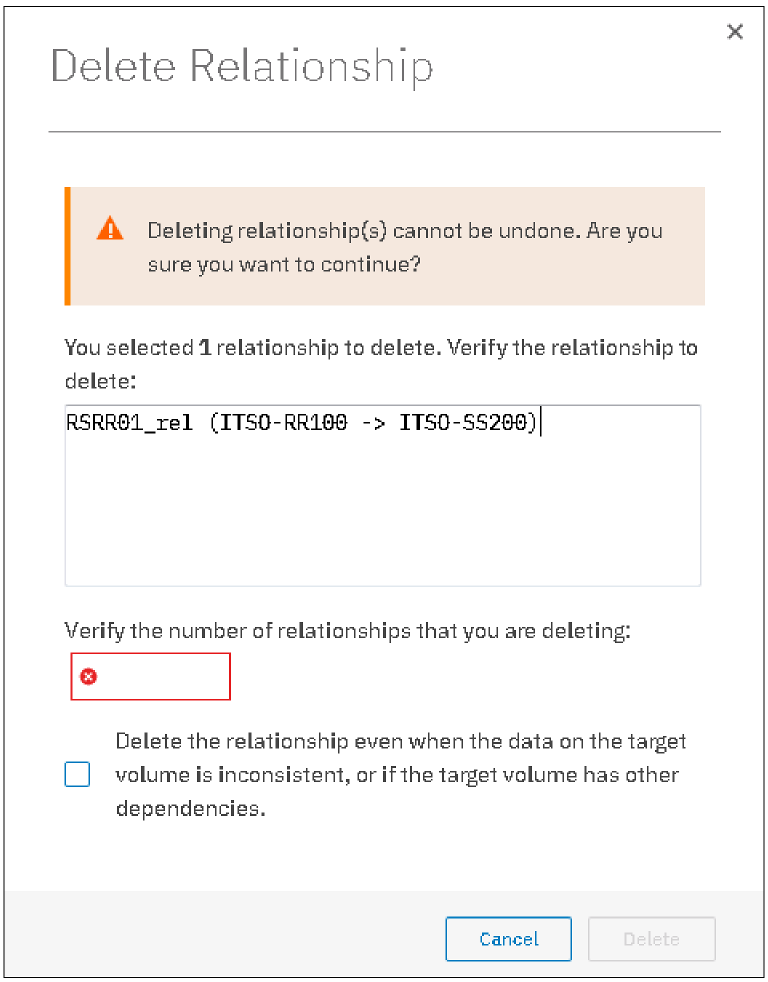


Figure 11-151 Confirmation of relationships deletion

11.9.15 Deleting a Consistency Group

To delete a remote copy consistency group, complete the following steps:

1. Open the **Copy Services** → → **Remote Copy** panel.
2. Right-click the consistency group that you want to delete and select **Delete**, as shown in Figure 11-152.

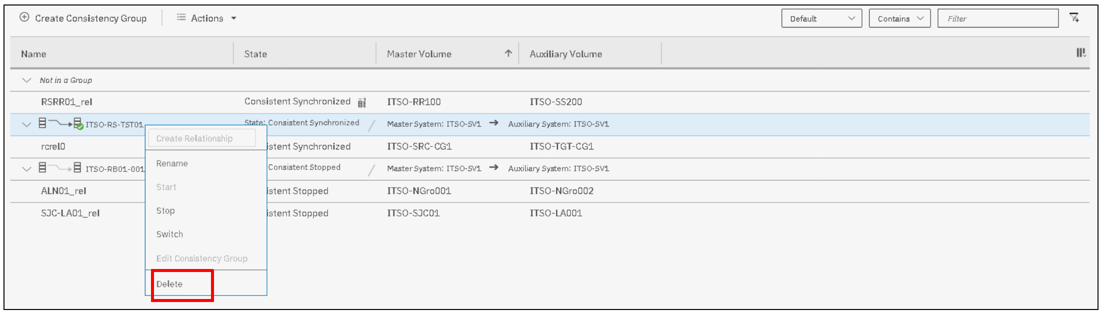


Figure 11-152 Deleting a Consistency Group

3. A confirmation message is displayed, as shown in Figure 11-153. Click **Yes**.

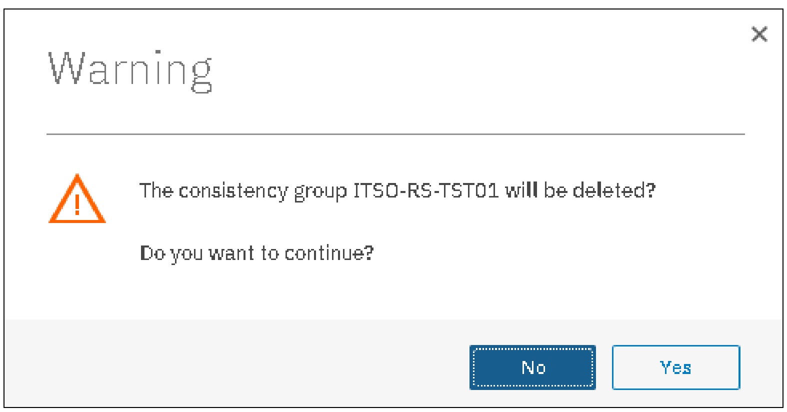


Figure 11-153 Confirmation of a Consistency Group deletion

> **Important:** Deleting a Consistency Group does not delete its remote copy mappings.

11.10 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 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 Remote Copy 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 volume capacity for the specified I/O group. So, for 256 KiB grains size, 2 TiB of total Metro Mirror, Global Mirror, or active-active volume capacity is provided.

See Table 11-14 to calculate the memory requirements and confirm that your system is able to accommodate the total installation size.

Table 11-14 Memory allocation for FlashCopy services

Minimum allocated bitmap space	**Default allocated bitmap space**	**Maximum allocated bitmap space**	**Minimum functionality when using the default values**[1]
0	20 MiB	512 MiB	40 TiB of remote mirroring volume capacity
[1] Remote copy includes Metro Mirror, Global Mirror, and active-active relationships.			

When you configure change volumes for use with Global Mirror, two internal FlashCopy mappings are created for each change volume.

For Metro Mirror, Global Mirror, and HyperSwap active-active relationships, two bitmaps exist. 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 use 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 IBM SAN Volume Controller system by opening the **Settings** → **System** panel and clicking the **Resources** menu, as shown in Figure 11-154.

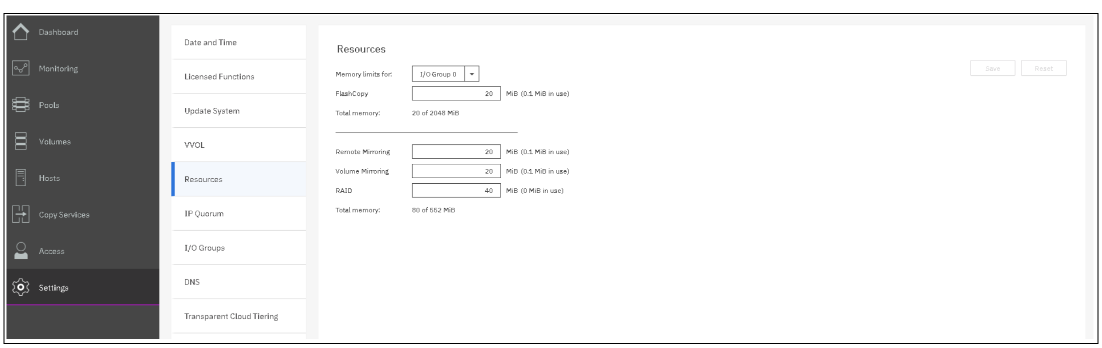


Figure 11-154 Modifying resources allocation

11.11 Troubleshooting remote copy

Remote copy (Metro Mirror and Global Mirror) has 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 will cause 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.

11.11.1 1920 error

A `1920` error is deliberately generated by the system and should be considered as a control mechanism. It occurs after `985003` (“Unable to find path to disk in the remote cluster (system) within the time-out 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. Global Mirror has a round-trip-time 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 milliseconds round trip. That is, it must have up to 40 or 125 milliseconds latency each way.

The primary component of your round-trip time 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 employs 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 Global Mirror 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 Global Mirror round-trip-time (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 existing 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.

In Figure 11-155, you can see why the 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.

Packet Size	Link Size	Serialization Delay (Time Required to Send Data)	Unit
64	256 Kbps	2.0E+03	microseconds
64	1.5 Mbps	3.4E+02	microseconds
64	100 Mbps	5.1E+00	microseconds
64	155 Mbps	3.3E+00	microseconds
64	622 Mbps	3.2E-01	microseconds
64	1 Gbps	5.1E-04	microseconds
64	10 Gbps	5.1E-05	microseconds
1500	256 Kbps	4.7E+04	microseconds
1500	1.5 Mbps	8.0E+03	microseconds
1500	100 Mbps	1.2E+02	microseconds
1500	155 Mbps	7.7E+01	microseconds
1500	622 Mbps	1.9E+01	microseconds
1500	1 Gbps	1.2E+01	microseconds
1500	10 Gbps	1.2E+00	microseconds
2148	256 Kbps	6.7E+04	microseconds
2148	1.5 Mbps	1.1E+04	microseconds
2148	100 Mbps	1.7E+02	microseconds
2148	155 Mbps	1.1E+02	microseconds
2148	622 Mbps	2.8E+01	microseconds
2148	1 Gbps	1.7E+01	microseconds
2148	10 Gbps	1.7E-03	microseconds

Figure 11-155 Effect of packet size (in bytes) versus the link size

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.

Figure 11-155 shows the orders of magnitude of difference between the link bandwidths. It is easy to see how `1920` errors can arise when your bandwidth is insufficient. Never use a TCP/IP ping to measure RTT for FCIP traffic.

In Figure 11-155 on page 600, 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 maximum transmission unit (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 could impact the host's response time. The system has no indication on if/when the relationship can be restarted. Therefore, the relationship cannot be restarted automatically and it needs to be done manually.

11.11.2 1720 error

The `1720` error (event ID `050020`) is the other problem remote copy 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 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 needs to 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 SAN Volume Controller node ports where each node port comes from a different IBM SAN Volume Controller 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 with IBM Spectrum Control and comparing against your sample interval reveals potential SAN congestion. If a zero buffer credit timer is more than 2% of the total time of the sample interval, it might cause problems.

Next, 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 could 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 the two IBM Storwize 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 the MM/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 remote copy problems, contact your IBM Support representative for a more complete analysis.

12

Encryption

Encryption protects against the potential exposure of sensitive user data that is stored on discarded, lost, or stolen storage devices. Storwize V7000 supports optional encryption of data at-rest.

This chapter includes the following topics:

12.1 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 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 the storage, the organization defines its needs and, if it is decided that data-at-rest encryption is required, 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 need to 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?

12.2 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 (FIPS-140) standard, but is not certified.
- Ciphertext stealing XTS-AES-256 is used for data encryption.
- AES 256 is used for master access keys.
- 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.

12.2.1 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 is used for encryption is chosen automatically by the system based on the placement of the data:

- Hardware encryption: Data is encrypted by using serial-attached SCSI (SAS) hardware. Used only for internal storage (drives).
- Software encryption: Data is encrypted by using nodes' CPU (encryption code uses AES-NI CPU instruction set). Used only for external storage.

Note: Software encryption is available in IBM Spectrum Virtualize code 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.

12.2.2 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 SAN only between IBM Spectrum Virtualize systems and external storage. Data is *not* encrypted when transferred on SAN interfaces under the following circumstances:

- Server-to-storage data transfer
- Remote copy (for example, Global or Metro Mirror)
- Intracluster 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 a SAN Volume Controller 2145-DH8 system, which stores hardware-encrypted data on internal disks. The data is mirrored to a remote Storwize V7000 Gen1 system using Remote Copy. The data flowing through the Remote Copy link is not encrypted. Because the Storwize V7000 Gen1 (2076-324) cannot perform any encryption activities, data on the Storwize V7000 Gen1 is not encrypted.

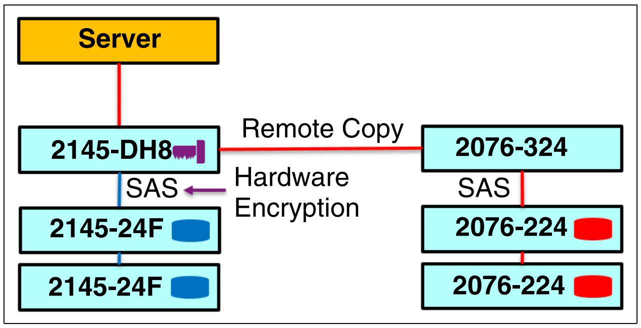


Figure 12-1 Encryption on single site

To enable encryption of both data copies, the Storwize V7000 Gen1 must be replaced by an encryption capable IBM Spectrum Virtualize system, as shown in Figure 12-2. After such replacement, both copies of data are encrypted, but the Remote Copy communication between both sites remains unencrypted.

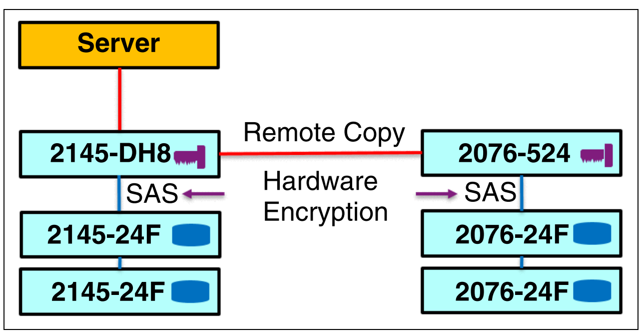


Figure 12-2 Encryption on both sites

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.

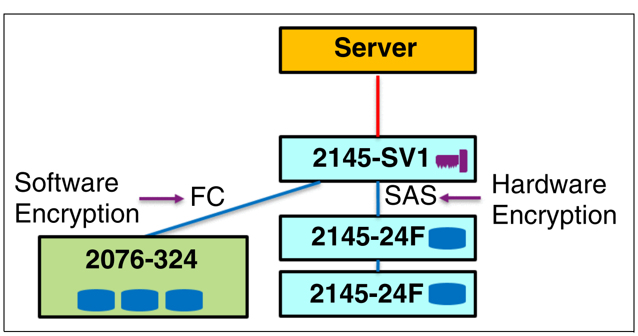


Figure 12-3 Example of software encryption and hardware encryption

Placement of hardware encryption and software encryption in the Storwize code stack are shown in Figure 12-4. The functions that are implemented in software are shown in blue. The external storage system is shown in yellow. The hardware encryption on the SAS chip is marked in pink. As compression is performed before encryption, it is possible to get benefits of compression for the encrypted data.

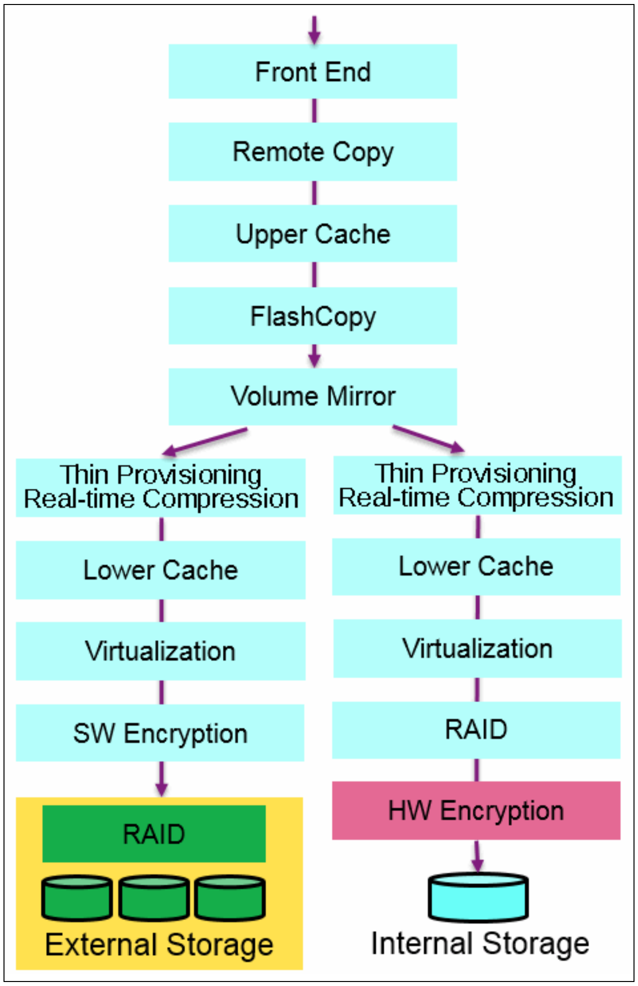


Figure 12-4 Encryption placement in the IBM Spectrum Virtualize software stack

Each volume copy can use different encryption methods (hardware and software). It is also allowed to 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.

12.2.3 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 (MDisk) 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.

12.2.4 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 enabled, the control enclosure must also be licensed.

No trial licenses for encryption exist on the basis that 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.

Contact your IBM marketing representative or IBM Business Partner to purchase an encryption license.

12.3 Activating encryption

Encryption is enabled at a system level and all of 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 in the IBM Spectrum Virtualize system.

Both methods are available during the initial system setup and when the system is in use.

12.3.1 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 printed on it. This code allows you to proceed using the automatic activation process.

If the automatic activation process fails or if you prefer the use of the manual activation process, see this web page 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 12.3.5, “Activate the license manually” on page 617.

12.3.2 Start activation process during initial system setup

One of the steps in the initial setup enables 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.

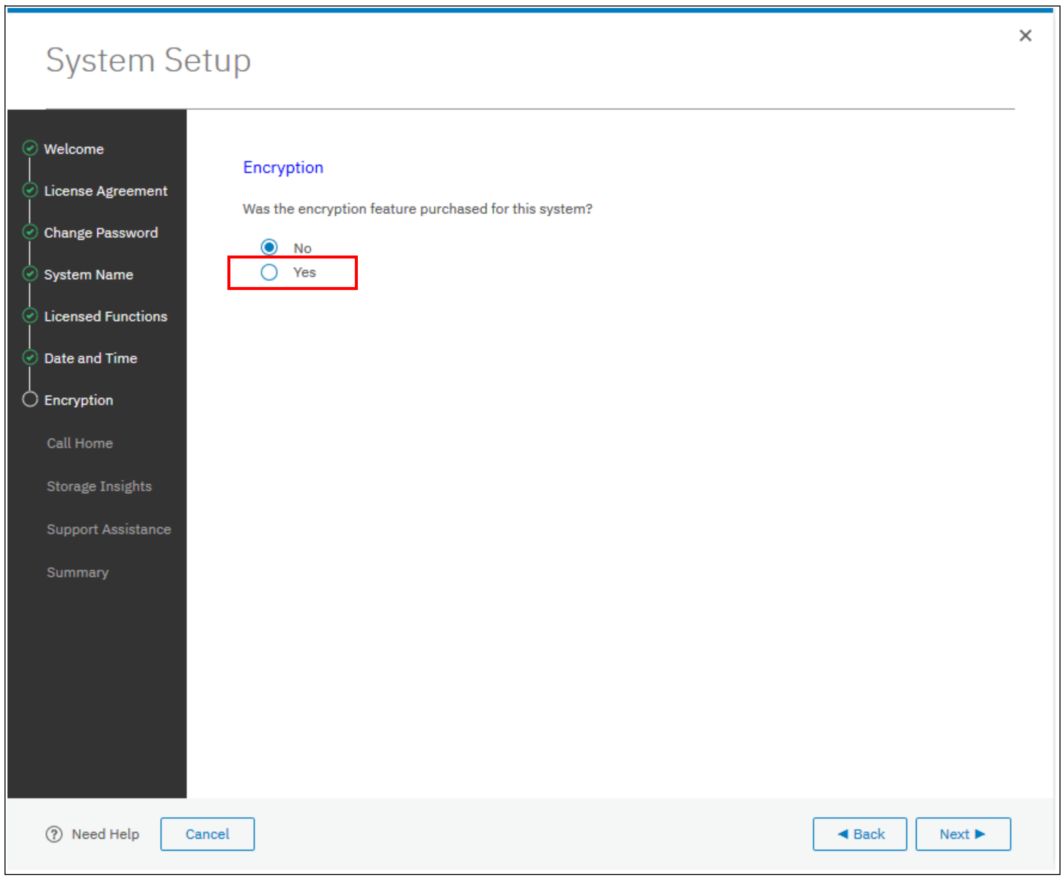


Figure 12-5 Encryption activation during initial system setup

The Encryption window displays information about your storage system, as shown in Figure 12-6.

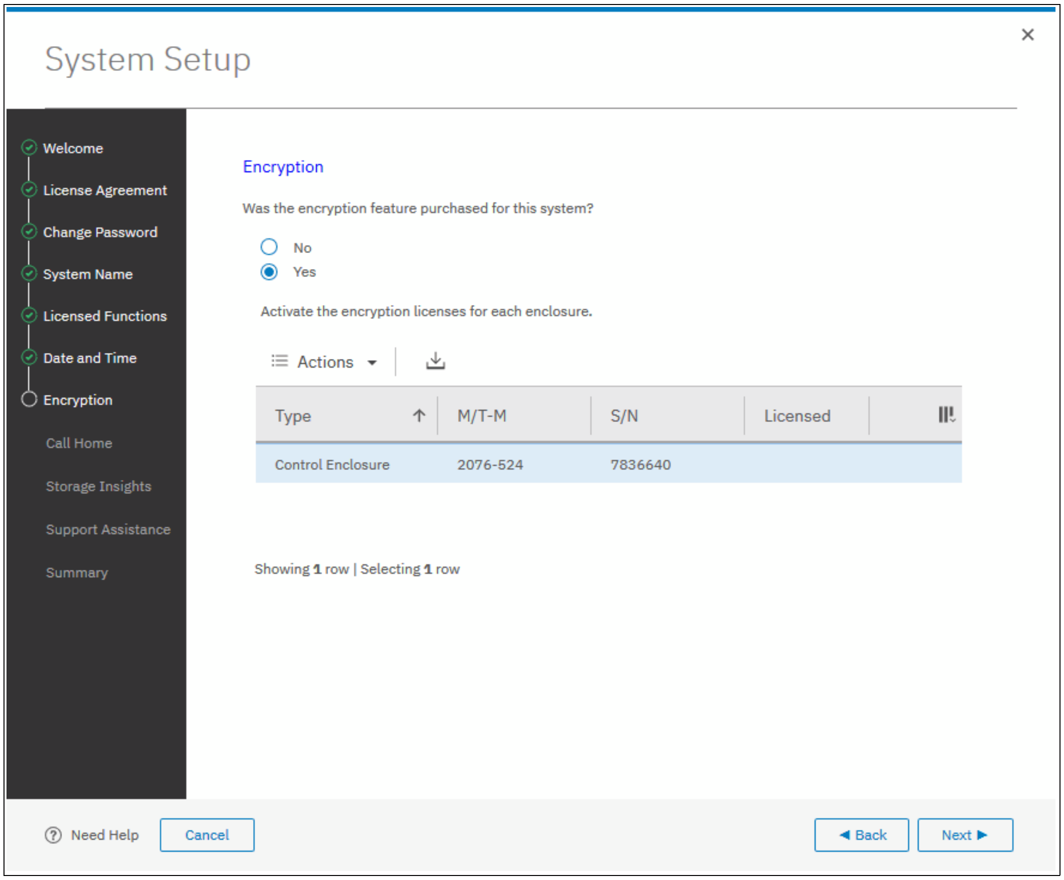


Figure 12-6 Information storage system during initial system setup

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 12.3.4, “Activate the license automatically” on page 614. For more information about how to complete a manual activation process, see 12.3.5, “Activate the license manually” on page 617.

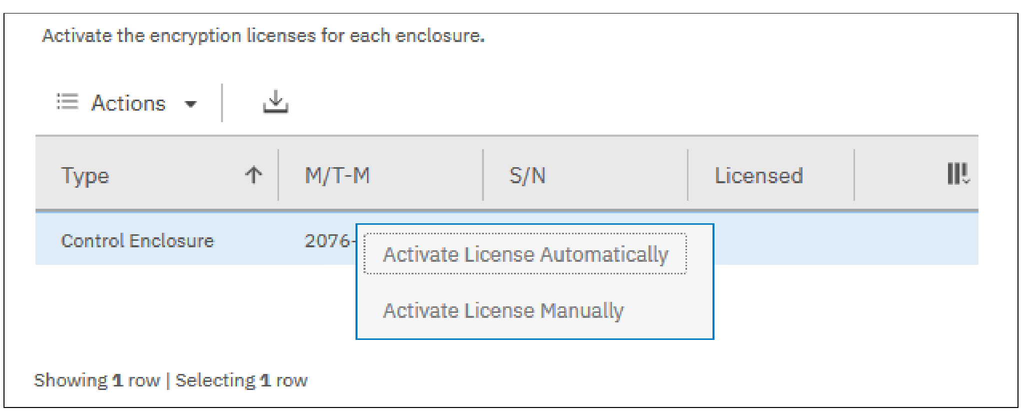


Figure 12-7 Selecting license activation method

3. After either activation process is complete, you can see a green check mark in the column 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.

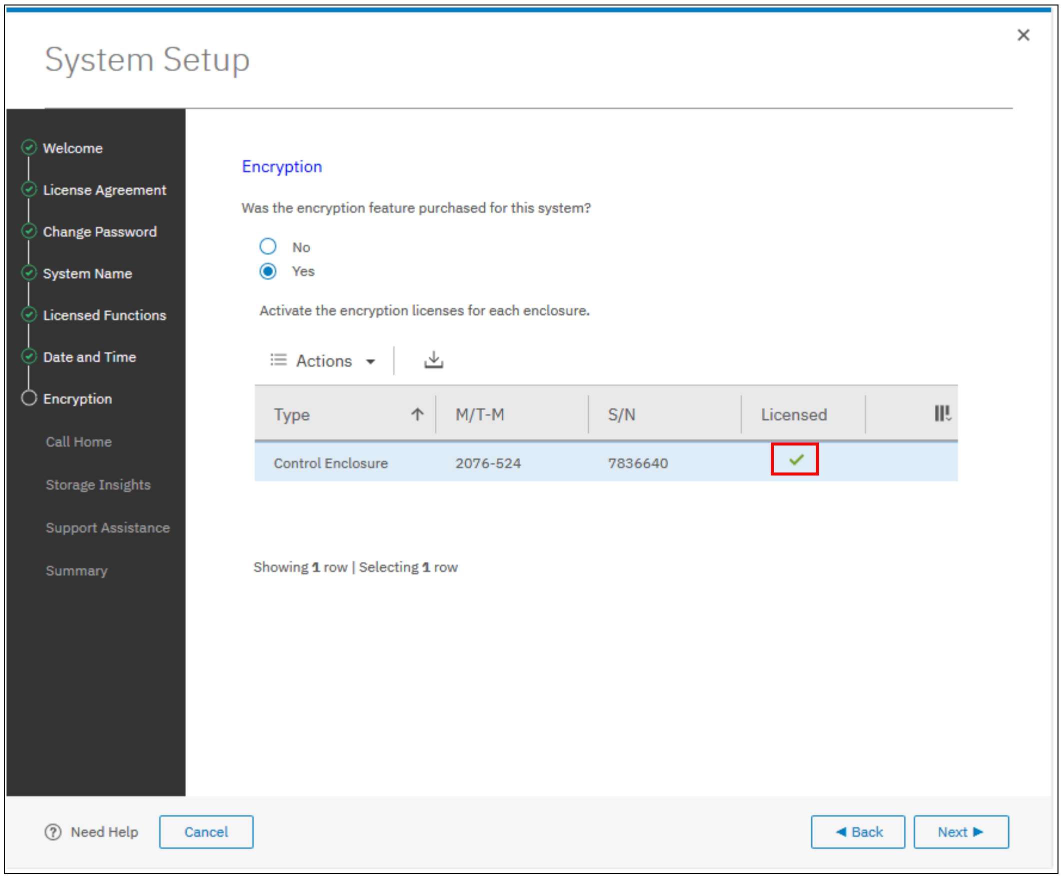


Figure 12-8 Successful encryption license activation during initial system setup

12.3.3 Start activation process on a running system

To activate encryption on a running system, complete the following steps:

1. Click **Settings** → **System** → **Licensed Functions**.
2. Click **Encryption Licenses**, as shown in Figure 12-9.

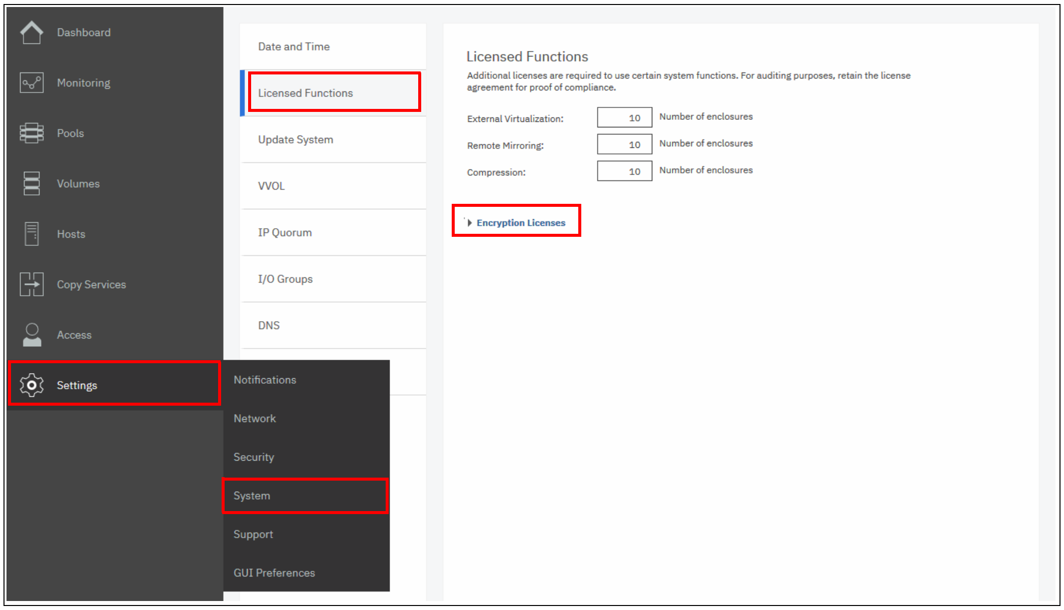


Figure 12-9 Expanding Encryption Licenses section on the Licensed Functions window

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 12.3.4, “Activate the license automatically” on page 614. For more information about how to complete a manual activation process, see 12.3.5, “Activate the license manually” on page 617.

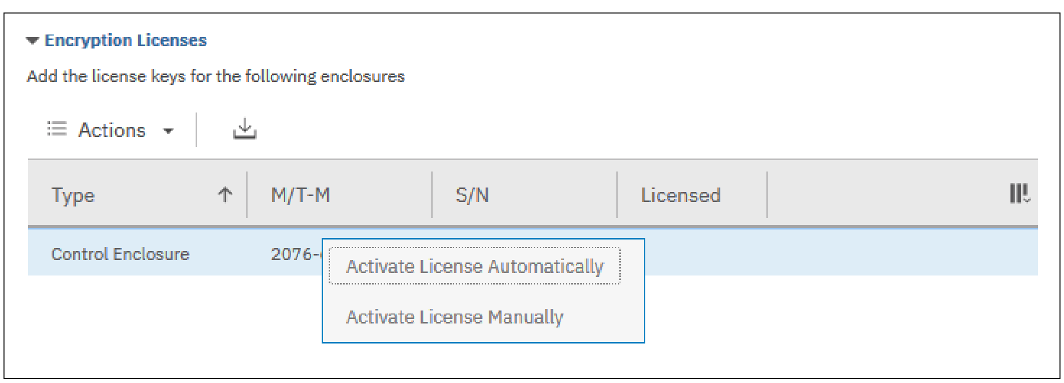


Figure 12-10 Select the Control Enclosure on which you want to enable the encryption

After either activation process is complete, you can see a green check mark in the column labeled Licensed for the control enclosure, as shown in Figure 12-11.

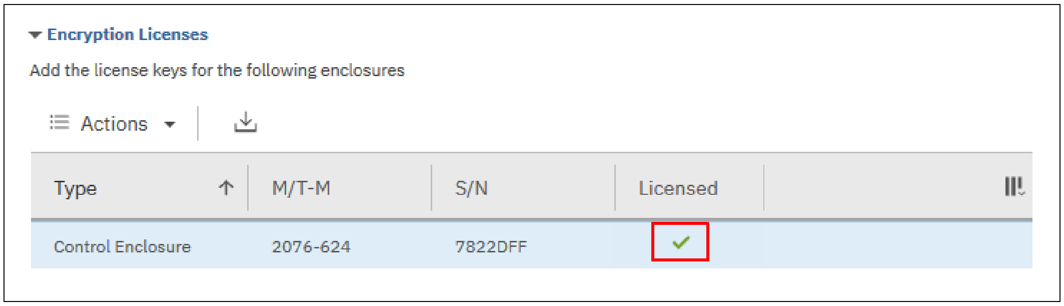


Figure 12-11 Successful encryption license activation on a running system

12.3.4 Activate 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 personal computer 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. Select **Activate License Automatically** to open the Activate License Automatically window, as shown in Figure 12-12.

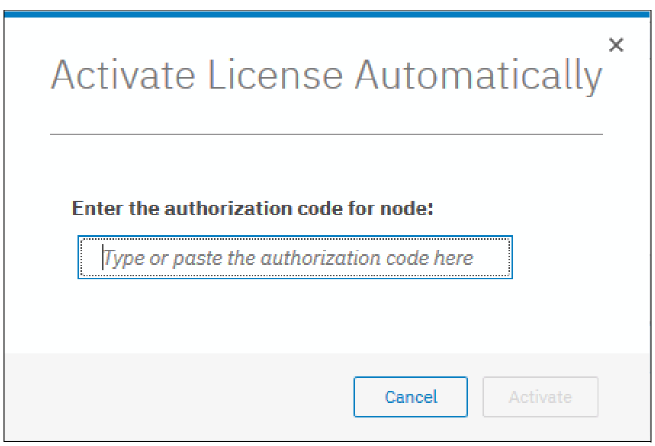


Figure 12-12 Encryption license Activate License Automatically window

2. Enter the authorization code that is specific to the control enclosure that you selected, as shown in Figure 12-13 on page 615. You can now click **Activate**.

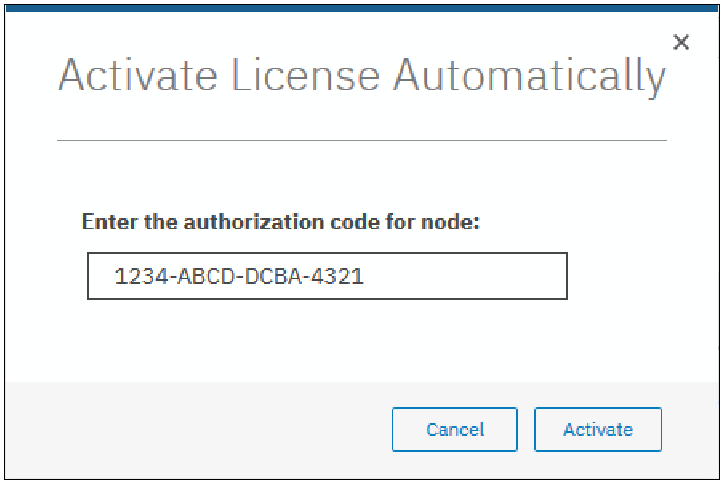


Figure 12-13 Entering an authorization code

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.

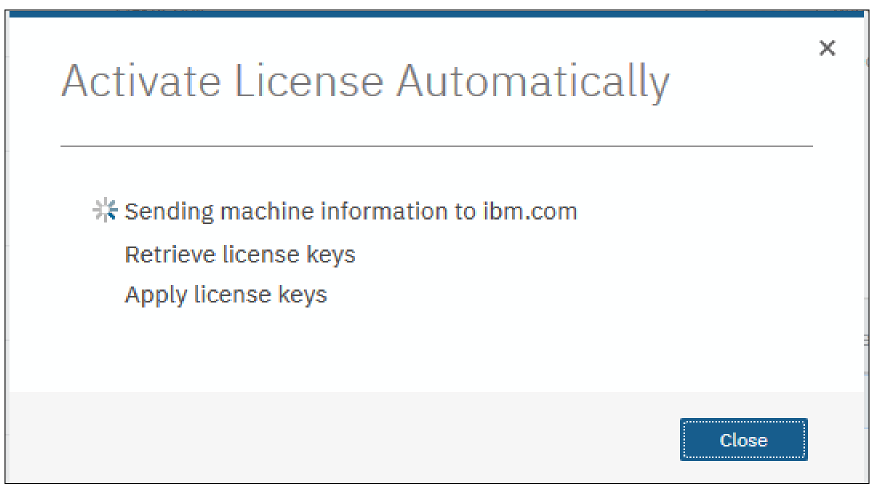


Figure 12-14 Activating encryption

After the license key is retrieved, it is automatically applied, as shown in Figure 12-15.

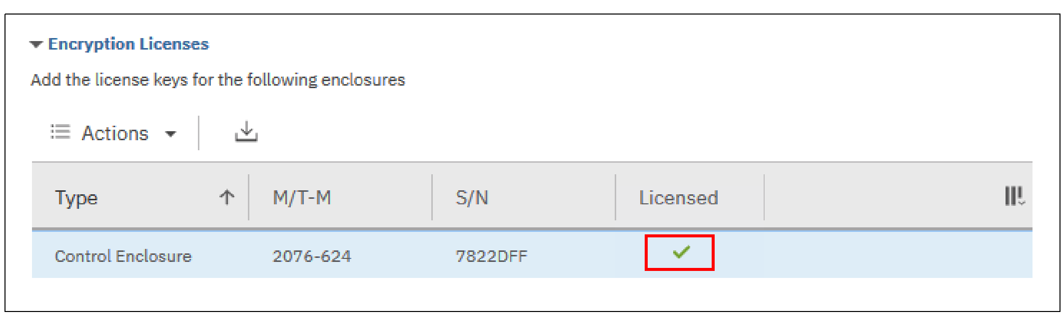


Figure 12-15 Successful encryption license activation

Problems with automatic license activation

If connections problems occur with the automatic license activation procedure, the system times out after 3 minutes with an error.

Check whether the personal computer that is used to connect to the Storwize V7000 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 12.3.5, “Activate the license manually” on page 617.

Although authorization codes and encryption license keys use the same format (four groups of four hexadecimal digits), you can only use each of them in the appropriate activation process. If you use a license key when the system expects an authorization code, the system displays an error message, as shown in Figure 12-16.

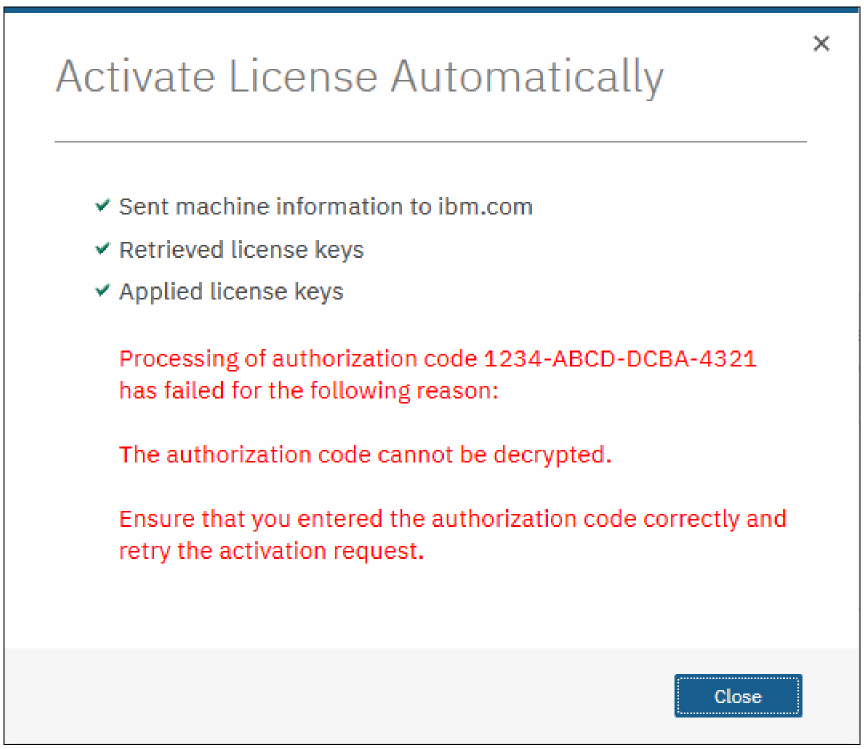


Figure 12-16 Authorization code failure

12.3.5 Activate the license manually

To manually activate the encryption license for a control enclosure, complete the following steps:

1. Select **Activate License Manually** to open the Manual Activation window, as shown in Figure 12-17.

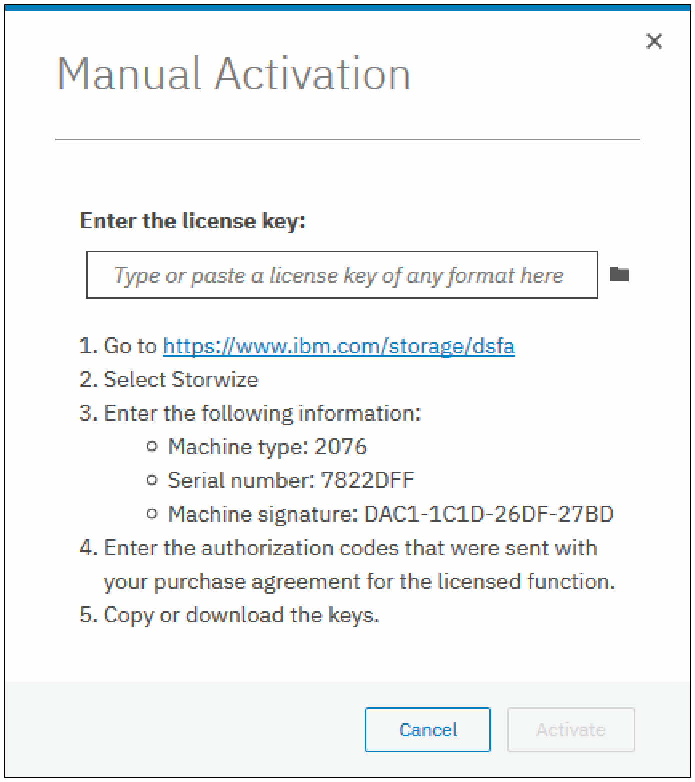


Figure 12-17 Manual encryption license activation window

2. If you have not done so already, 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 12.3.1, “Obtaining an encryption license” on page 609.

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**.

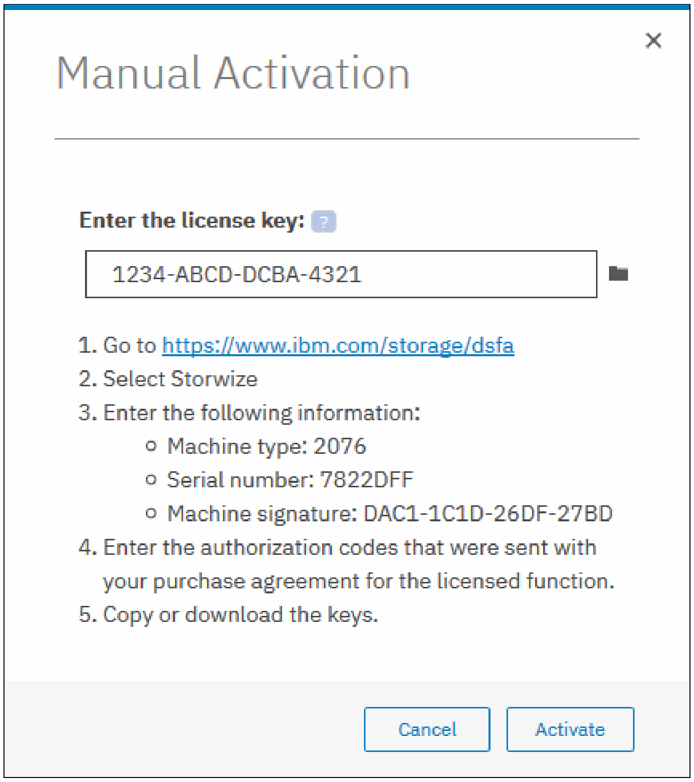


Figure 12-18 Entering an encryption license key

After the task completes successfully, the GUI shows that encryption is licensed for the specified control enclosure, as shown in Figure 12-19.

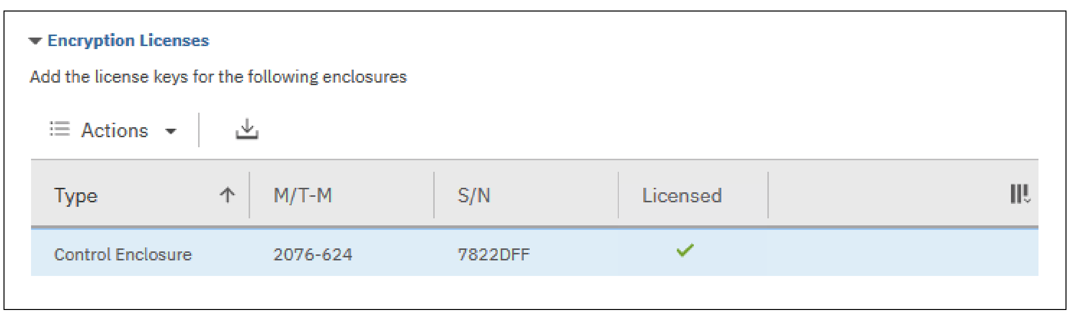


Figure 12-19 Successful encryption license activation

Problems with manual license activation

Although authorization codes and encryption license keys use the same format (four groups of four hexadecimal digits), you can only use each of them 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.

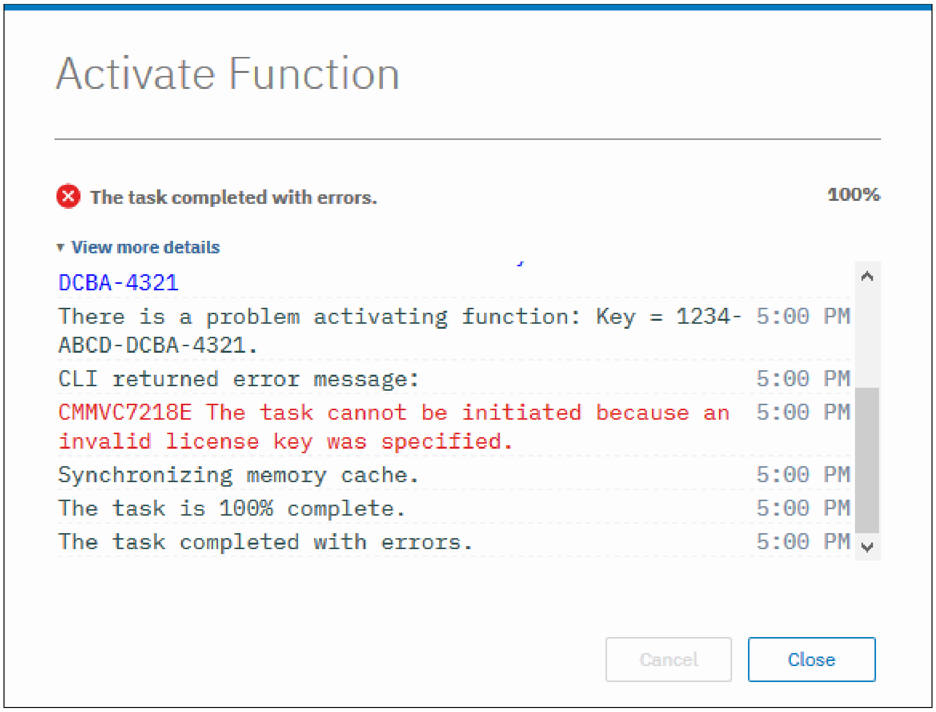


Figure 12-20 License key failure

12.4 Enabling encryption

This section describes the process to create and store system master access key copies, also referred to 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 (SKLM), introduced in IBM Spectrum Virtualize V7.8.
- Gemalto SafeNet KeySecure, introduced in IBM Spectrum Virtualize V8.2.

IBM Spectrum Virtualize code 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 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 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 security of the encryption keys. Pay special attention to ensure secure creation, management, and storage of the encryption keys.

12.4.1 Starting the Enable Encryption wizard

After the license activation step is successfully completed on IBM Storwize V7000 control enclosures, you can now enable encryption. You can enable encryption after completion of the initial system setup by using the GUI or CLI.

There are two ways in the GUI to start the Enable Encryption wizard. It can be started by clicking **Run Task** next to Enable Encryption on the Suggested Tasks window, as shown in Figure 12-21.

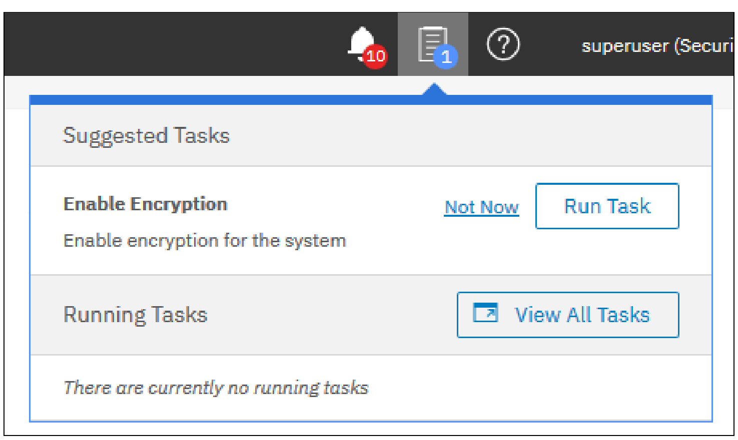


Figure 12-21 Enable Encryption from the Suggested Tasks *window*

You can also click **Settings** → **Security** → **Encryption** and then, click **Enable Encryption**, as shown in Figure 12-22.

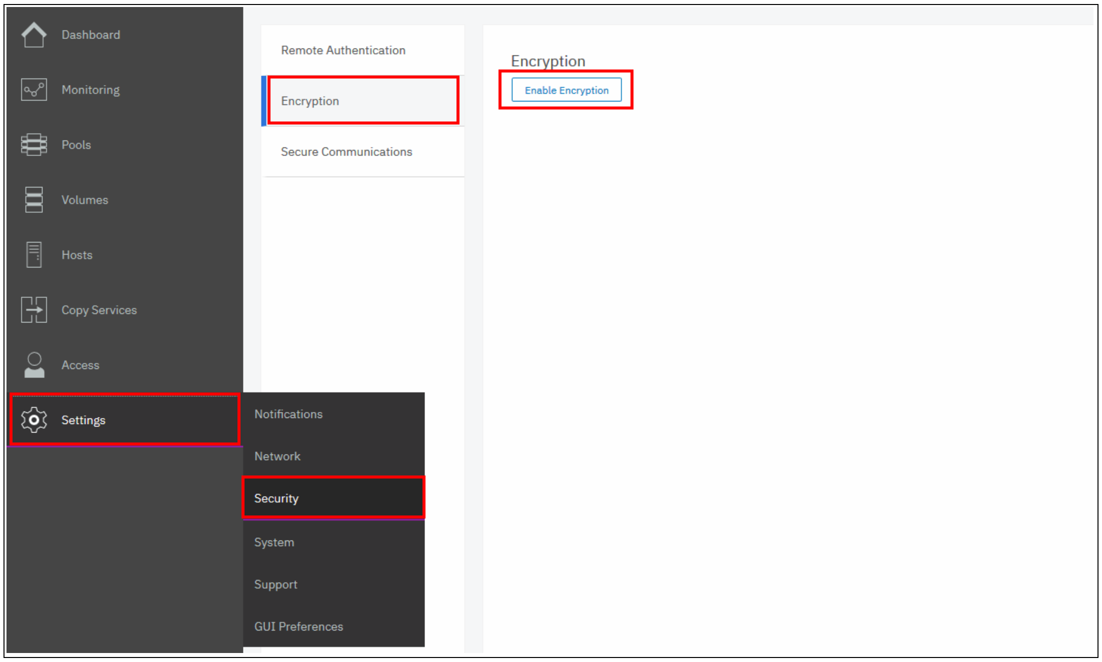


Figure 12-22 Enable Encryption from the Security pane

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. You can enable either or both providers.

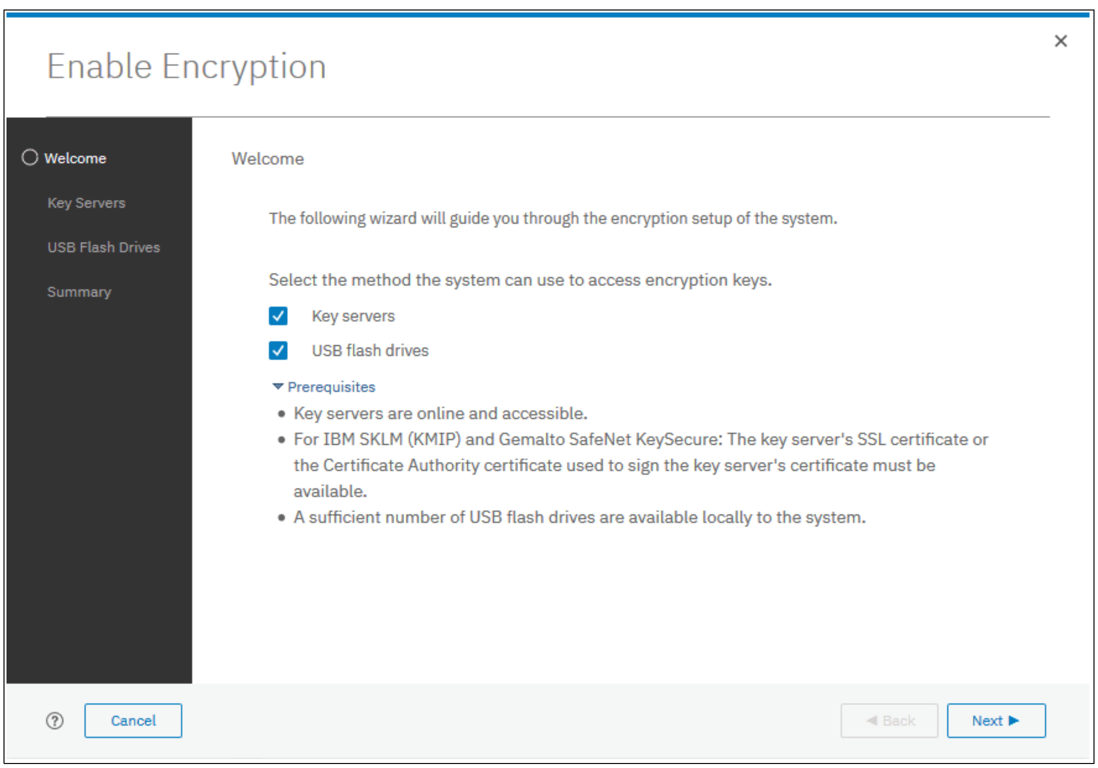


Figure 12-23 Enable Encryption wizard Welcome window

The next section presents a scenario in which both encryption key providers are enabled at the same time. For more information about how to enable encryption by using only USB flash drives, see 12.4.2, “Enabling encryption using USB flash drives” on page 622.

For more information about how to enable encryption by using key servers as the sole encryption key provider, see 12.4.3, “Enabling encryption using key servers” on page 627.

12.4.2 Enabling encryption using USB flash drives

> **Note:** The system needs at least three USB flash drives to be present 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 name 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, 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 12-26 on page 625 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 reboot or power loss. No USB flash drive is required during a warm reboot, such as a node exiting service mode or a single node reboot. The data center power-on procedure needs to ensure that USB flash drives containing encryption keys are plugged into the storage system before it is powered on.

During power-on, insert 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.

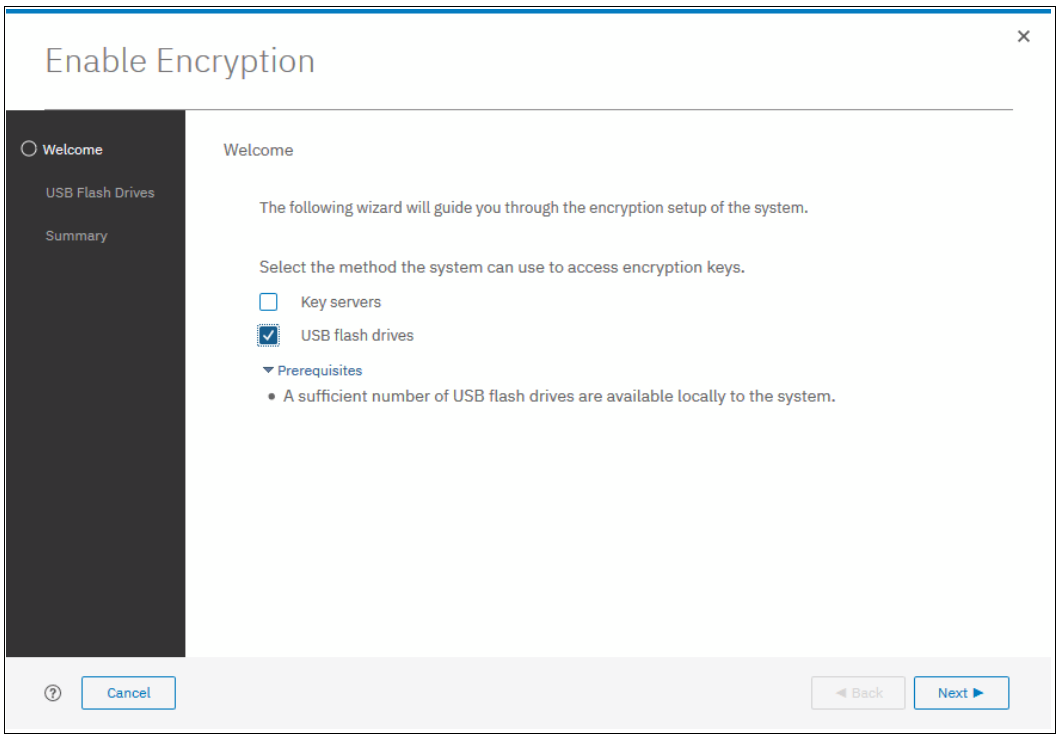


Figure 12-24 Selecting USB flash drives in the Enable Encryption wizard

2. If there are fewer than three USB flash drives inserted into the system, you are prompted to insert more drives, as shown in Figure 12-25. The system reports how many more drives must be inserted.

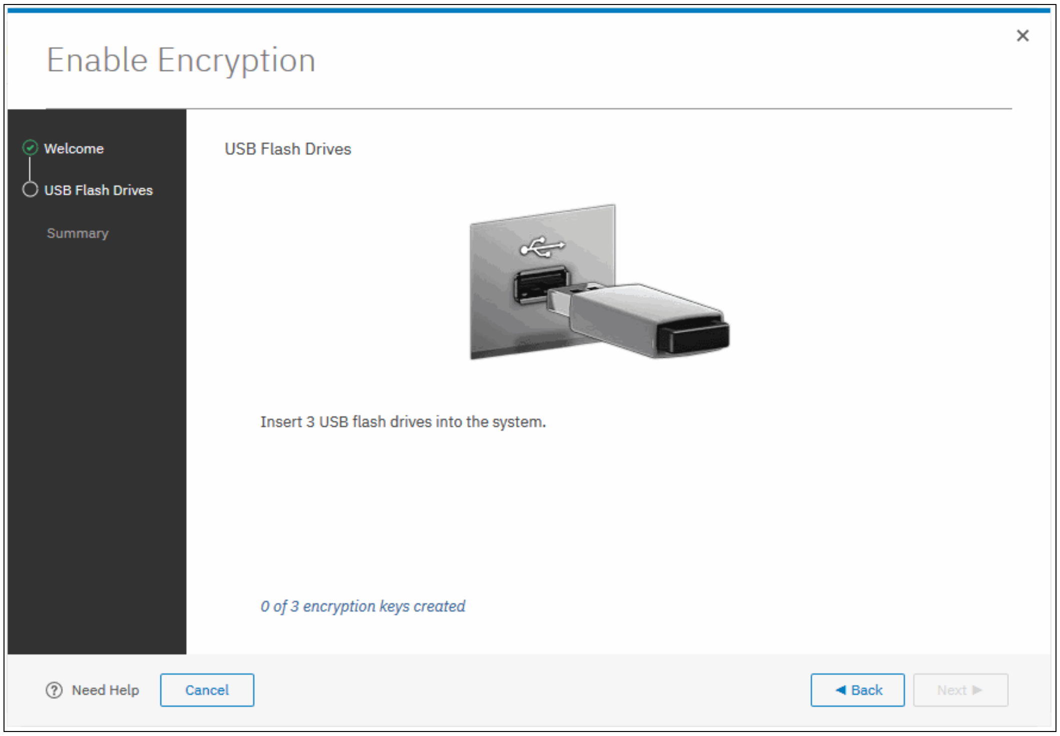


Figure 12-25 Waiting for USB flash drives to be inserted

> **Note:** The **Next** option remains disabled until at least three USB flash drives are detected.

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 on the USB flash drives, as shown in Figure 12-26.

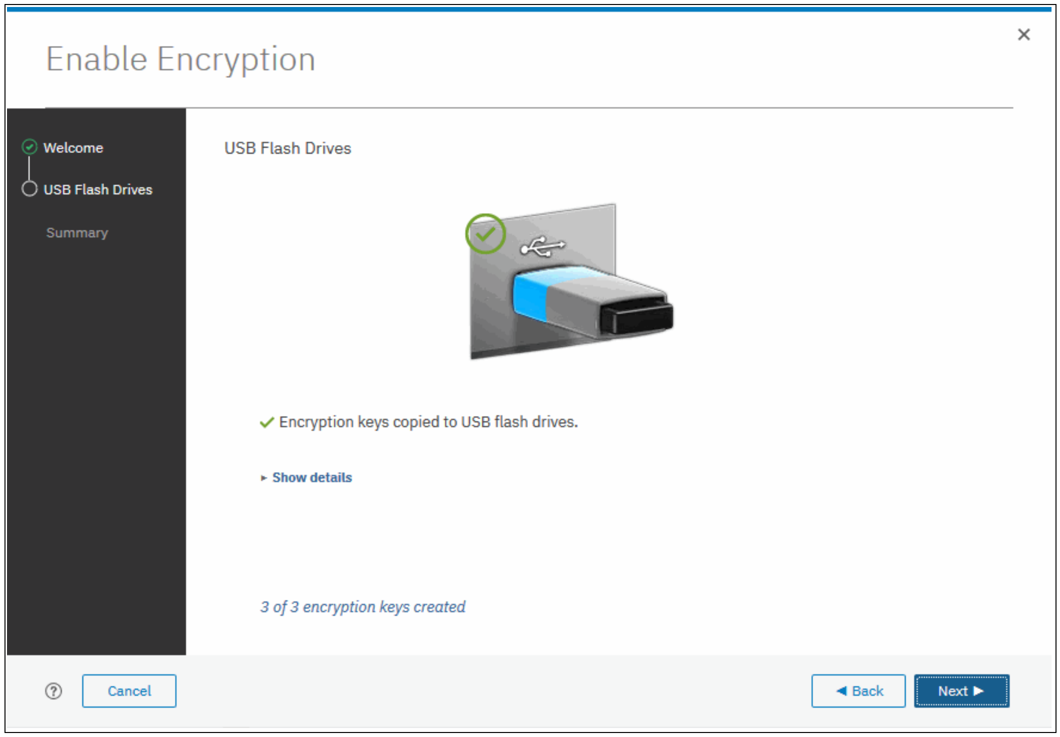


Figure 12-26 Writing the master access key to USB flash drives

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 12-27. Click **Finish** to finalize the encryption enablement.

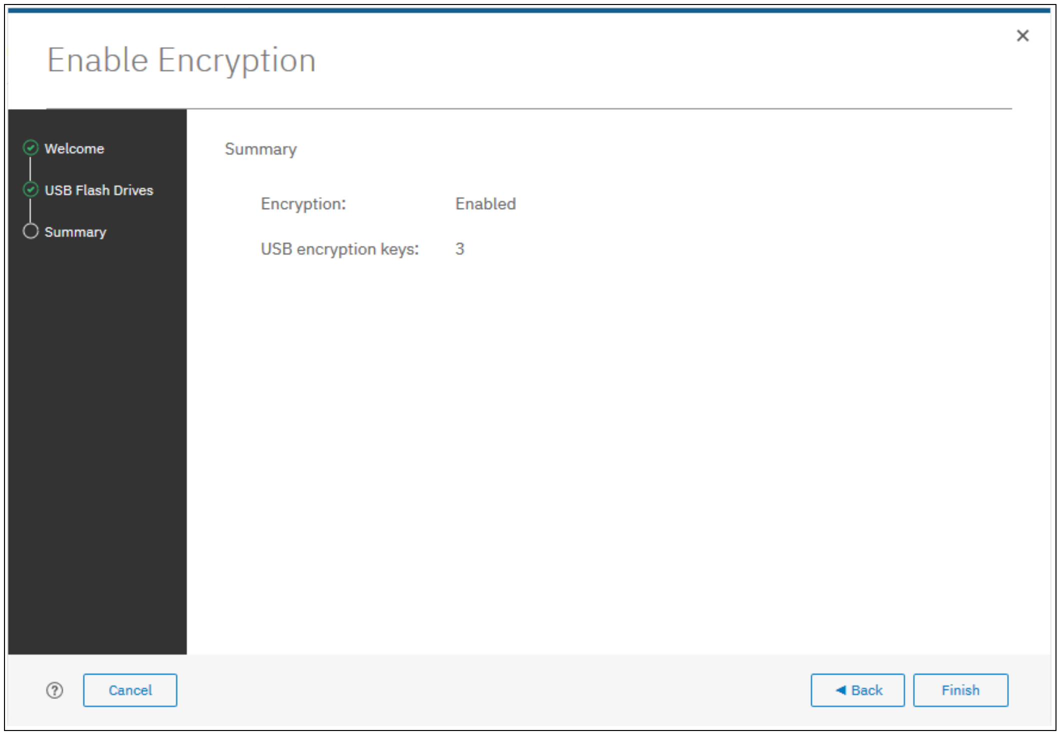


Figure 12-27 Commit the encryption enablement

You receive a message confirming that the encryption is now enabled on the system, as shown in Figure 12-28.

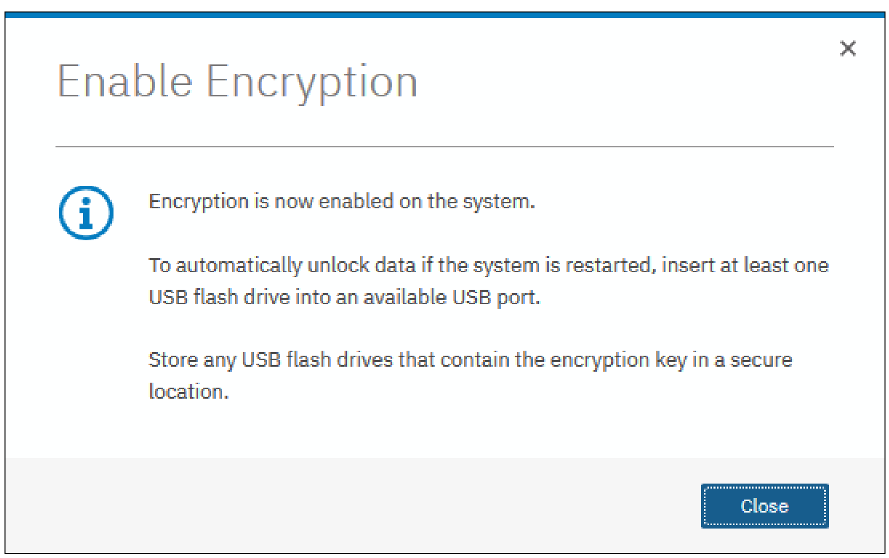


Figure 12-28 Encryption enabled message using USB flash drives

5. You can confirm that encryption is enabled and verify which key providers are in use by clicking **Settings** → **Security** → **Encryption**, as shown in Figure 12-29.

Figure 12-29 Encryption view showing using USB flash drives as the enabled provider

12.4.3 Enabling encryption 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 use of the following key servers as encryption key providers:

- IBM Security Key Lifecycle Manager (SKLM)
- 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.

SKLM and KeySecure SafeNet support Key Management Interoperability Protocol (KMIP), which is a standard for management of cryptographic keys.

Note: Make sure that the key management server functionality is fully independent from encrypted storage that has encryption 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 code V8.1 and later supports up to four key server objects that are defined in parallel. But, only one key server type (SKLM 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 need to migrate from your current key server to USB encryption, and then migrate from USB to the other type of key server.

Enabling encryption using SKLM

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 SKLM server before you enable encryption on the storage system:

- Configure the SKLM server to use Transport Layer Security version 1.2. 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 TLSv1.2) in the SKLM 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 a `SPECTRUM_VIRT` device group for IBM Spectrum Virtualize systems.

For more information about completing these tasks, see the SKLM documentation at IBM Knowledge Center.

Access to the key server storing the correct master access key is required to enable encryption for the cluster after a system restart, such as a system-wide reboot or power loss. Access to the key server is not required during a warm reboot, such as a node exiting service mode or a single node reboot. The data center power-on procedure must ensure key server availability before the storage system using encryption is booted.

To enable encryption by using an SKLM key server, complete the following steps:

1. Ensure that service IPs are configured on all your nodes.
2. In the Enable Encryption wizard Welcome tab, select **Key servers** and click **Next**, as shown in Figure 12-30.

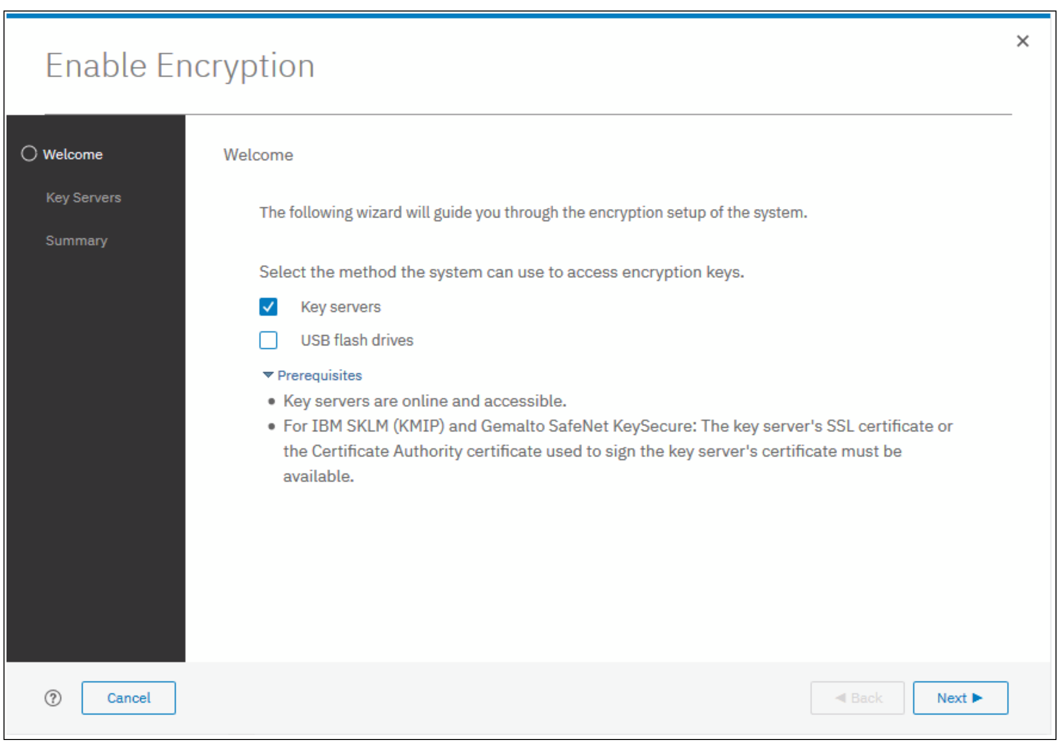


Figure 12-30 Selecting Key server as the only provider in the Enable Encryption wizard

3. Select IBM SKLM (with KMIP) as the key server type, as shown in Figure 12-31.

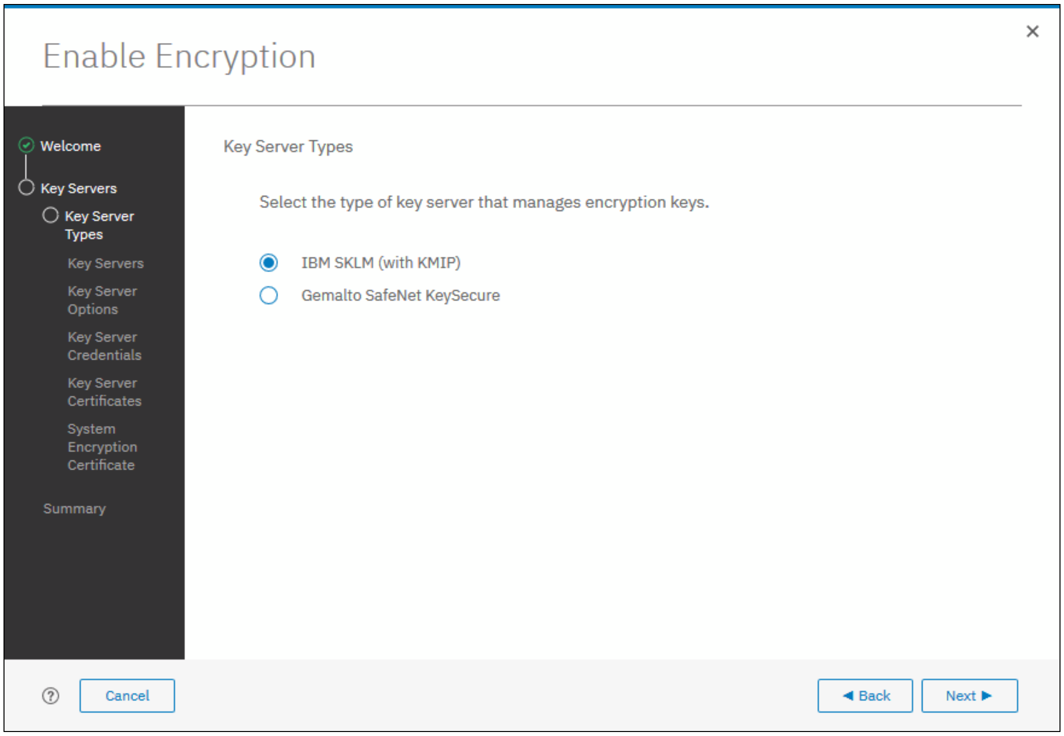


Figure 12-31 Selecting SKLM as key server type

4. The wizard moves to the Key Servers tab, as shown in Figure 12-32 on page 630. Enter the name and **IP address** of the key servers. Note that the first key server specified must be the primary SKLM key server.

> **Note:** The supported versions of IBM Security Key Lifecycle Manager (up to V3.0, which was the latest code version available at the time of this writing) differentiate between the primary and secondary key server role. The Primary SKLM server as defined on the Key Servers window of the Enable Encryption wizard must be the server defined as the primary by SKLM 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 SKLM configuration is shown in Figure 12-32 on page 630.

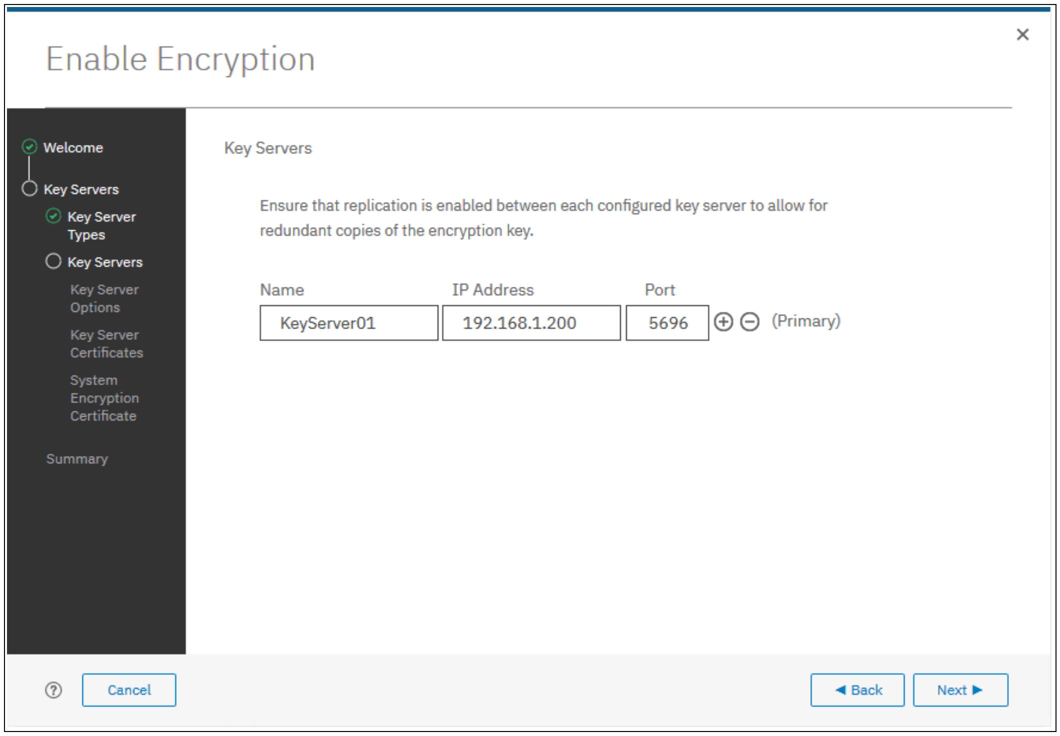


Figure 12-32 Configuration of the primary SKLM server

5. If you want to add secondary SKLM servers, click the "**+**" symbol and enter the data for secondary SKLM servers, as shown on Figure 12-33. You can define up to four SKLM servers. Click **Next** when you are done.

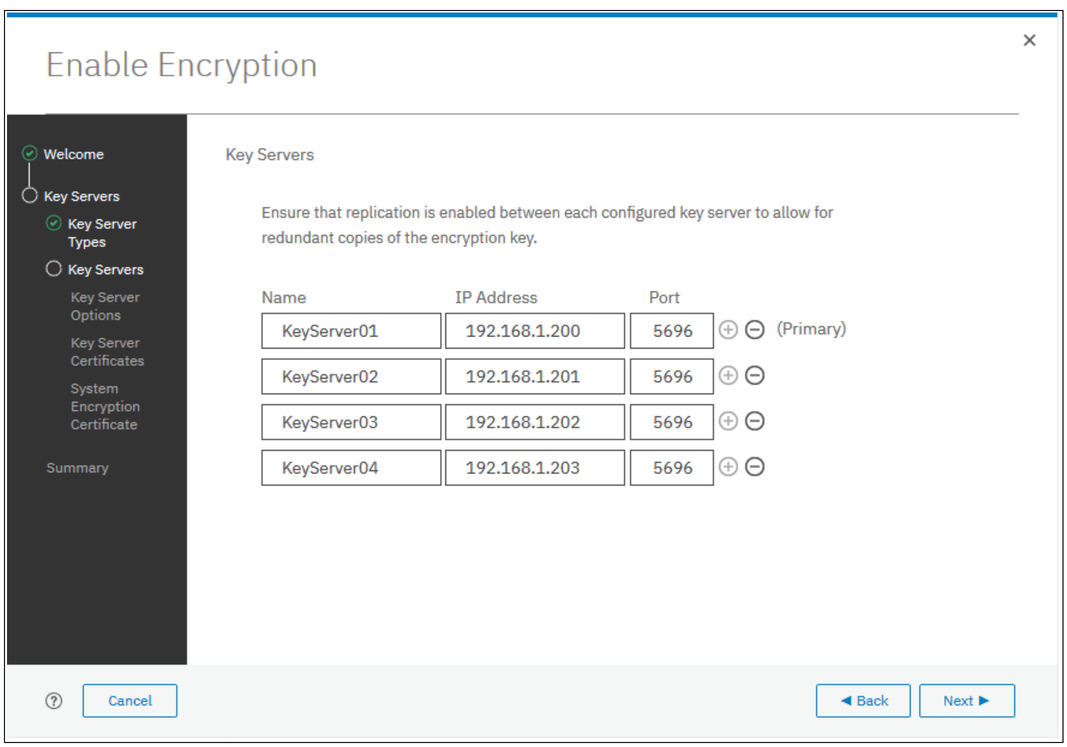


Figure 12-33 Configuring multiple SKLM servers

6. The next window in the wizard is a reminder that SPECTRUM_VIRT device group dedicated for IBM Spectrum Virtualize systems must exist on the SKLM key servers. Make sure that this device group exists and click **Next** to continue, as shown in Figure 12-34.

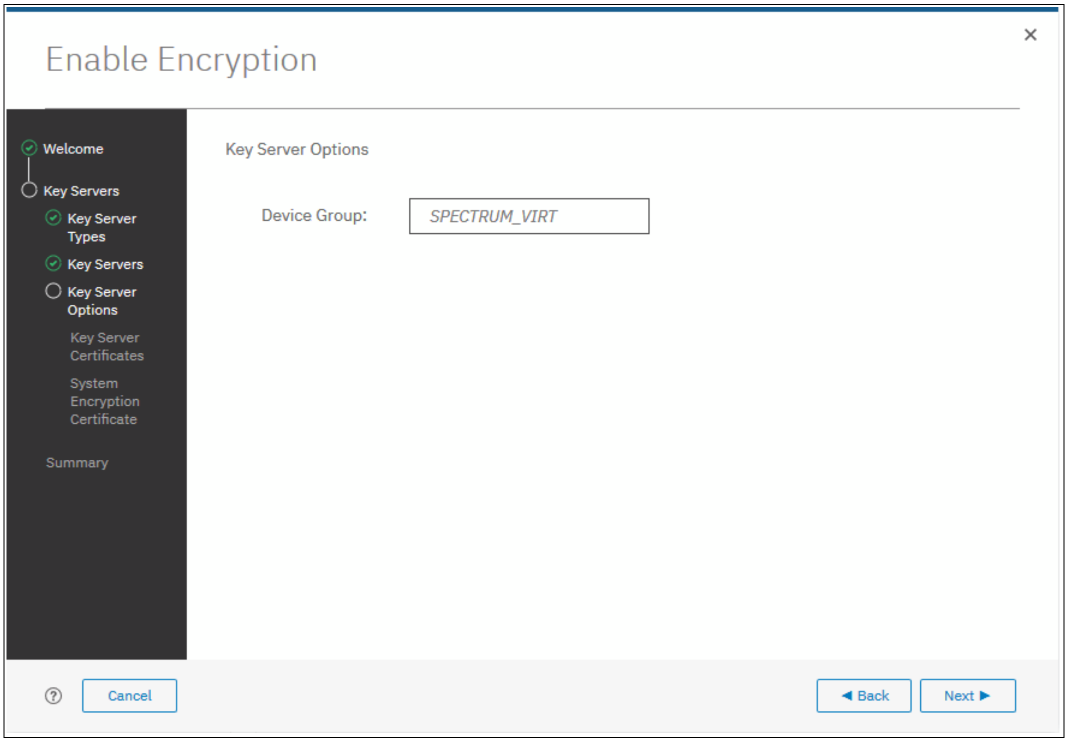


Figure 12-34 Checking key server device group

7. Enable secure communication between the IBM Spectrum Virtualize system and the SKLM key servers by uploading the key server certificate (from a trusted third party or a self-signed certificate), or by uploading the public SSL certificate of each key server directly.

   After uploading any of the certificates in the window that us shown in Figure 12-35 on page 632, click **Next**.

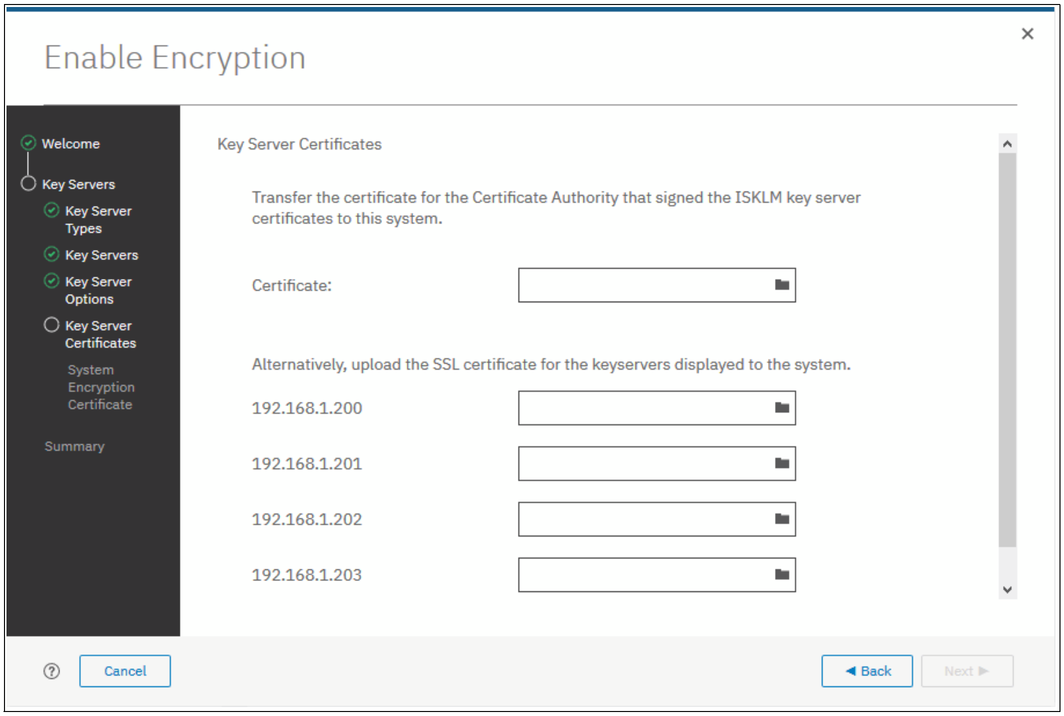


Figure 12-35 Uploading key servers or Certificate Authority SSL certificate

8. Configure the SKLM 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 12-36. Install this certificate in the SKLM key server in the SPECTRUM_VIRT device group.

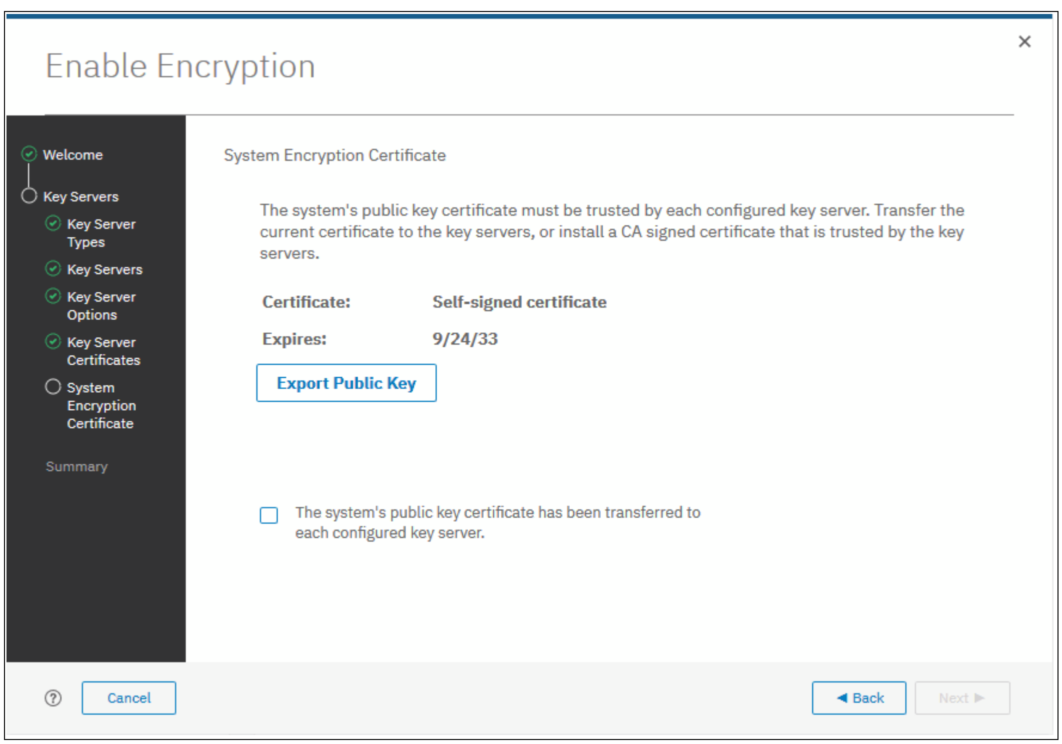


Figure 12-36 Downloading the IBM Spectrum Virtualize SSL certificate

9. When the IBM Spectrum Virtualize system public key certificate is installed on the SKLM key servers, acknowledge this installation by selecting the box that is indicated in Figure 12-37 and click **Next**.

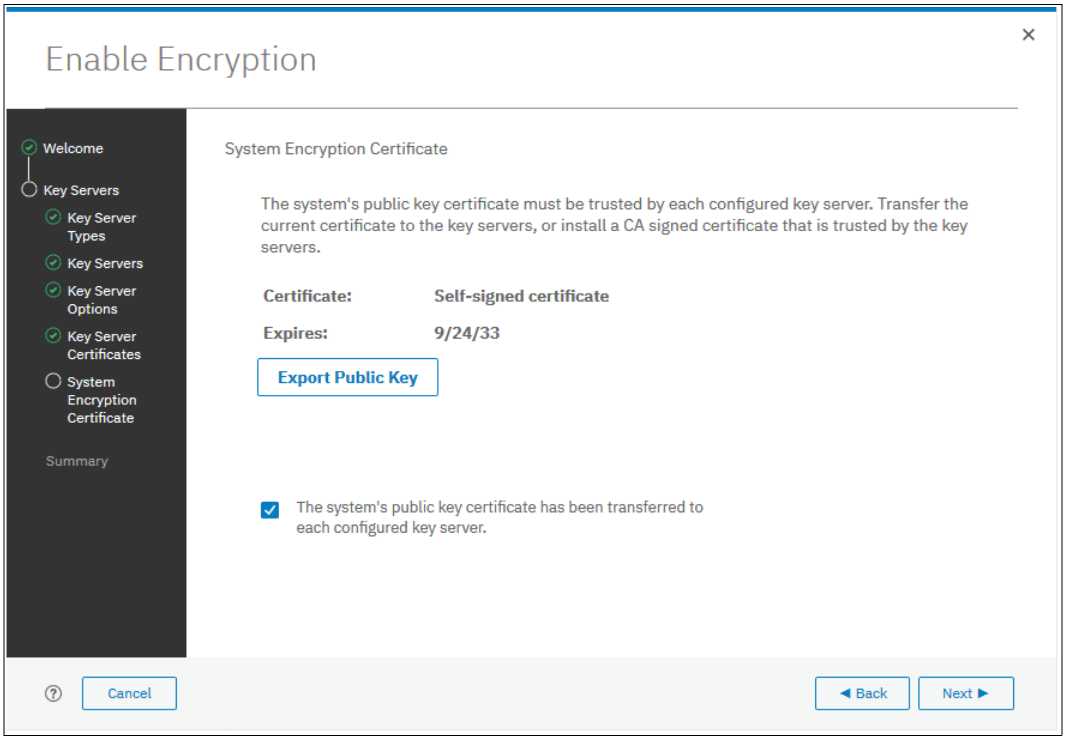


Figure 12-37 Acknowledge IBM Spectrum Virtualize public key certificate transfer

10. The key server configuration is shown in the Summary tab, as shown in Figure 12-38 on page 634. Click **Finish** to create the key server object and finalize the encryption enablement.

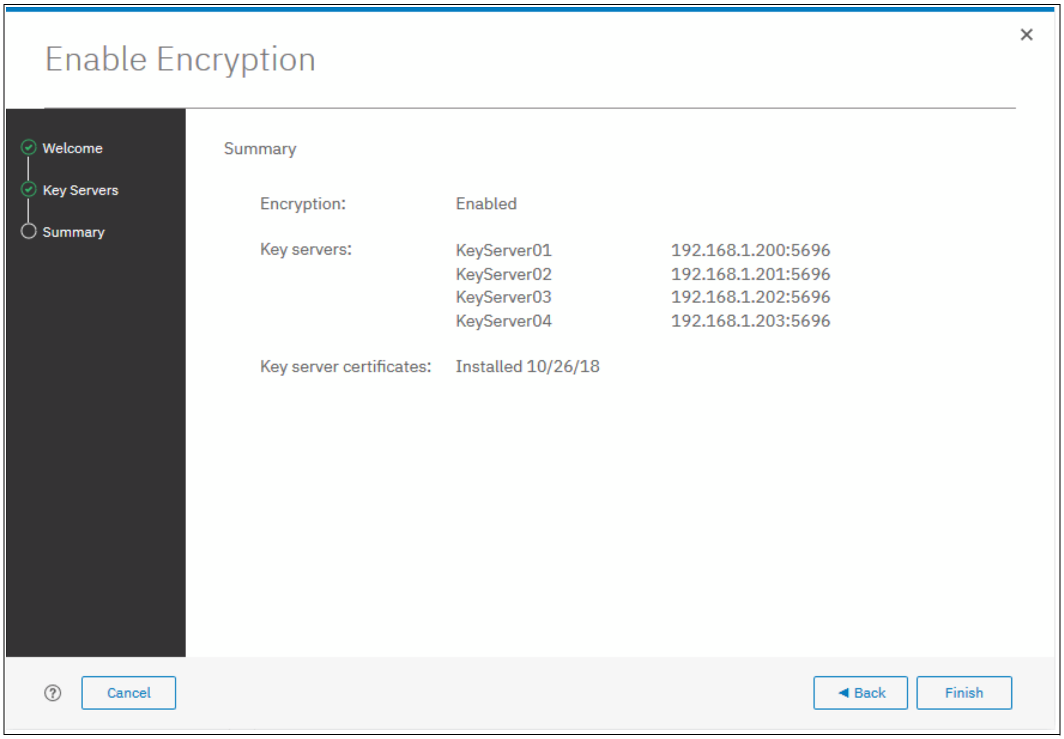


Figure 12-38 Finish enabling encryption using SKLM key servers

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, as shown in Figure 12-39. Click **Close**.

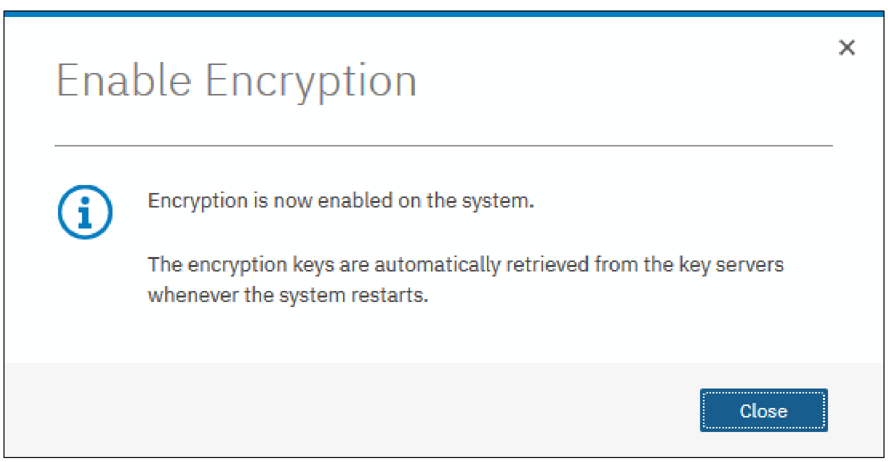


Figure 12-39 Encryption enabled message using an SKLM key server

12. Confirm that encryption is enabled in **Settings** → **Security** → **Encryption**, as shown in Figure 12-40. Note the `Online` state that indicates which SKLM servers are detected as available by the system.

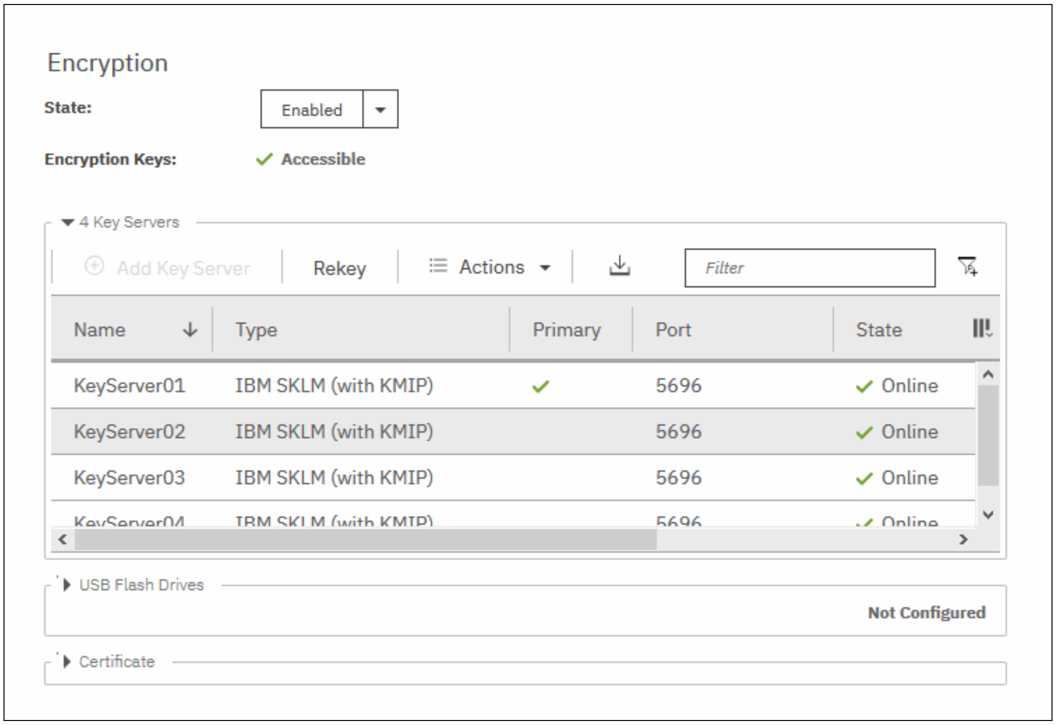


Figure 12-40 Encryption enabled with only SKLM servers as encryption key providers

Enabling encryption using SafeNet KeySecure

IBM Spectrum Virtualize V8.2.1 introduces 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 (SKLM).

IBM Spectrum Virtualize supports Gelmato SafeNet KeySecure version 8.3.0 and later, and only using 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 SKLM key servers configured at the same time in IBM Spectrum Virtualize. It is also not possible to migrate directly from one type of key server to another (from SKLM 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 KMIP protocol as IBM SKLM does, an option is available to configure the user name and password for IBM Spectrum Virtualize and KeySecure server authentication, which is not possible when the configuration is performed with SKLM.

The certificate for client authentication in SafeNet KeySecure can be self-signed or signed by a Certificate Authority.

To enable encryption in IBM Spectrum Virtualize by using a Gemalto SafeNet KeySecure key server, complete the following steps:

1. Ensure that the service IPs are configured on all of your nodes.
2. In the Enable Encryption wizard Welcome tab, select **Key servers** and click **Next**, as shown in Figure 12-41.

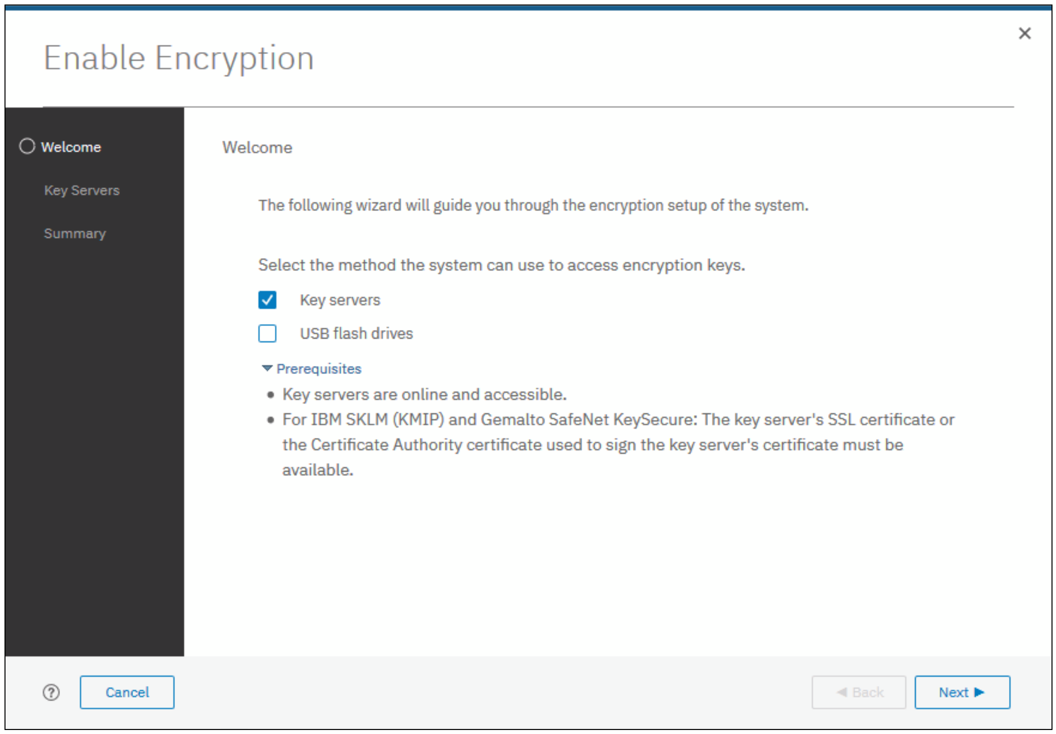


Figure 12-41 Selecting Key servers as the only provider in the Enable Encryption wizard

3. In the next window, you can choose between IBM SKLM or Gemalto SefeNet KeySecure server types, as shown in Figure 12-42. Select **Gemalto SefeNet KeySecure** and click **Next**.

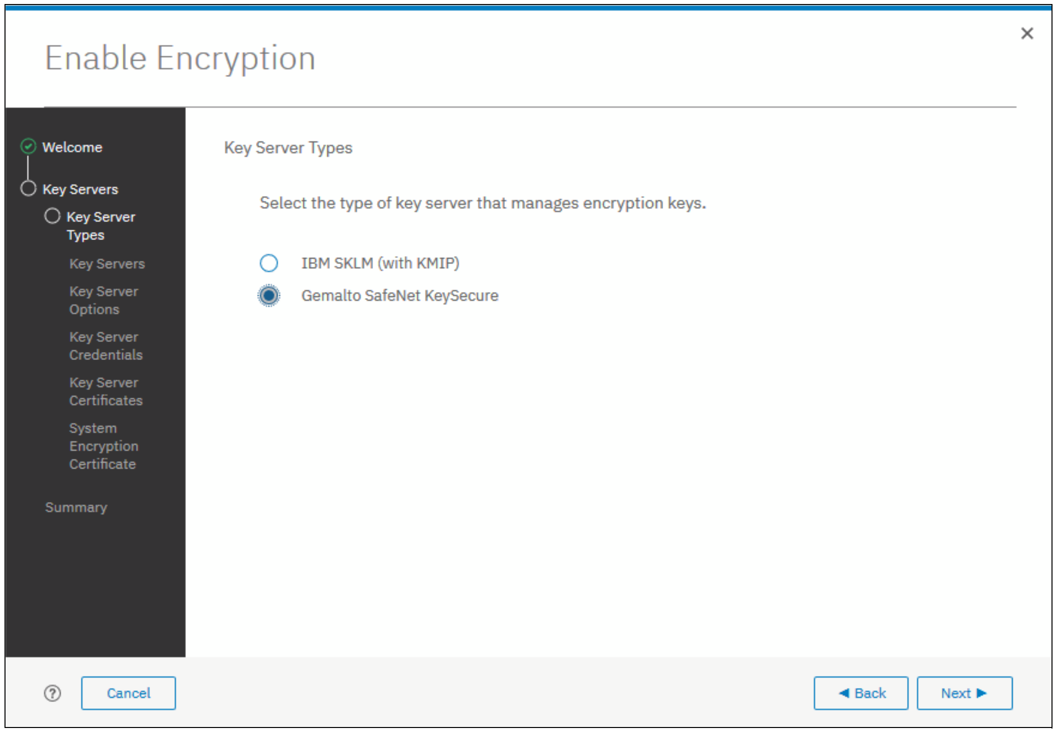


Figure 12-42 Selecting Gemalto SafeNet KeySecure as key server type

4. Add up to four SafeNet KeySecure servers in the next wizard window, as shown in Figure 12-43 on page 638. For each key server, enter the name, IP address and TCP port for KMIP protocol (default value is 5696). Note that server name is only a label, so it does not need to be the real host name 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.

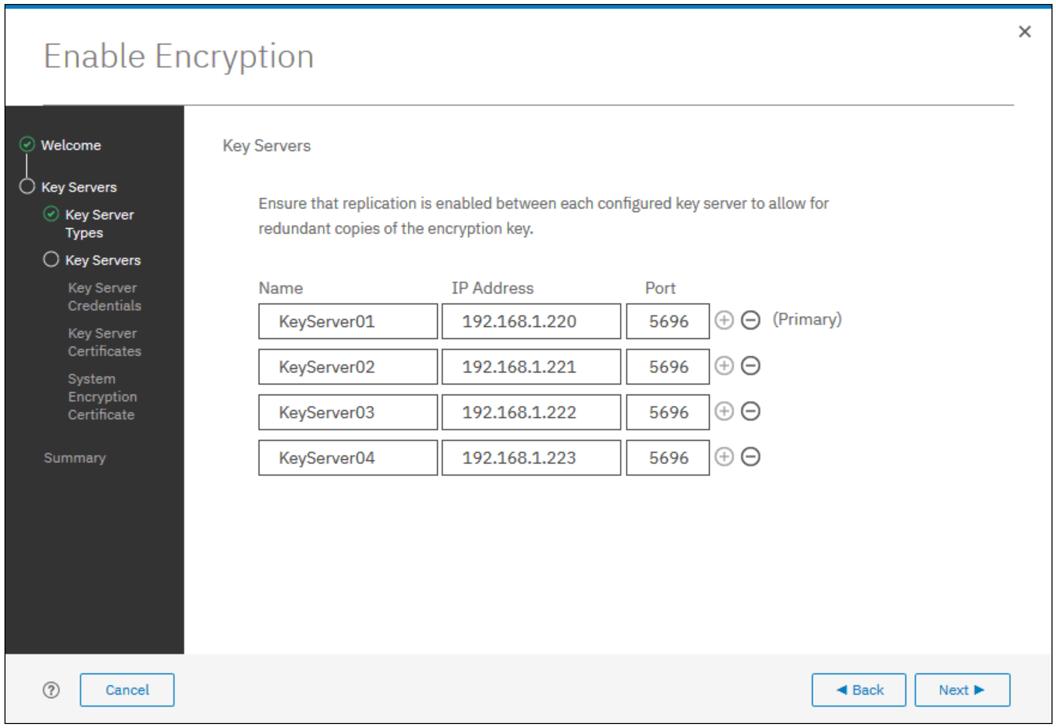


Figure 12-43 Configuring multiple SafeNet KeySecure servers

5. The next window in the wizard prompts for key servers credentials (user name and password), as shown in Figure 12-44. This setting is optional because it depends on how SafeNet KeySecure servers are configured.

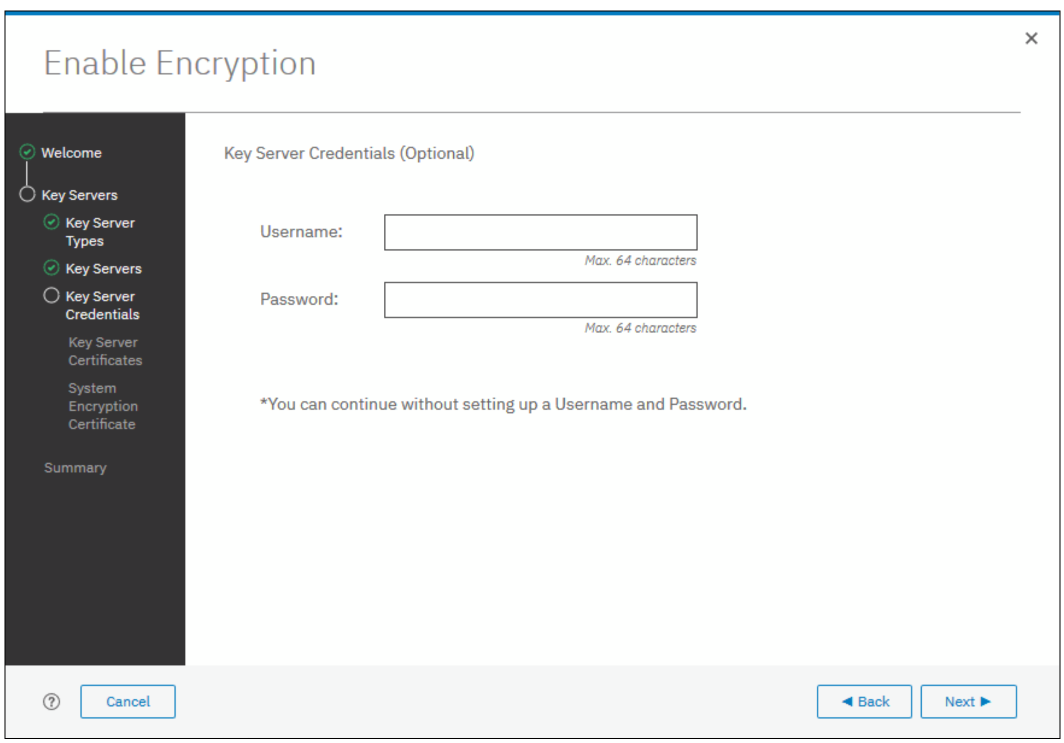


Figure 12-44 Key server credentials input (optional)

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 or a self-signed certificate), or by uploading the SSL certificate of each key server directly. After uploading any of the certificates in the window that is shown in Figure 12-45, click **Next**.

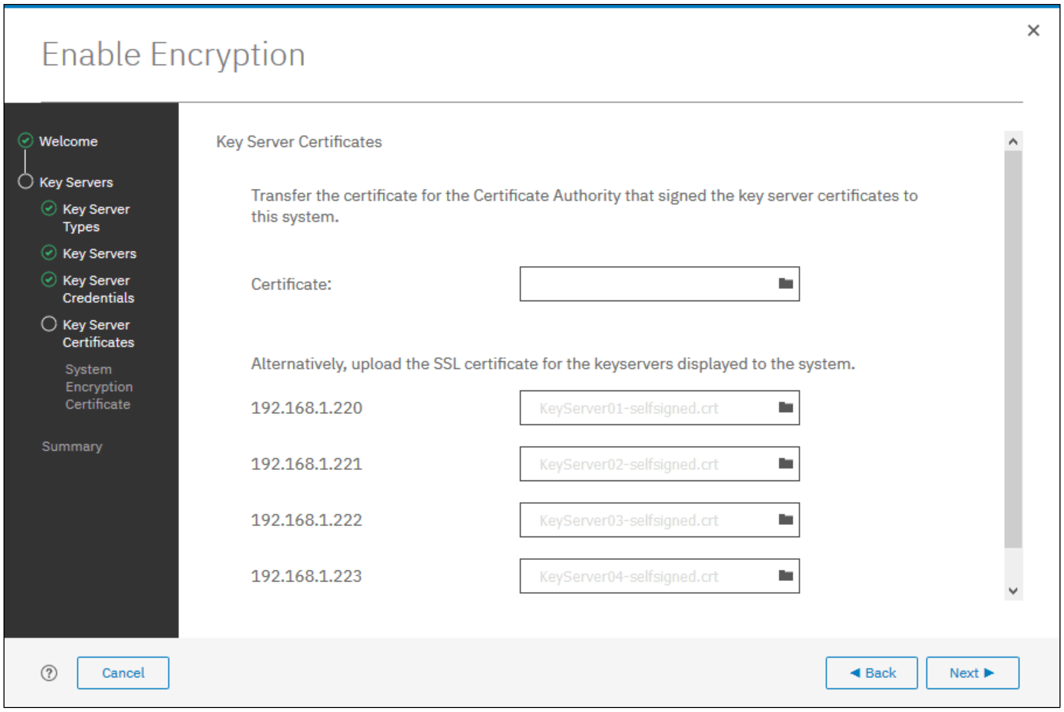


Figure 12-45 Uploading SafeNet KeySecure key servers certificate

7. 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 12-45. After adding the public key certificate to the key servers, select the box and click **Next**.

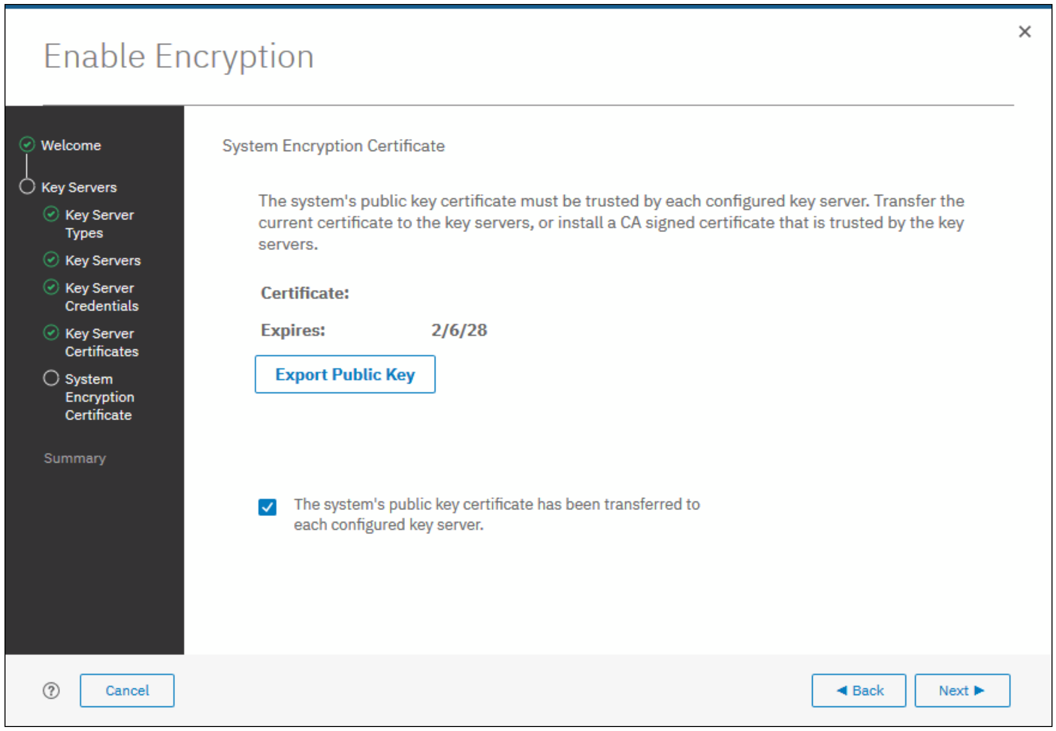


Figure 12-46 Downloading the IBM Spectrum Virtualize SSL certificate

8. The key server configuration is shown in the Summary tab, as shown in Figure 12-47. Click **Finish** to create the key server object and finalize the encryption enablement.

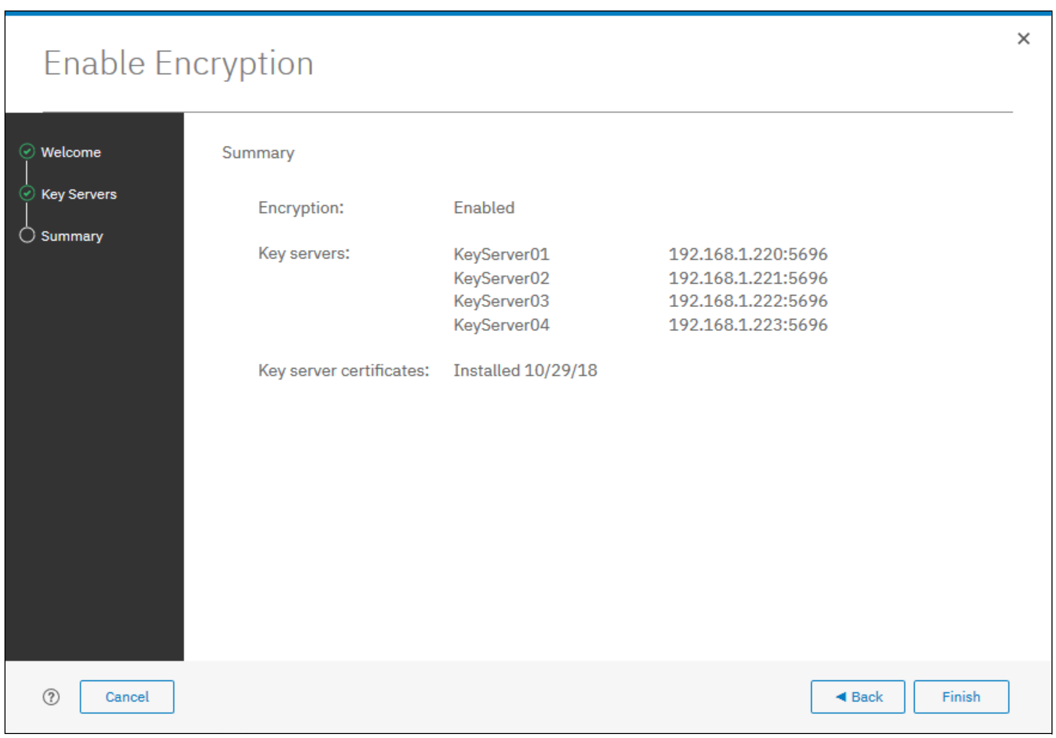


Figure 12-47 Finish enabling encryption using SafeNet KeySecure key servers

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, as shown in Figure 12-48. Click **Close**.

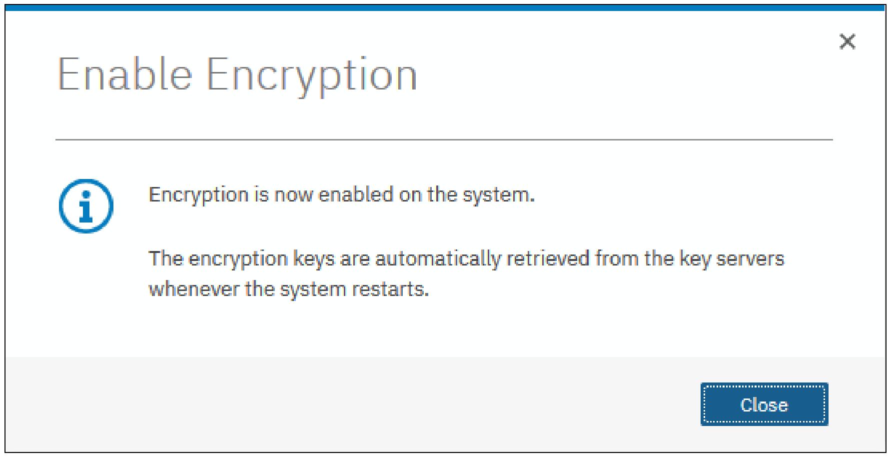


Figure 12-48 Encryption enabled using SafeNet KeySecure key servers

10. Confirm that encryption is enabled in **Settings** → **Security** → **Encryption**, as shown in Figure 12-49. Check whether the four servers are shown as online, which indicate that all four SafeNet KeySecure servers are detected as available by the system.

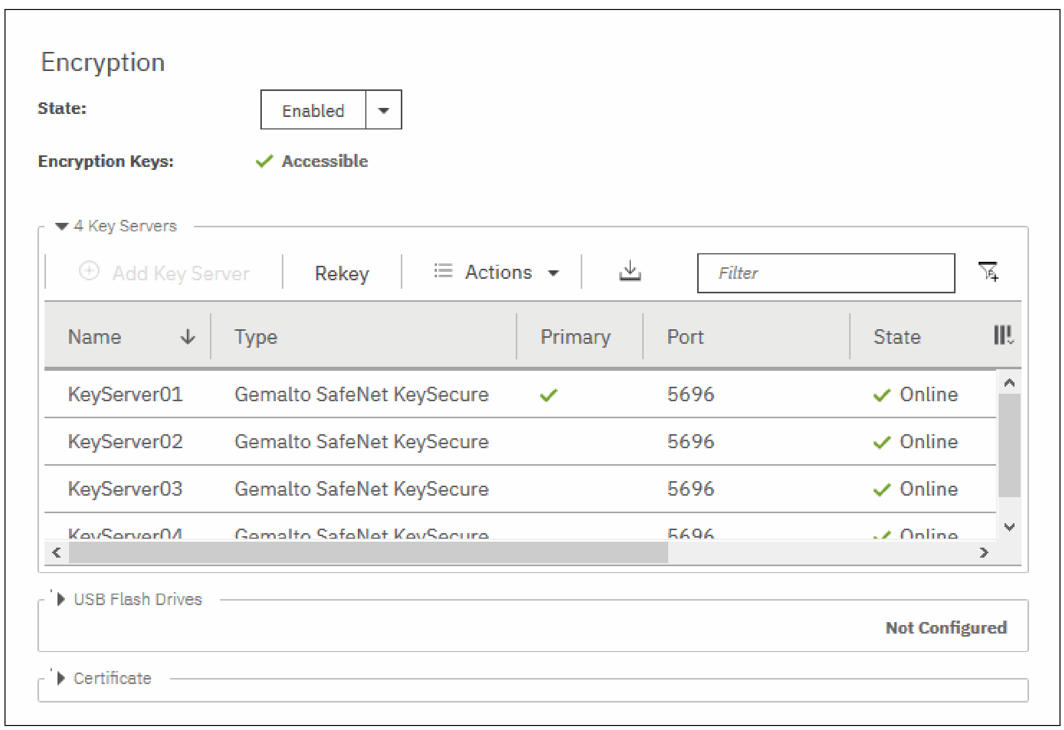


Figure 12-49 Encryption enabled with 4 SafeNet KeySecure key servers

12.4.4 Enabling encryption using both providers

IBM Spectrum Virtualize allows parallel use of both USB flash drive and one type of key server (SKLM or SafeNet KeySecure) as encryption key providers. It is possible to configure both providers in a single run of encryption enable wizard. To perform this configuration, the system must meet requirements of both key server (SKLM of SafeNet KeySecure) and USB flash drive encryption key providers.

Note: Make sure that the key management server functionality is fully independent from an encrypted storage that has encryption 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 code V8.1 and later supports up to four key server objects that are defined in parallel.

Before you start to enable encryption by using both USB flash drives and a key servers, confirm the requirements that are described in section 12.4.2, “Enabling encryption using USB flash drives” on page 622, and 12.4.3, “Enabling encryption using key servers” on page 627.

To enable encryption by using a key server and USB flash drive, complete the following steps:

1. Ensure that you have service IPs 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 12-50.

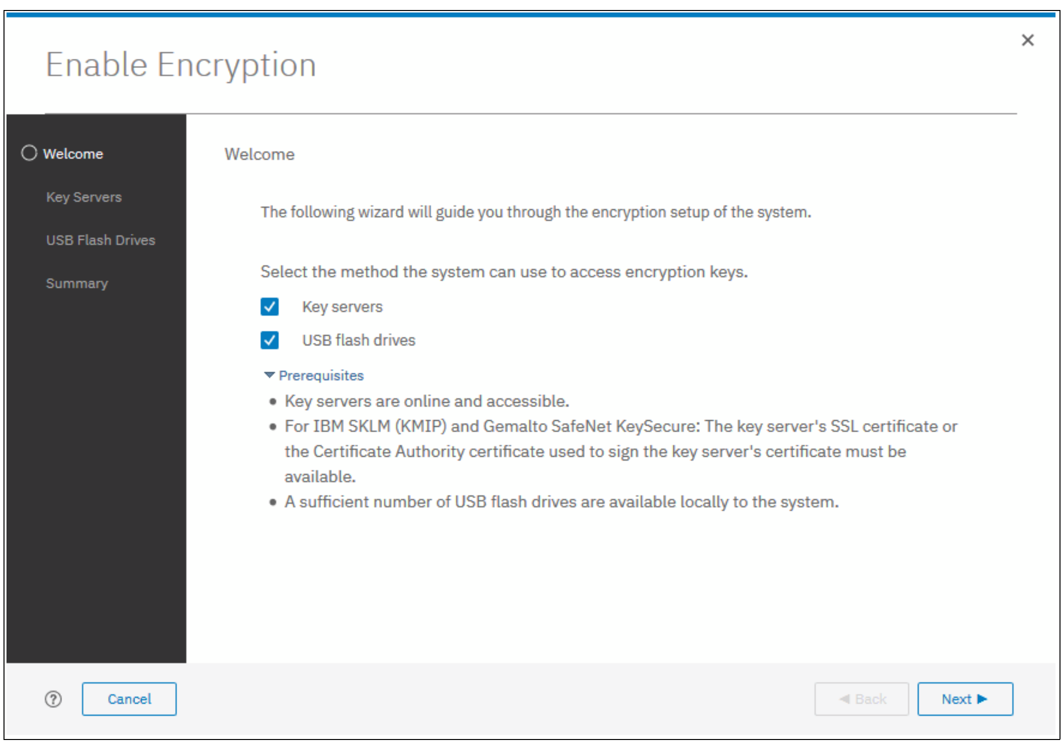


Figure 12-50 Selecting Key servers and USB flash drives in the Enable Encryption wizard

3. The wizard moves to the Key Server Types window, as shown in Figure 12-51. Then, select the key server type that manages the encryption keys.

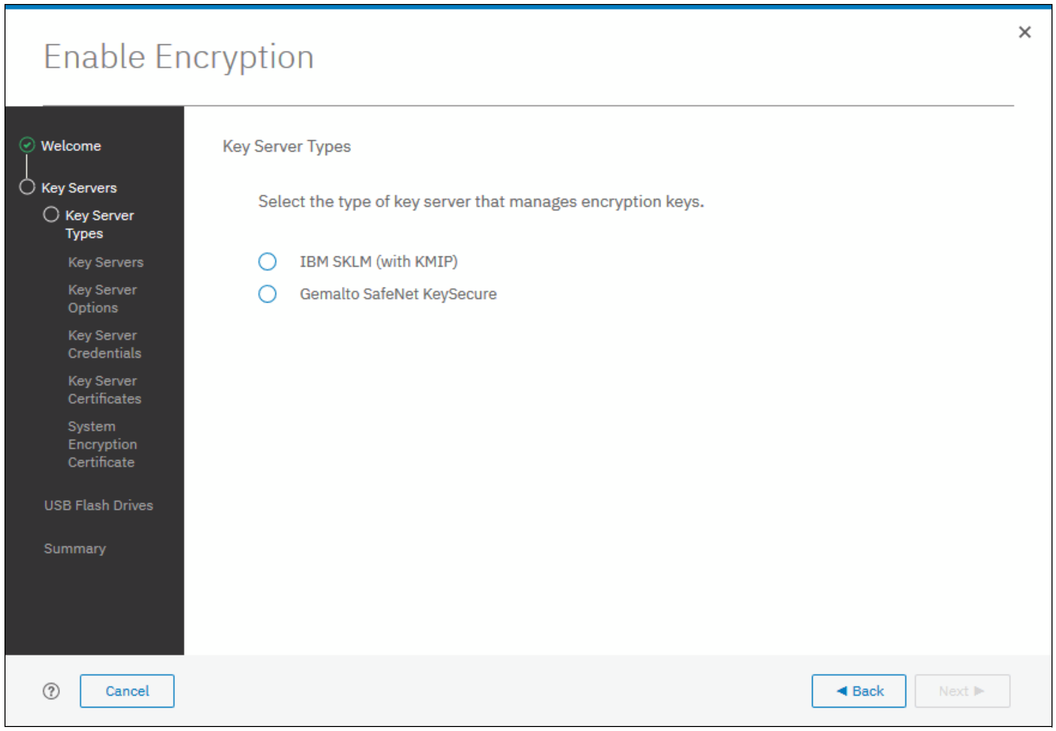


Figure 12-51 Selecting the key server type

The next windows that are displayed are the same as shown in 12.4.3, “Enabling encryption using key servers” on page 627, depending on the type of key server selected.

When the key servers details are entered, the USB flash drive encryption configuration is displayed. In this step, master encryption key copies are stored in the USB flash drives. If fewer than three drives are detected, the system requests plugging in more USB flash drives, as shown on Figure 12-52. You cannot proceed until the required minimum number of USB flash drives is detected by the system.

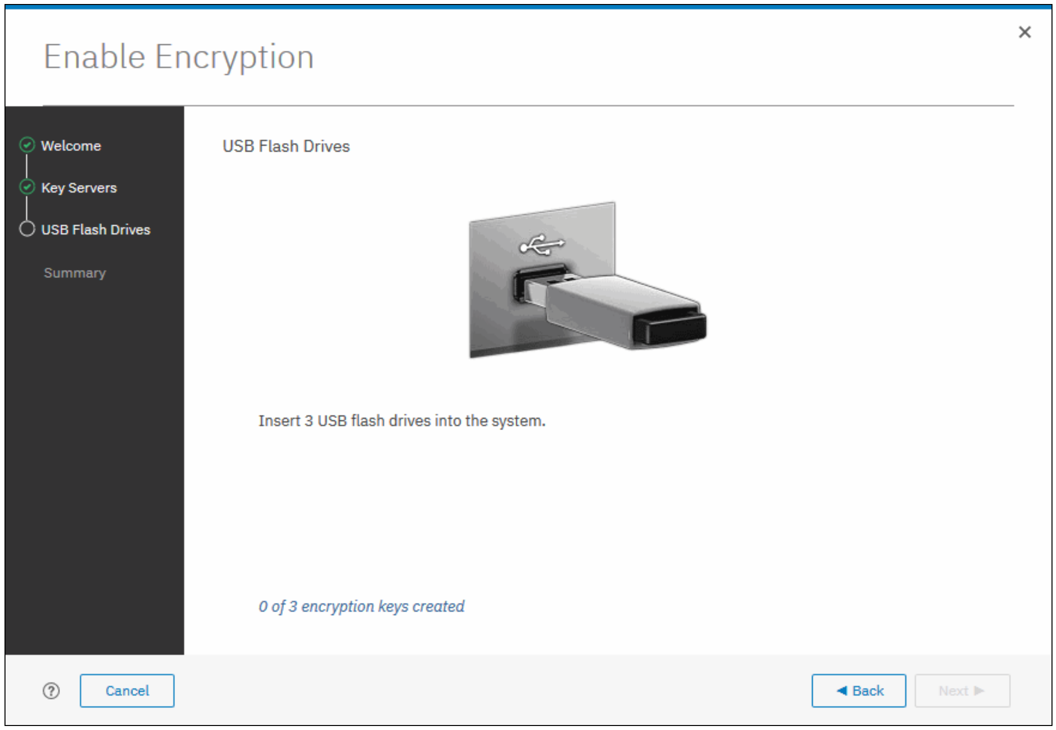


Figure 12-52 Prompt to insert USB flash drives

After at least three USB flash drives are detected, the system writes master access key to each of the drives. Note that the system attempts to write the encryption key to any flash drive it detects. Therefore, it is crucial to maintain physical security of the system during this procedure. After the keys are successfully copied to at least three USB flash drives, the system displays a window, as shown on Figure 12-53.

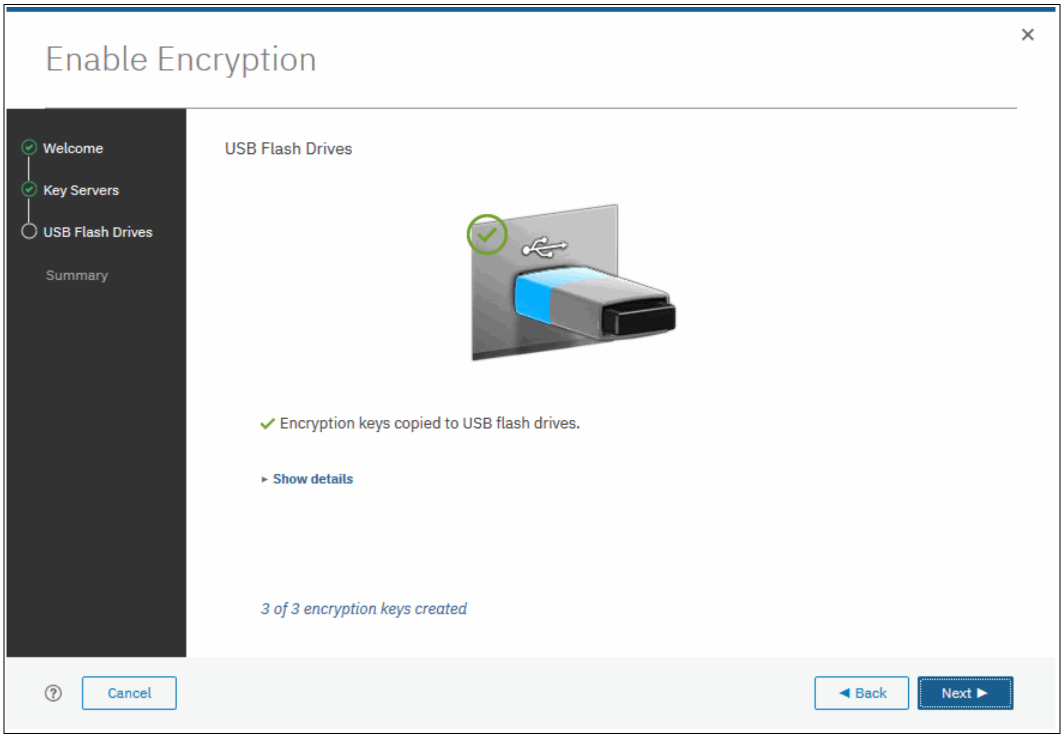


Figure 12-53 Master Access Key successfully copied to USB flash drives

4. After copying encryption keys to USB flash drives, the next window is shown with the summary of the configuration that is implemented on the system (see Figure 12-54 on page 646). Click **Finish** to create the key server object and finalize the encryption enablement.

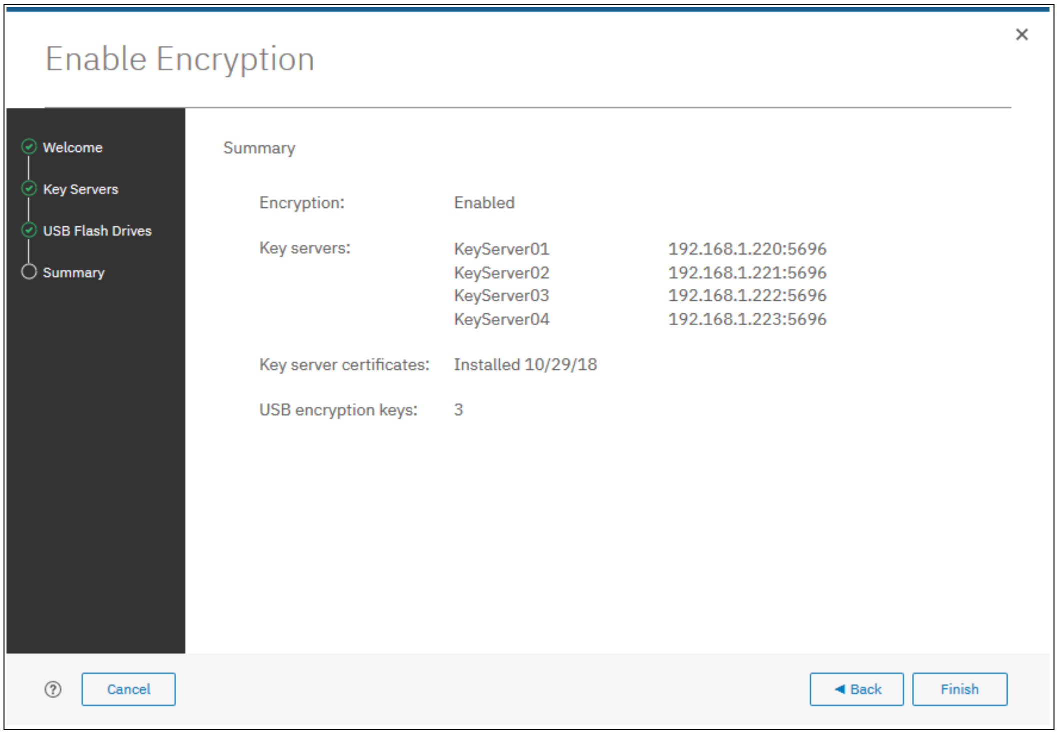


Figure 12-54 Encryption configuration summary in two providers scenario

5. 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 (see Figure 12-55).

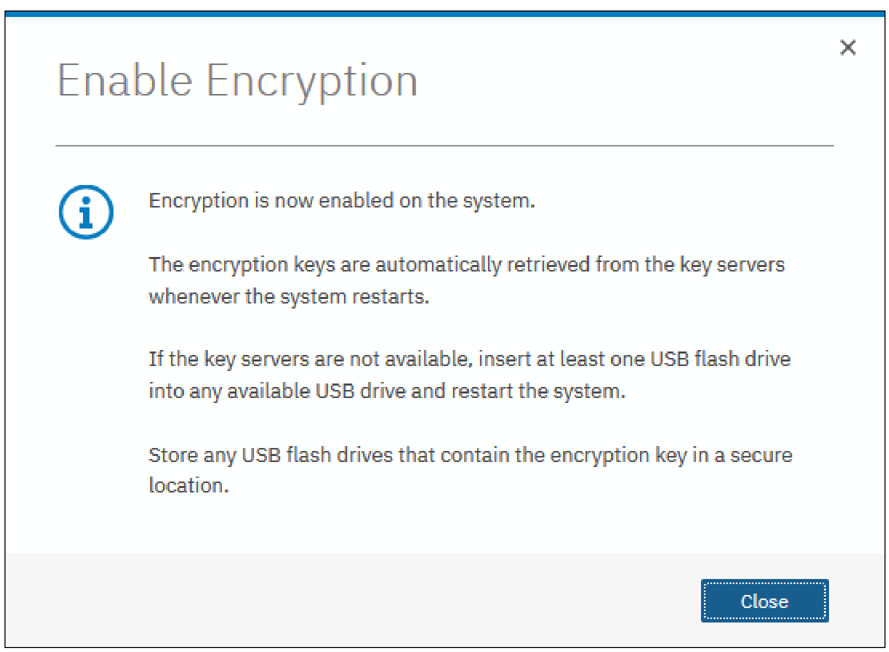


Figure 12-55 Encryption enabled message using both encryption key providers

6. You can confirm that encryption is enabled and verify which key providers are in use by going to **Settings** → **Security** → **Encryption**, as shown in Figure 12-56. Note the state *Online* of key servers and state *Validated* of USB ports where USB flash drives are inserted to make sure they are properly configured.

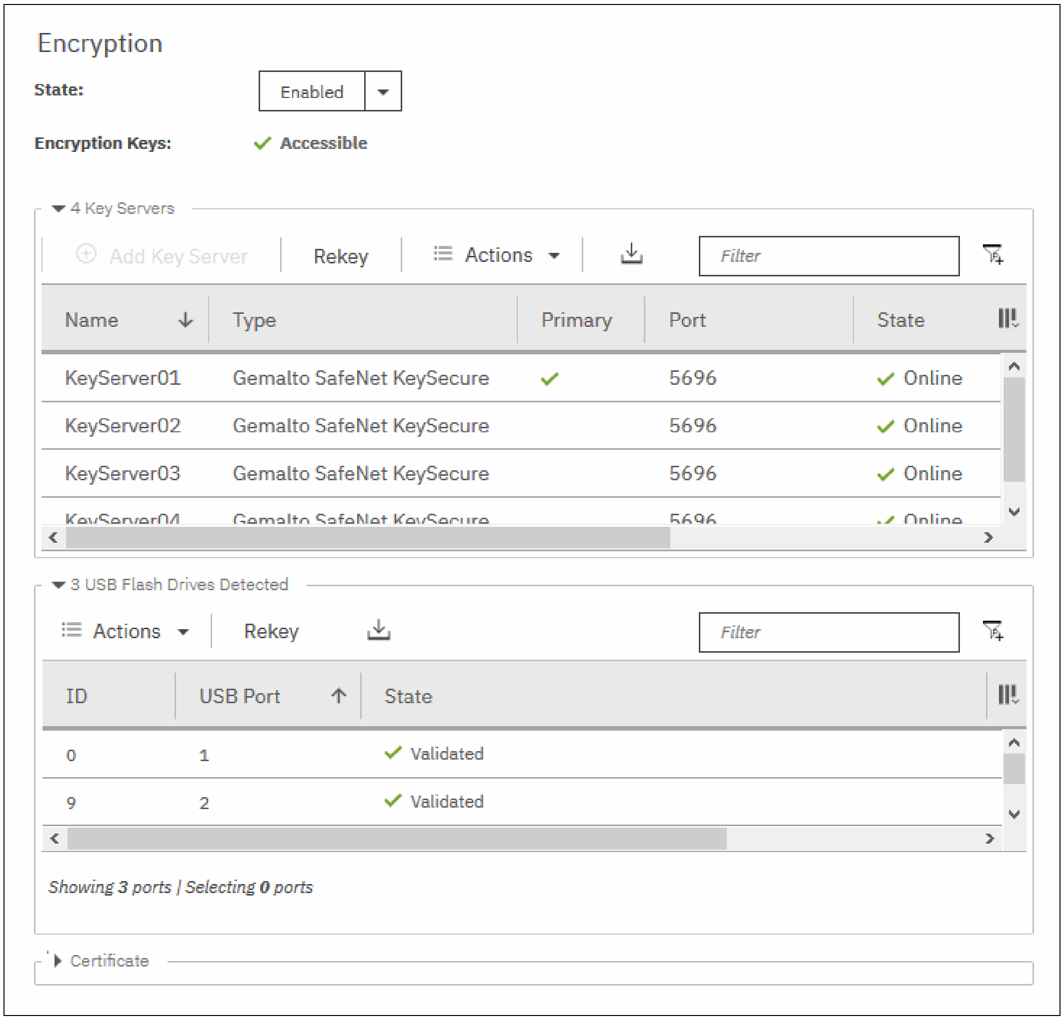


Figure 12-56 Encryption enabled with both USB flash drives and key servers

12.5 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 IBM Spectrum Virtualize code version earlier than V7.8.0, you must rekey the master encryption key before you can enable second encryption provider when you upgrade to code version V8.1 or later.

12.5.1 Adding key servers as a second provider

If the storage system is configured with the USB flash drive provider, it is possible to configure SKLM or SafeNet KeySecure servers as a second provider. To enable key servers as a second provider, complete the following steps:

1. Go to **Settings** → **Security** → **Encryption**. Expand the Key Servers section and click **Configure**, as shown in Figure 12-57. To enable 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.

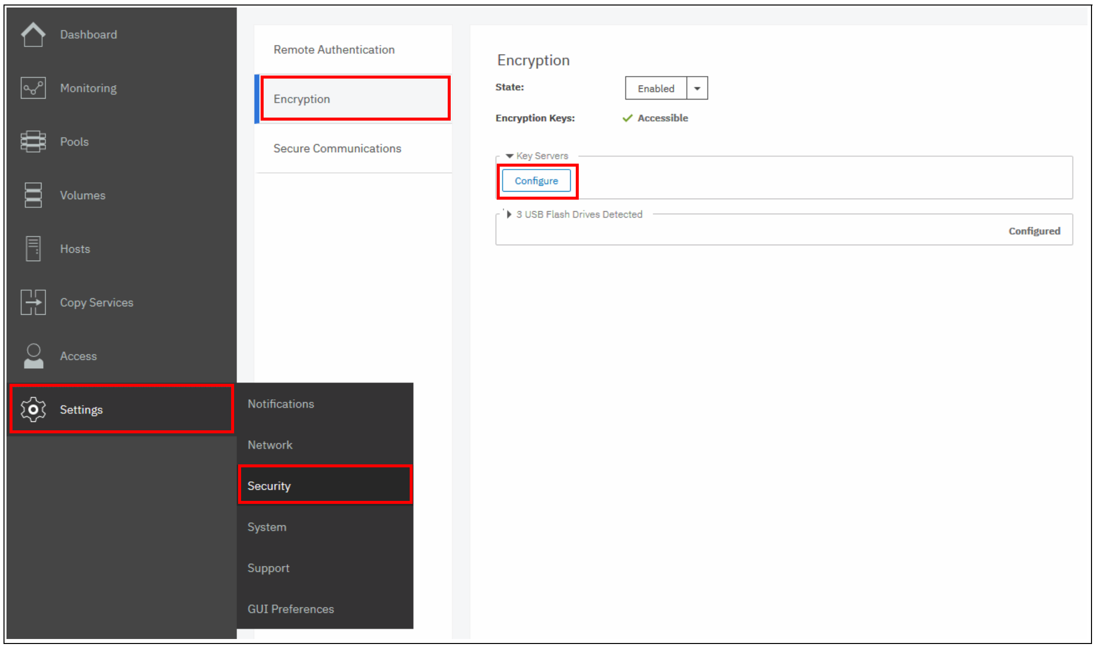


Figure 12-57 Enable key servers as a second provider

2. Complete the steps that are required to configure the key server provider, as described in 12.4.3, “Enabling encryption using key servers” on page 627. 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 12-58 on page 649.

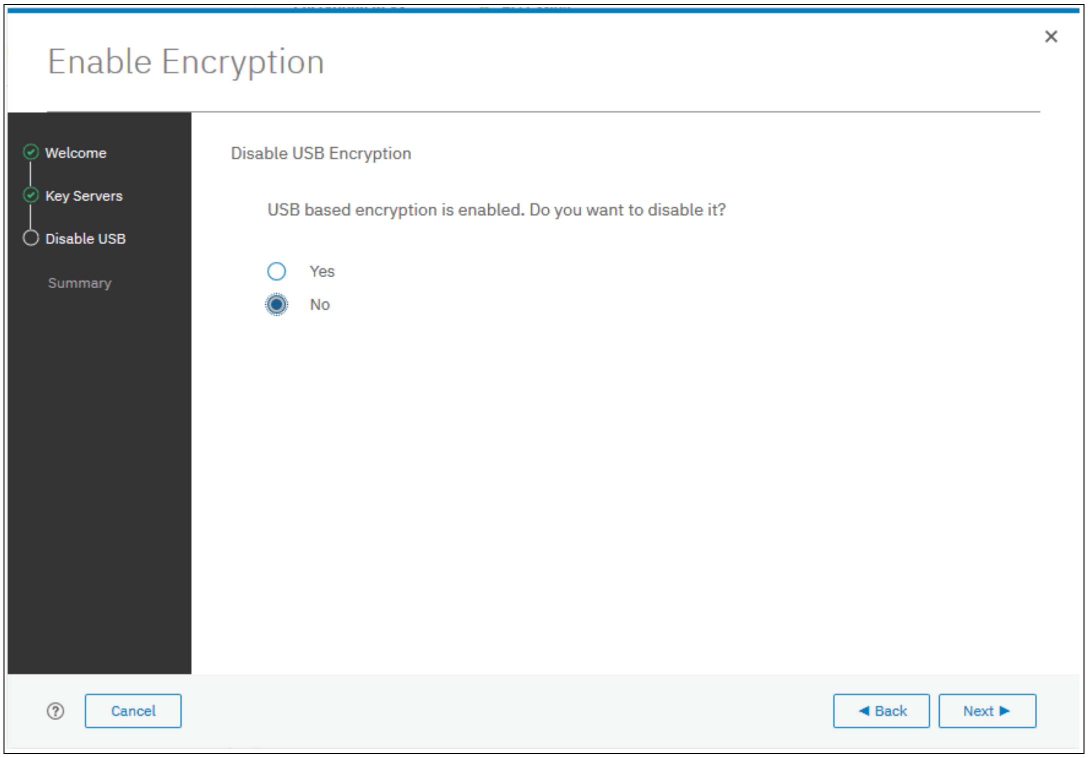


Figure 12-58   Do not disable USB flash drive encryption key provider

This choice is confirmed on the summary window before the configuration is committed, as shown in Figure 12-59.

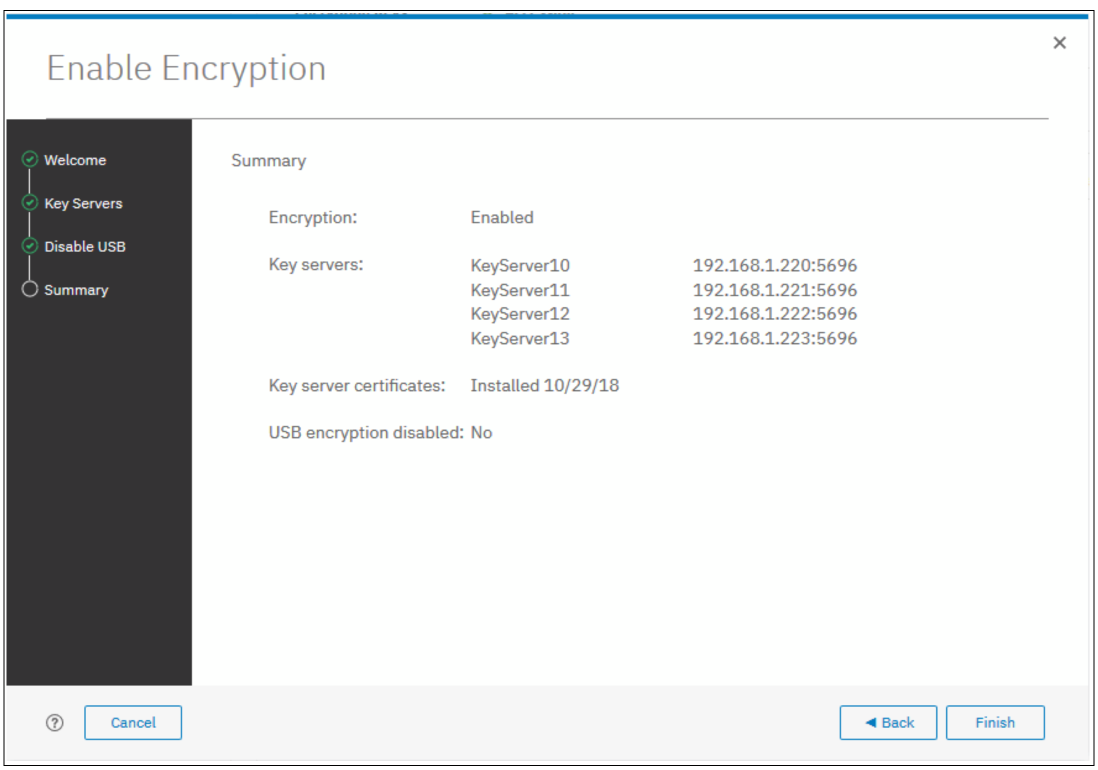


Figure 12-59   Configuration summary before committing

3. After you click **Finish**, the system configures keys servers as a second encryption key provider. Successful completion of the task is confirmed by a message, as shown in Figure 12-60. Click **Close**.

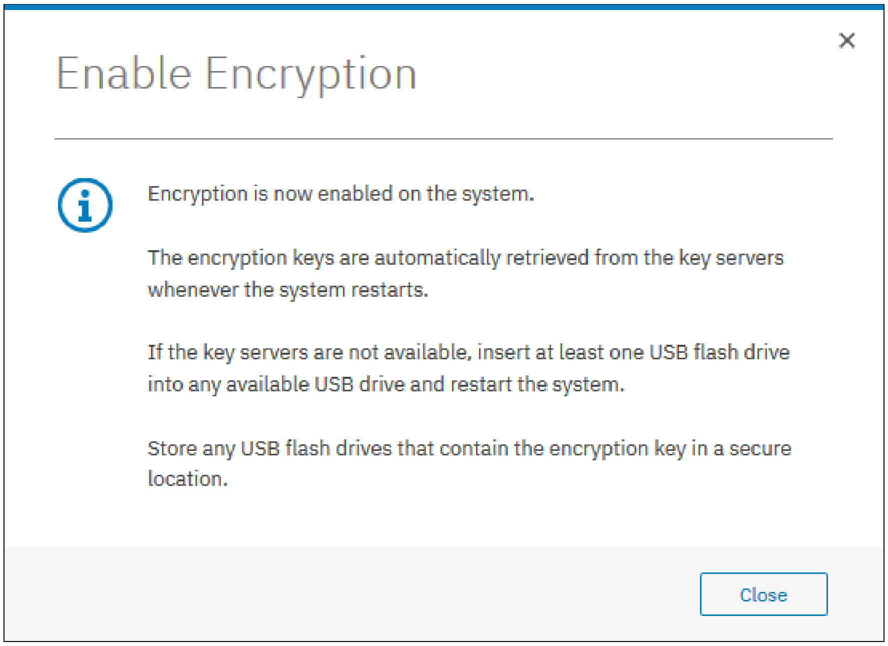


Figure 12-60 Confirmation of successful configuration of two encryption key providers

4. You can confirm that encryption is enabled and verify which key providers are in use by going to **Settings** → **Security** → **Encryption**, as shown in Figure 12-61. Note the `Online` state of key servers and `Validated` state of USB ports where USB flash drives are inserted to make sure they are properly configured.

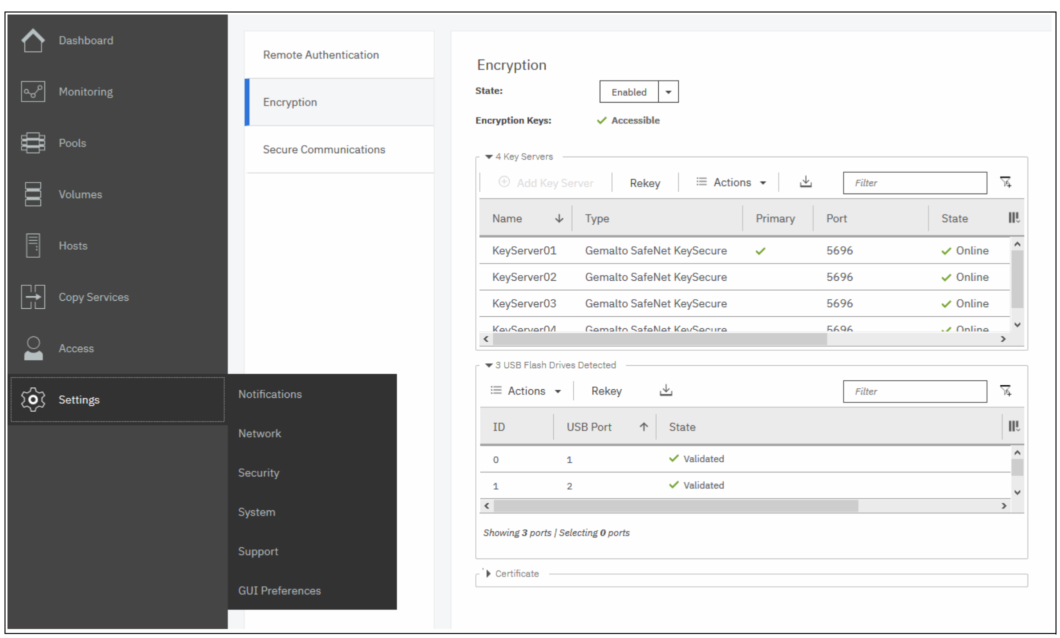


Figure 12-61 Encryption enabled with two key providers available

12.5.2 Adding USB flash drives as a second provider

If the storage system is configured with an SKLM or 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. Go to **Settings** → **Security** → **Encryption**. Expand the USB Flash Drives section and click **Configure**, as shown in Figure 12-62. To enable USB flash drives as a second provider, the system must access key servers with the current master access key.

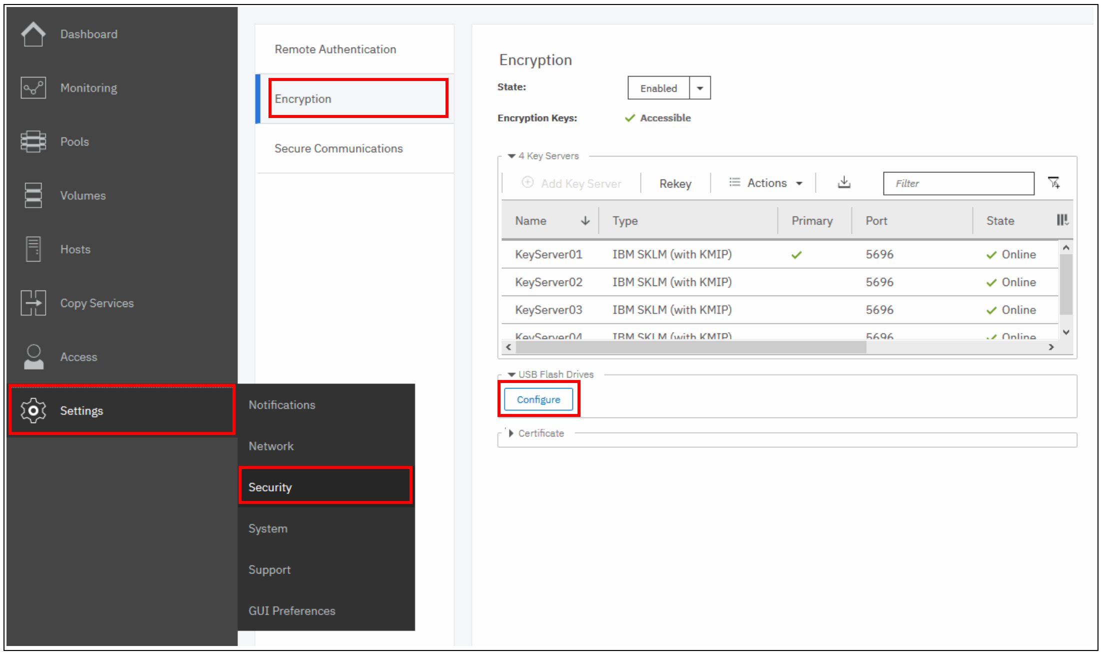


Figure 12-62 Enable USB flash drives as a second encryption key provider

2. After you click **Configure**, you are presented with a wizard similar to the one that is described in 12.4.2, “Enabling encryption using USB flash drives” on page 622. Note that 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, as shown in Figure 12-63.

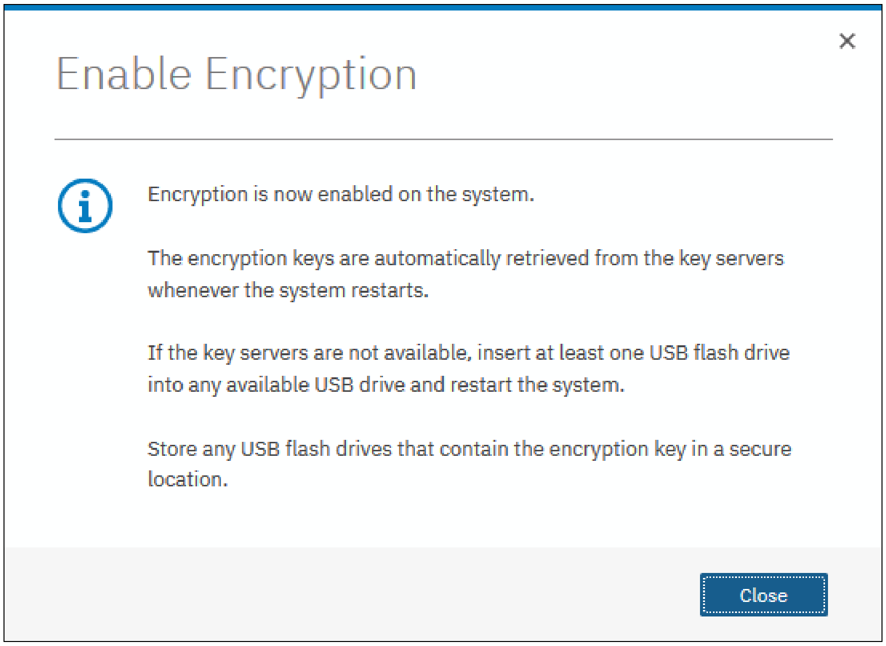


Figure 12-63 Confirmation of successful configuration of two encryption key providers

3. You can confirm that encryption is enabled and verify which key providers are in use by going to **Settings** → **Security** → **Encryption**, as shown in Figure 12-64. Note the state `Online` state of key servers and `Validated` state of USB ports where USB flash drives are inserted to make sure they are properly configured.

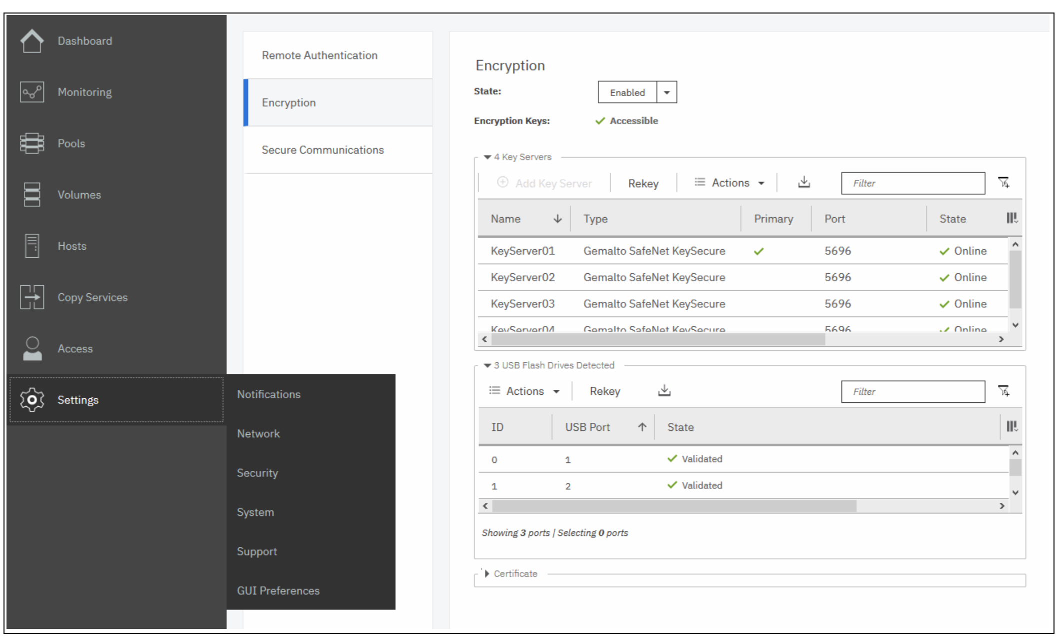


Figure 12-64 Encryption enabled with two key providers available

12.6 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 allows migration from configuration using only USB flash drive provider to key servers provider, and vice versa.

If you want to migrate from one key server type to another (for example, migrating from SKLM to SafeNet KeySecure or vice-versa), direct migration is not possible. In this case, it is required first to migrate from the current key server type to a USB flash drive, and then migrate to the other type of key server.

12.6.1 Changing from USB flash drive provider to encryption key server

The system is designed to facilitate changing from USB flash drives encryption key provider to encryption key server provider. If you follow the steps that are described in 12.5.1, “Adding key servers as a second provider” on page 648, but when completing step 2 on page 648, select **Yes** instead of **No** (see Figure 12-65). This action causes de-activation of the USB flash drives provider, and the procedure completes with only key servers configured as key provider.

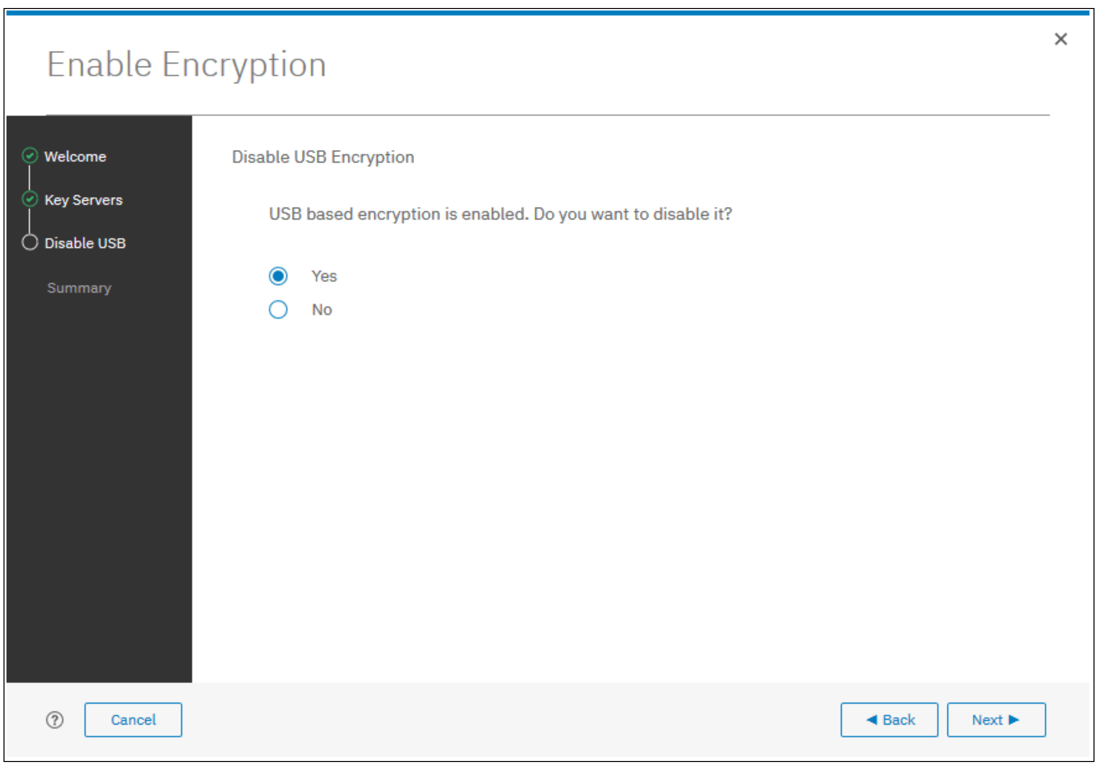


Figure 12-65 Disable USB flash drive provider while changing to SKLM provider

12.6.2 Changing from encryption key server to USB flash drive provider

Change in the other direction (that is, from the use of encryption key servers provider to 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 described in 12.5.2, “Adding USB flash drives as a second provider” on page 651.

Then, issue the following command in the CLI:

```
chencryption -usb validate
```

To make sure that USB drives contain the correct master access key, disable the encryption key server provider by issuing the following command:

```
chencryption -keyserver disable
```

This command disables the encryption key server provider, which effectively migrates your system from encryption key server to USB flash drive provider.

12.6.3 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.

If you want to migrate from one type of key server to another, you first need to 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 SKLM key servers (as shown in Figure 12-66) to SafeNet KeySecure servers.

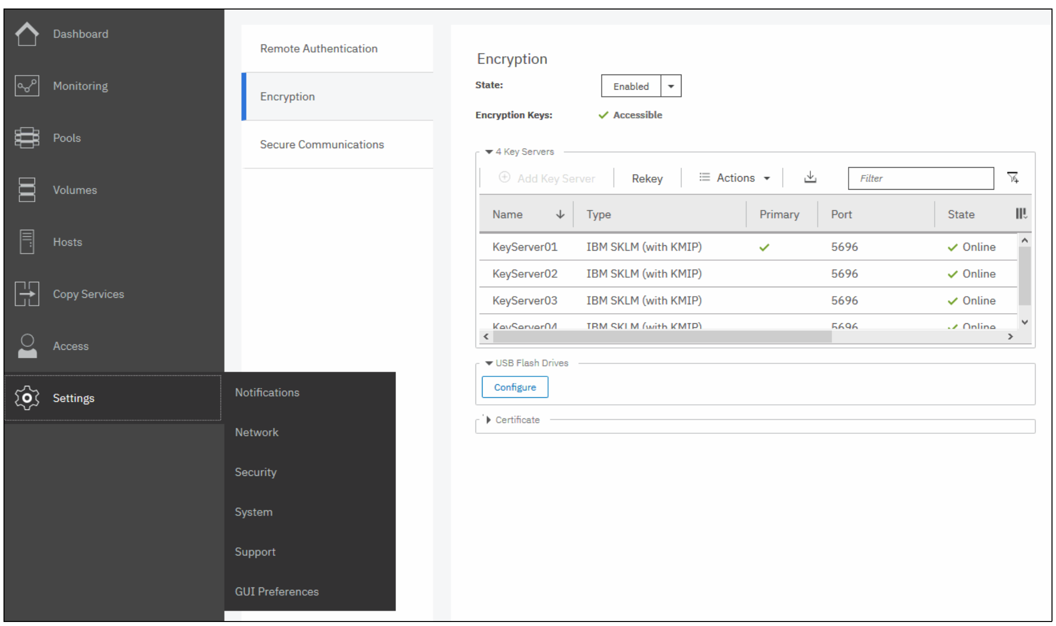


Figure 12-66 IBM Storwize V7000 encryption configured with IBM SKLM servers

Complete the following steps to migrate to Gemalto SafeNet KeySecure:

1. Migrate from key server encryption to USB flash drives encryption, as described in 12.6.2, “Changing from encryption key server to USB flash drive provider” on page 653. After this step, only USB flash drives encryption are configured, as shown in Figure 12-67 on page 655.

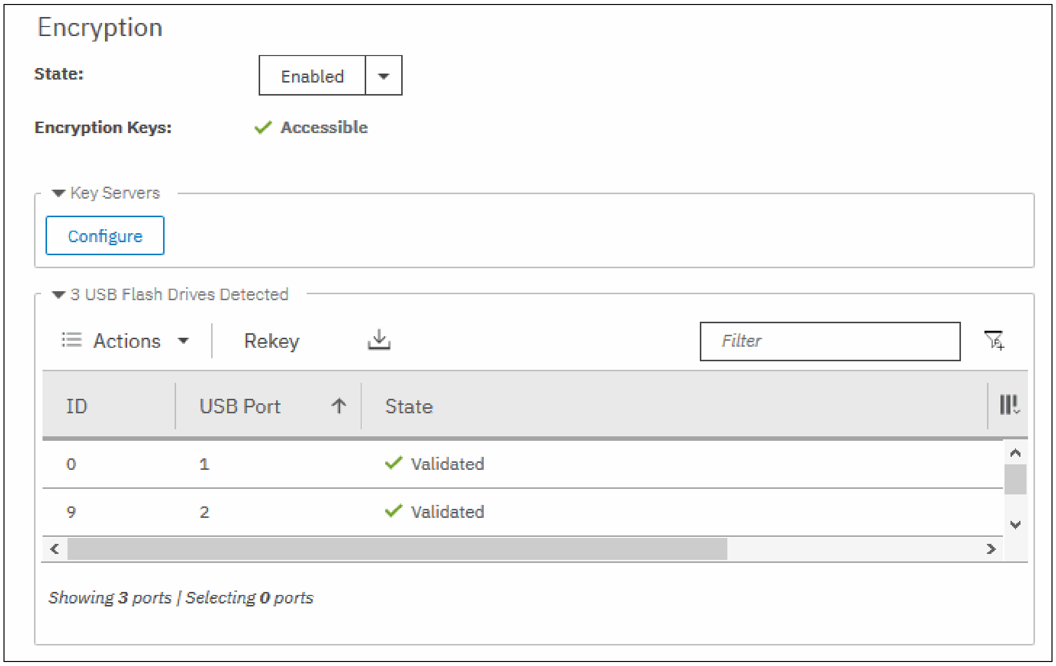


Figure 12-67 IBM Storwize V7000 encryption configured with USB Flash Drives

2. Migrate from USB flash drives encryption to the other key server type encryption (in this example, Gemalto SafeNet KeySecure), following the steps that are described in 12.6.1, “Changing from USB flash drive provider to encryption key server” on page 653. After completing this step, the other key server type are configured as encryption provider in IBM Spectrum Virtualize, as shown in Figure 12-68.

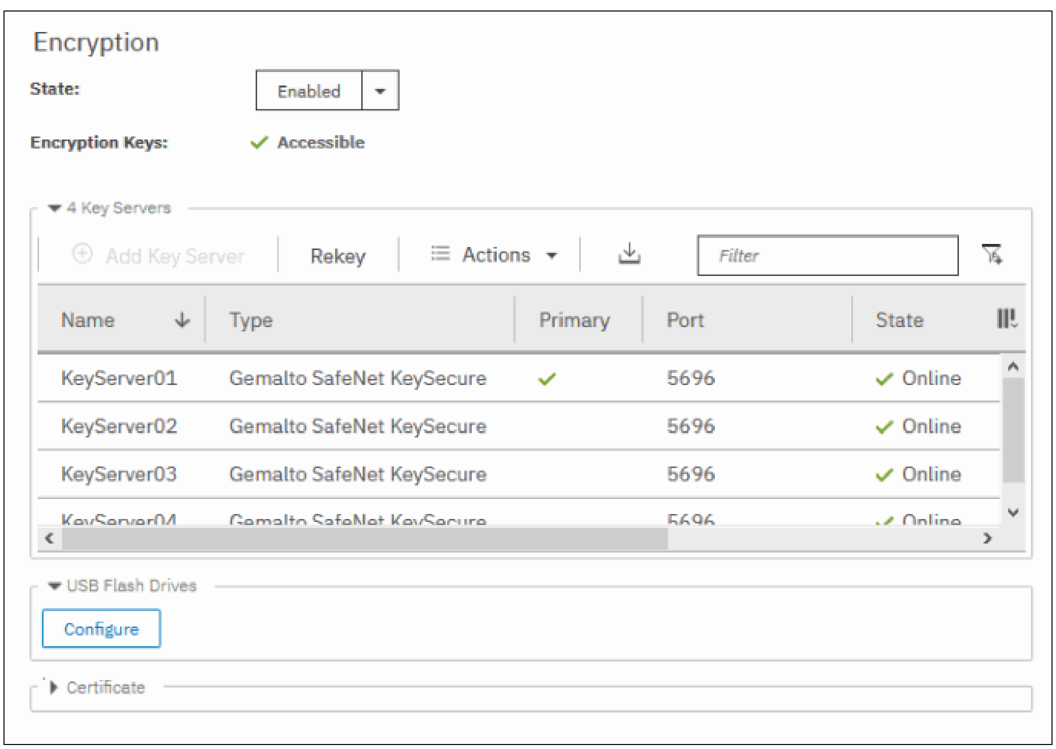


Figure 12-68 IBM Storwize V7000 encryption configured with SafeNet KeySecure

12.7 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 kept on the USB flash drives or by losing all SKLM servers), you can recover from this situation by disabling the provider to which you lost the access. To disable the unavailable provider, you must have access to a valid master access key on the remaining provider.

If you have lost access to the encryption key server provider, issue the following command:

```
chencryption -keyserver disable
```

If you have lost access to the USB flash drives provider, issue 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 12.5, “Configuring more providers” on page 647.

> **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 protected by the master access key.

12.8 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 RAID is available in IBM Spectrum Virtualize code 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, 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.

12.8.1 Encrypted pools

For more information about how to open the Create Pool window, see Chapter 6, “Storage pools” on page 191. After encryption is enabled, any new pool is created by default as encrypted, as shown in Figure 12-69.

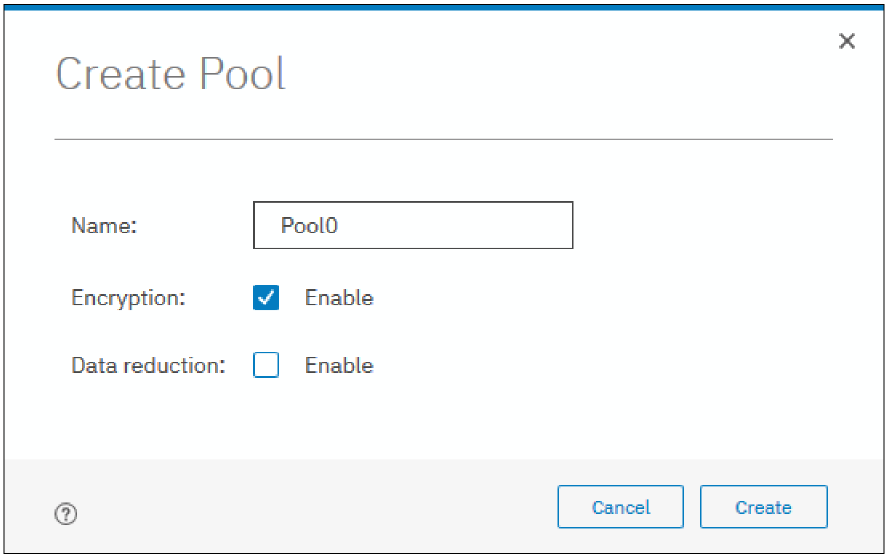


Figure 12-69 Create Pool window basic

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 pool encryption status. Click **Pools** → **Pools**, and then, click **Actions** → **Customize Columns** → **Encryption**, as shown in Figure 12-70.

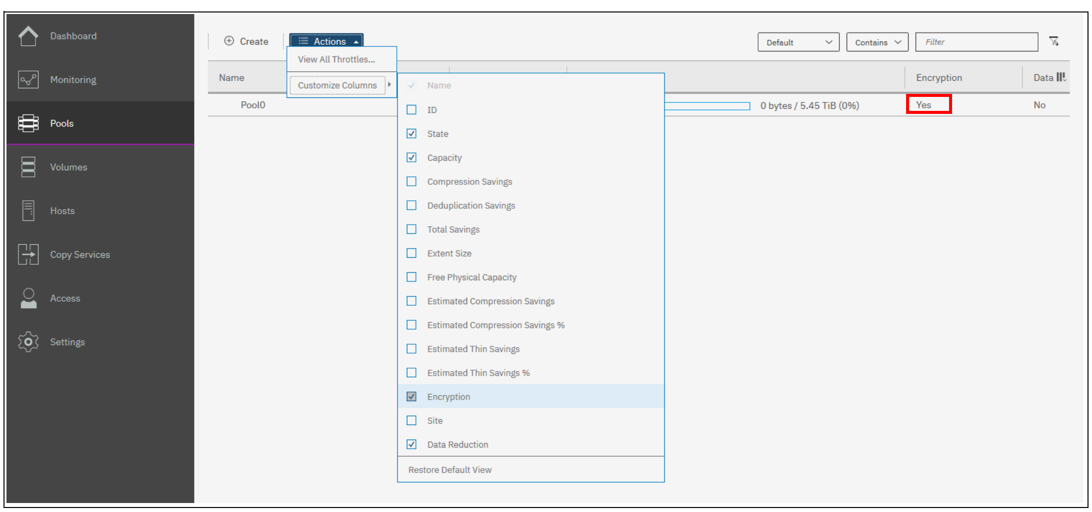


Figure 12-70 Pool encryption state

If you create an unencrypted pool but you 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.

More information about how to add encrypted storage to encrypted pools is described next. You can mix and match storage encryption types in a pool. Figure 12-71 shows an example of an encrypted pool that contains storage by using different encryption methods.

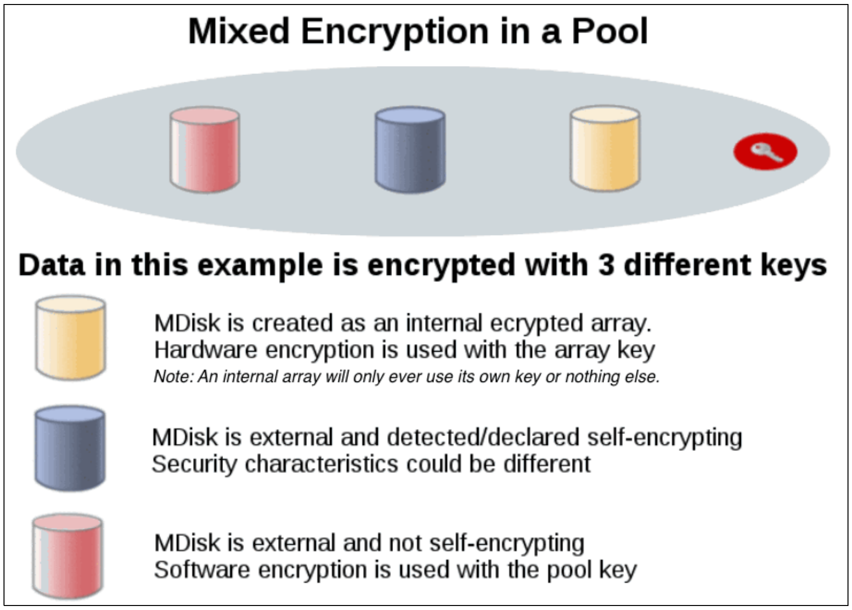


Figure 12-71 Mix and match encryption in a pool

12.8.2 Encrypted child pools

For more information about how to open the Create Child Pool window, see Chapter 6, "Storage pools" on page 191. 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 12-72.

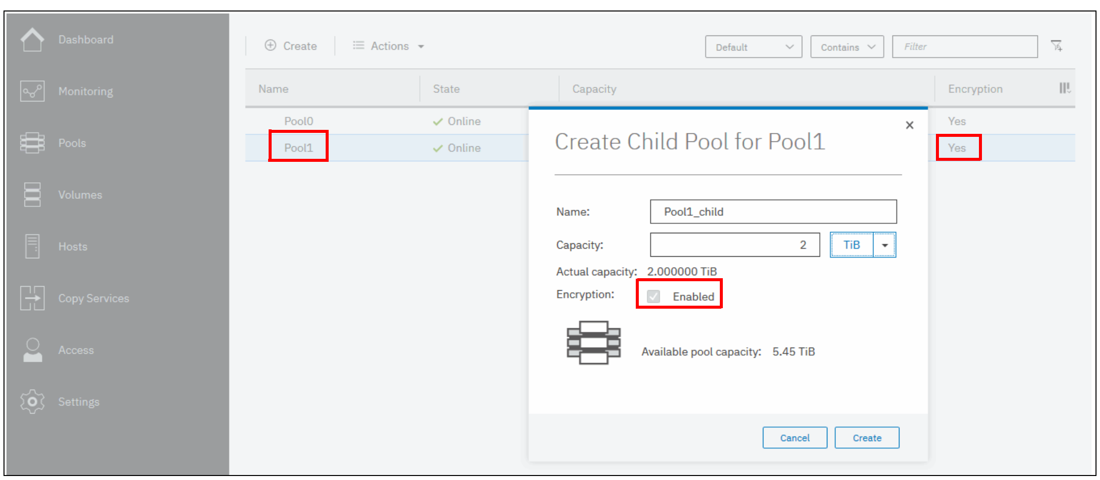


Figure 12-72 Create a child pool of an encrypted parent pool

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 array is in the unencrypted pool, when you try to create a child pool and select the option to set as encrypted, it is created as unencrypted.
- All V7000 Storwize 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.

If you modify Pools view, you see the encryption status of child pools, as shown in Figure 12-73. The example shows an encrypted child pool with non-encrypted parent pool.

.

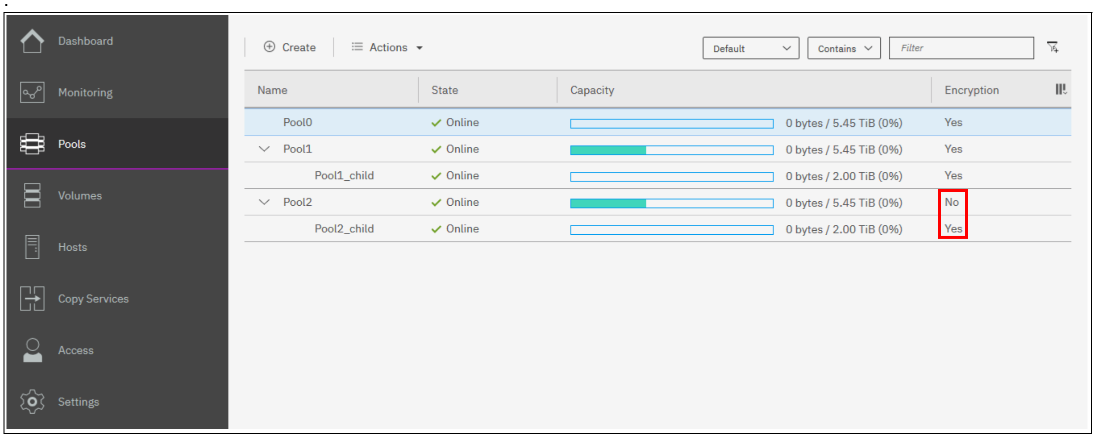


Figure 12-73 Child pool encryption state

12.8.3 Encrypted arrays

For more information about how to add internal storage to a pool, see Chapter 6, “Storage pools” on page 191. After encryption is enabled, all newly built arrays are hardware encrypted by default. In this case, the graphical user interface (GUI) does not allow you to create an unencrypted array. To create an unencrypted array, the command-line interface (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 using CLI

```
IBM_Storwize:ITSO-V7k:superuser>svctask mkarray -drive 6:4 -level raid1 -sparegoal
0 -strip 256 -encrypt no Pool2
MDisk, id [2], successfully created
IBM_Storwize:ITSO-V7k:superuser>
```

Note: It is not possible to add unencrypted arrays to an encrypted pool.

You can customize MDisks by Pools view to show array encryption status. Click **Pools** → **MDisk by Pools**, and then, click **Actions** → **Customize Columns** → **Encryption**. You also can right-click the table header to customize columns and select **Encryption**, as shown in Figure 12-74.

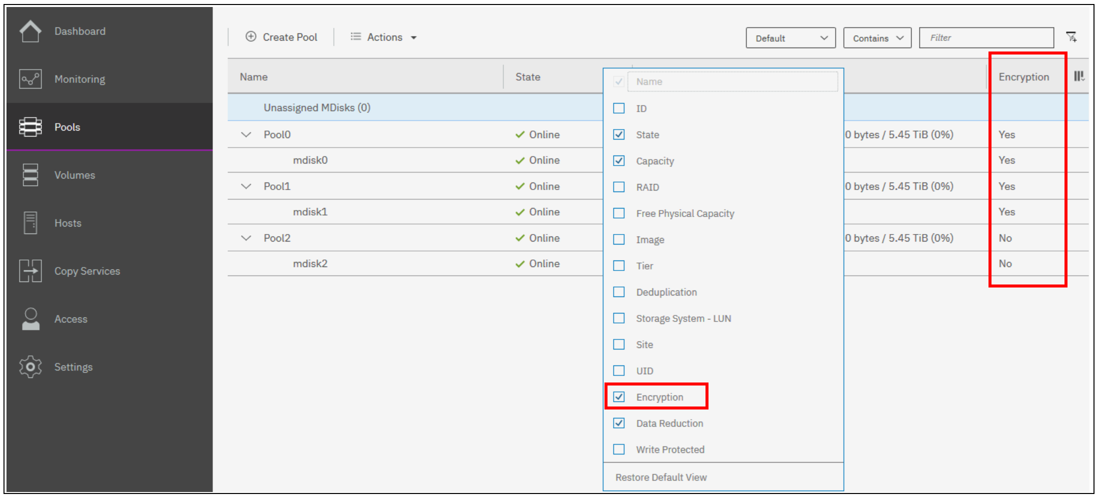


Figure 12-74 Array encryption state

You can also check the encryption state of an array by reviewing its drives in **Pools** → **Internal Storage** view. The internal drives that are associated with an encrypted array are assigned an encrypted property that can be seen, as shown in Figure 12-75.

Drive ID ↑	Capacity	Use	Status	MDisk Name	Slot ID	Encrypted
0	5.46 TiB	Member	✓ Online	mdisk0	11	✓
1	5.46 TiB	Spare	✓ Online		1	
2	5.46 TiB	Member	✓ Online	mdisk0	9	✓
3	5.46 TiB	Member	✓ Online	mdisk1	2	✓
4	5.46 TiB	Member	✓ Online	mdisk2	6	
5	5.46 TiB	Member	✓ Online	mdisk1	10	✓
6	5.46 TiB	Member	✓ Online	mdisk2	5	

Figure 12-75 Drive encryption state

12.8.4 Encrypted MDisks

For more information about how to add external storage to a pool, see Chapter 6, “Storage pools” on page 191. 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 or declared 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 show the object encryption state by clicking **Pools** → **MDisk by Pools**. Figure 12-76 shows an example in which self-encrypting MDisk is in an unencrypted pool.

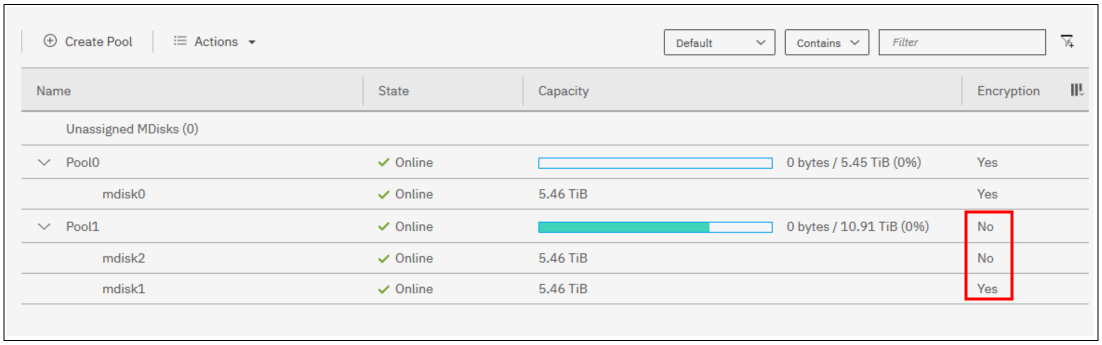


Figure 12-76 MDisk encryption state

When working with MDisks encryption, take extra care when configuring 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 only marks disk space 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, it contains stale data from its previous life.

Another mistake that can occur is to misconfigure an external MDisk as self-encrypting, while in reality it is not self-encrypting. In that case, the data written to this MDisk is not encrypted by Storwize V7000 because Storwize V7000 is convinced that the MDisk encrypts the data. At the same time, the MDisk does not encrypt the data because it is not self-encrypting, so you end up with unencrypted data on an extent in an encrypted pool.

However, all data that is written to any MDisk that is a part of correctly configured encrypted pool is going to be encrypted.

Self-encrypting 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 waste of resources, such as CPU time. However, when used improperly, it might lead to unencrypted data at-rest.

To declare an MDisk as self-encrypting, select **Externally encrypted** when adding external storage in the **Assign Storage** view, as shown in Figure 12-77.

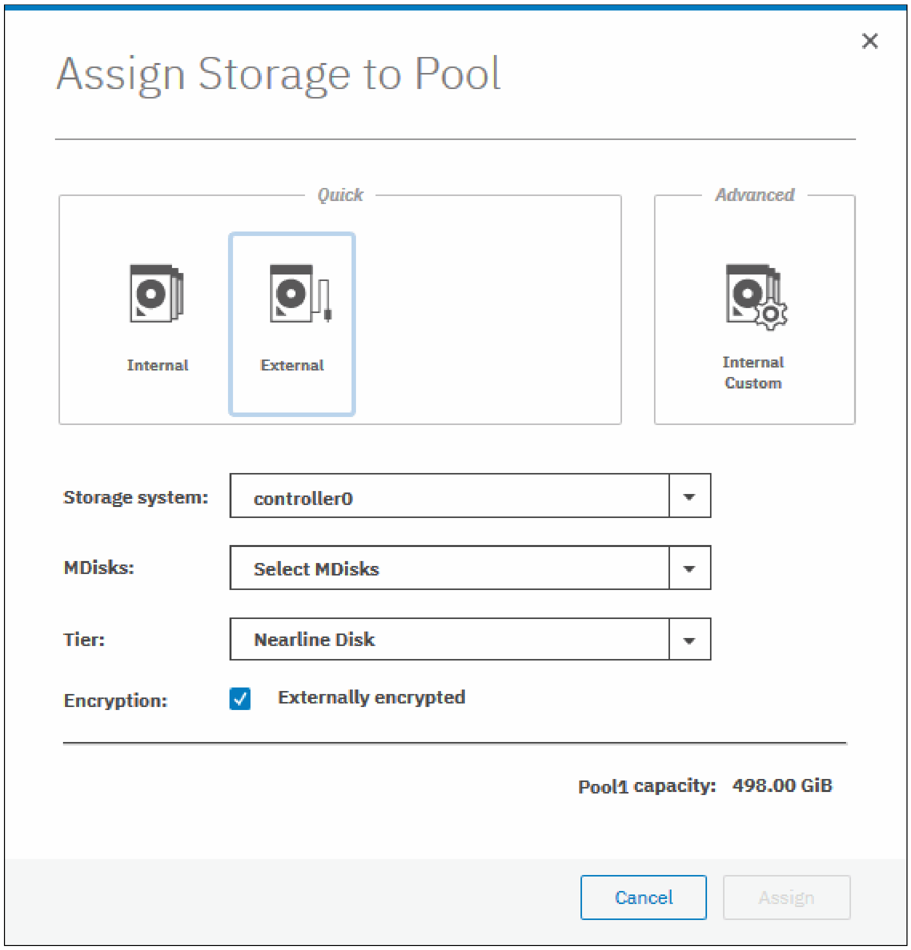
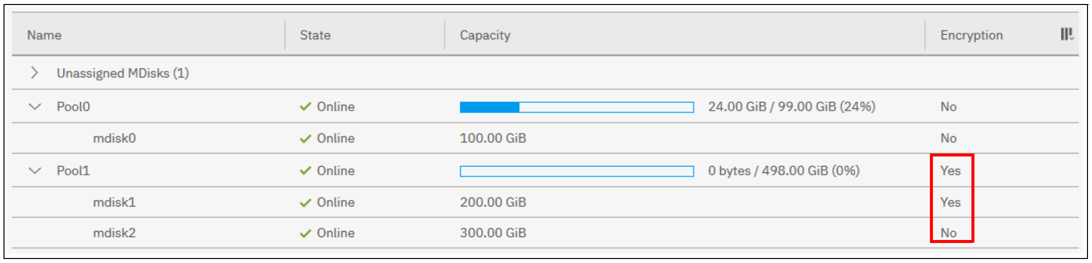


Figure 12-77 Declaring MDisk as externally encrypted

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. For these MDisks, the **Externally encrypted** option that is shown in Figure 12-77 is not selected. However, when added, they are still considered as self-encrypting.

Note: You can override the external encryption setting of a detected MDisk as self-encrypting and configure it as unencrypted by using the CLI command **`chmdisk -encrypt no`**. However, only do so if you plan to decrypt the data on the backend or if the backend uses inadequate data encryption.

To check whether an MDisk was detected or declared as self-encrypting, click **Pools** → **MDisk by Pools** and verify the information in the Encryption column, as shown in Figure 12-78.



Name	State	Capacity		Encryption
Unassigned MDisks (1)				
Pool0	Online		24.00 GiB / 99.00 GiB (24%)	No
mdisk0	Online	100.00 GiB		No
Pool1	Online		0 bytes / 498.00 GiB (0%)	Yes
mdisk1	Online	200.00 GiB		Yes
mdisk2	Online	300.00 GiB		No

Figure 12-78 MDisk self-encryption state

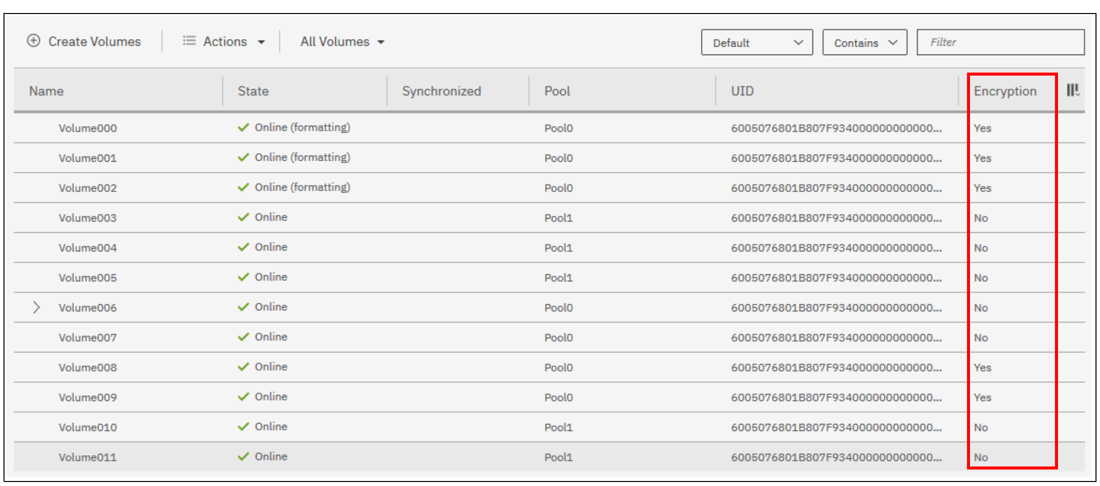
The value that is shown in the Encryption column shows the property of objects in respective rows. That means that in the configuration that is shown in Figure 12-78, `Pool1` is encrypted, so every volume created from this pool is encrypted. However, that pool is formed by two MDisks, out of which one is self-encrypting and one is not. Therefore, a value of `No` next to `mdisk2` does not imply that encryption of `Pool1` is in any way compromised. It only indicates that encryption of the data placed on `mdisk2` is done by using software encryption. Data that is placed on `mdisk1` 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 member of an unencrypted pool. However, you cannot change the self-encrypting attribute of an MDisk after it is added to an encrypted pool.

12.8.5 Encrypted volumes

For more information about how to create and manage volumes, see Chapter 7, “Volumes” on page 241. The encryption status of a volume depends on the pool encryption status. Volumes that are created in an encrypted pool are automatically encrypted.

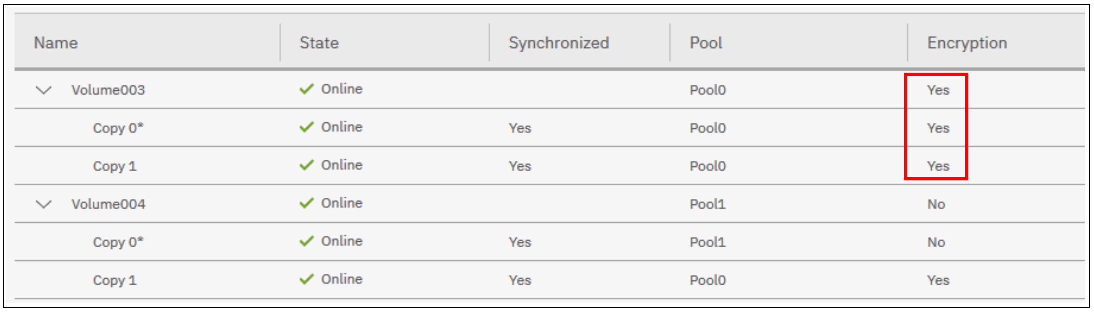
You can modify Volumes view to show if the volume is encrypted. Click **Volumes** → **Volumes**. Then, click **Actions** → **Customize Columns** → **Encryption** to customize the view to show volumes encryption status, as shown in Figure 12-79.



Name	State	Synchronized	Pool	UID	Encryption
Volume000	Online (formatting)		Pool0	6005076801B807F9340000000000000...	Yes
Volume001	Online (formatting)		Pool0	6005076801B807F9340000000000000...	Yes
Volume002	Online (formatting)		Pool0	6005076801B807F9340000000000000...	Yes
Volume003	Online		Pool1	6005076801B807F9340000000000000...	No
Volume004	Online		Pool1	6005076801B807F9340000000000000...	No
Volume005	Online		Pool1	6005076801B807F9340000000000000...	No
Volume006	Online		Pool0	6005076801B807F9340000000000000...	No
Volume007	Online		Pool0	6005076801B807F9340000000000000...	No
Volume008	Online		Pool0	6005076801B807F9340000000000000...	Yes
Volume009	Online		Pool0	6005076801B807F9340000000000000...	Yes
Volume010	Online		Pool1	6005076801B807F9340000000000000...	No
Volume011	Online		Pool1	6005076801B807F9340000000000000...	No

Figure 12-79 Volume view customization

A volume is reported as encrypted only if all the volume copies are encrypted, as shown in Figure 12-80.



Name	State	Synchronized	Pool	Encryption
Volume003	Online		Pool0	Yes
Copy 0*	Online	Yes	Pool0	Yes
Copy 1	Online	Yes	Pool0	Yes
Volume004	Online		Pool1	No
Copy 0*	Online	Yes	Pool1	No
Copy 1	Online	Yes	Pool0	Yes

Figure 12-80 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 12-81.

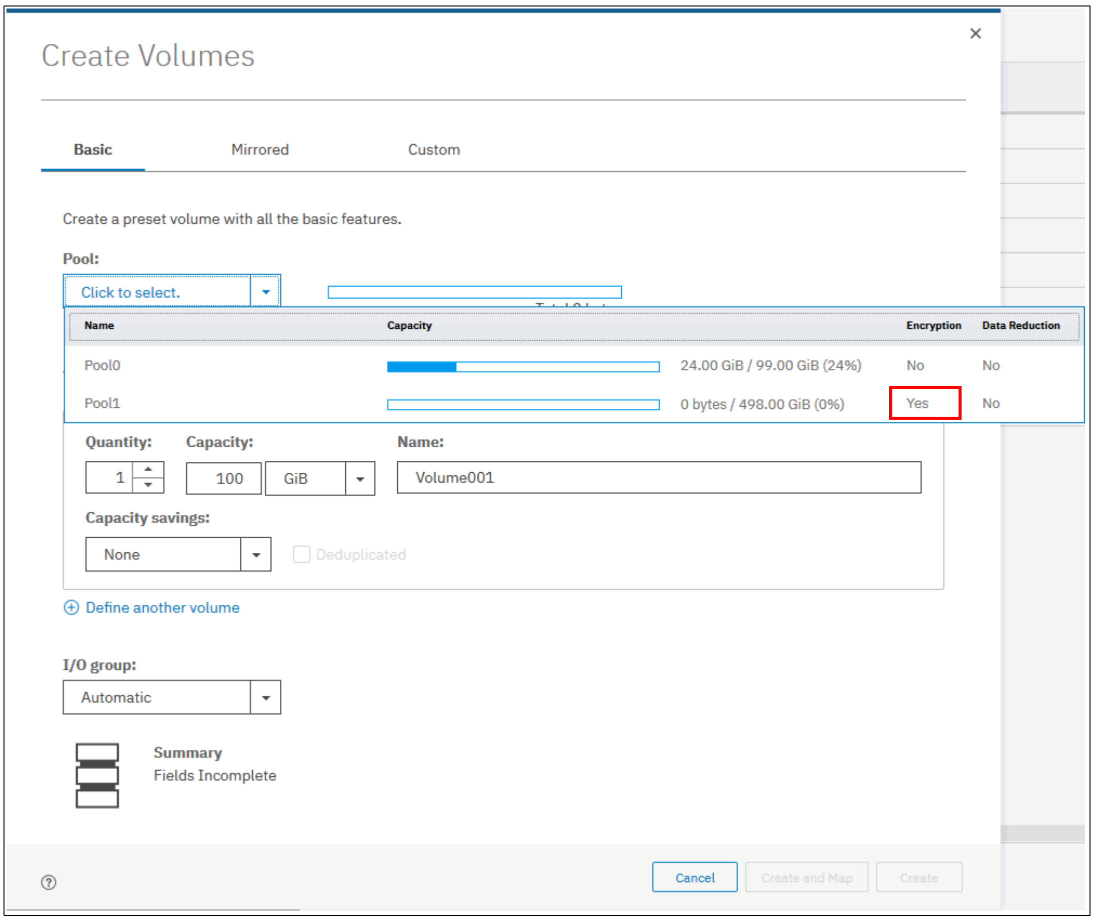


Figure 12-81 Create an encrypted volume by selecting an encrypted pool

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 7, “Volumes” on page 241.

12.8.6 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.

12.9 Rekeying an encryption-enabled system

Changing the master access key is a security requirement. *Rekeying* is the process of replacing 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 defined in your organization's security policy and whenever you suspect that the key might have been compromised.

If you have both USB and key server 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.

12.9.1 Rekeying using a key server

Ensure that all the configured key servers can be reached by the system and that service IPs are configured on all your nodes.

To rekey the master access key kept on the key server provider, complete the following steps:

1. Click **Settings** → **Security** → **Encryption**. Ensure that Encryption Keys shows that all configured SKLM servers are reported as **Accessible**, as shown in Figure 12-82 on page 666. Click **Key Servers** to expand the section.

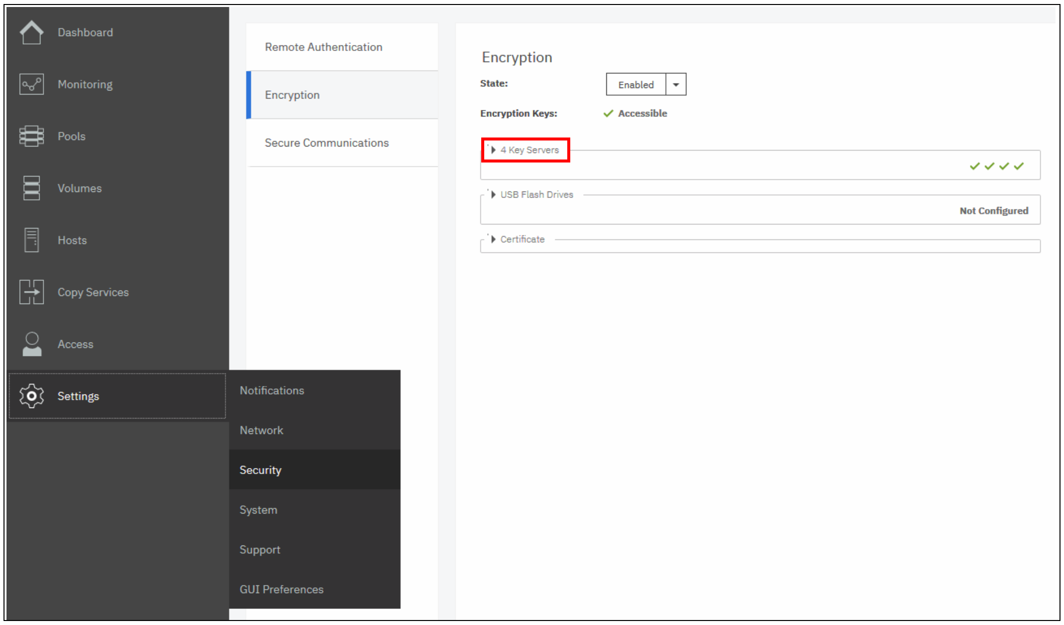


Figure 12-82 Locate Key Servers section on Encryption window

2. Click **Rekey**, as shown in Figure 12-83.

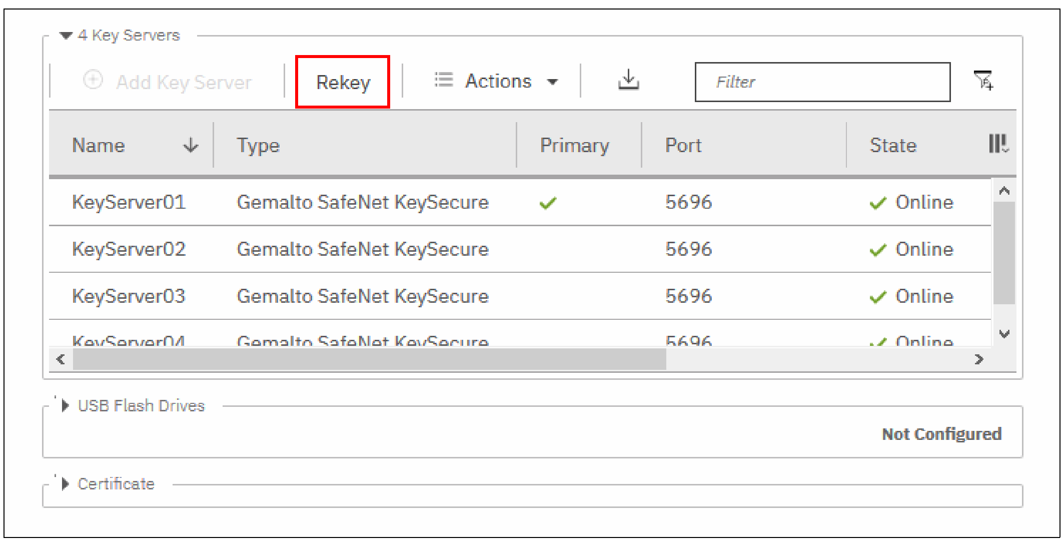


Figure 12-83 Start rekey on SKLM key server

3. Click **Yes** in the next window to confirm the rekey operation, as shown in Figure 12-84.

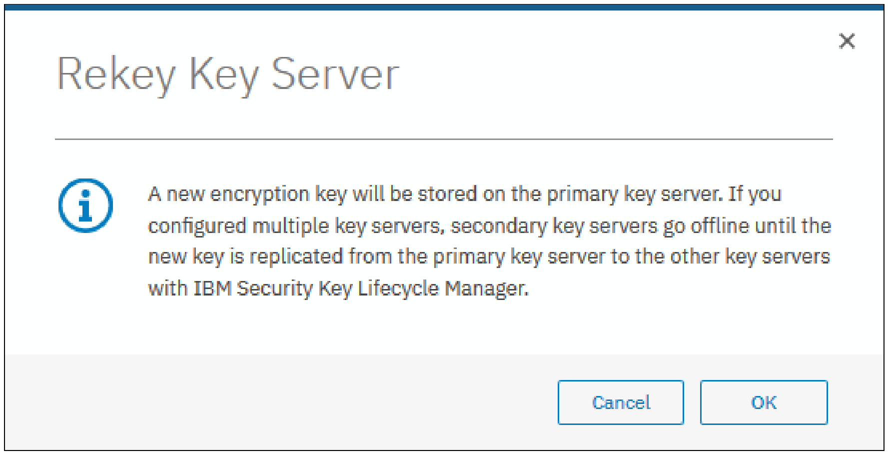


Figure 12-84 Confirm key server rekey operation

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, as shown in Figure 12-85.

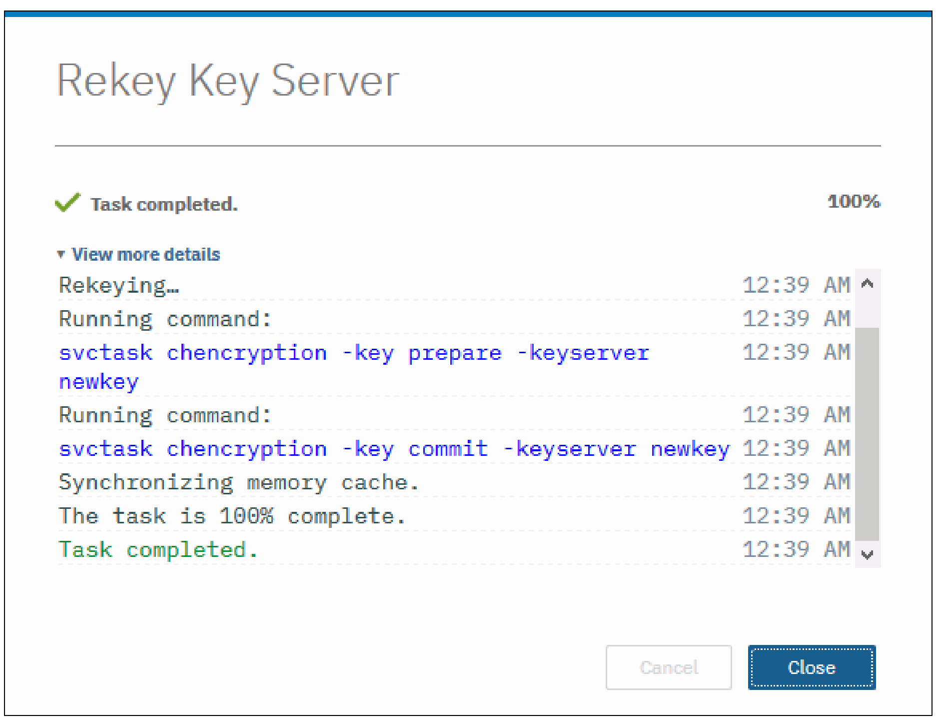


Figure 12-85 Successful key server rekey operation

12.9.2 Rekeying 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 stored on other media. Take the same precautions to securely store all copies of the new encryption key as when you were enabling encryption for the first time.

To rekey the master access key on USB flash drives, complete the following steps:

1. Click **Settings** → **Security** → **Encryption**. Click **USB Flash Drives** to expand the section, as shown in Figure 12-86.

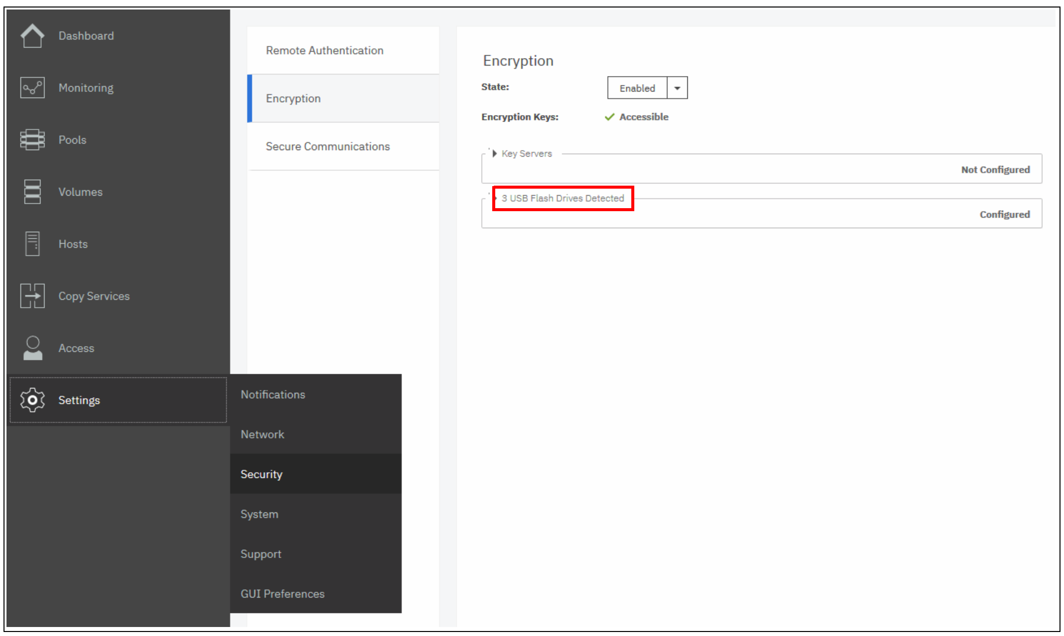


Figure 12-86 Locate USB Flash Drive section in the Encryption view

2. Verify that all USB drives plugged into the system are detected and show as `Validated`, as shown in Figure 12-87. Click **Rekey**. You need at least three USB flash drives, with at least one reported as `Validated` to process with rekey.

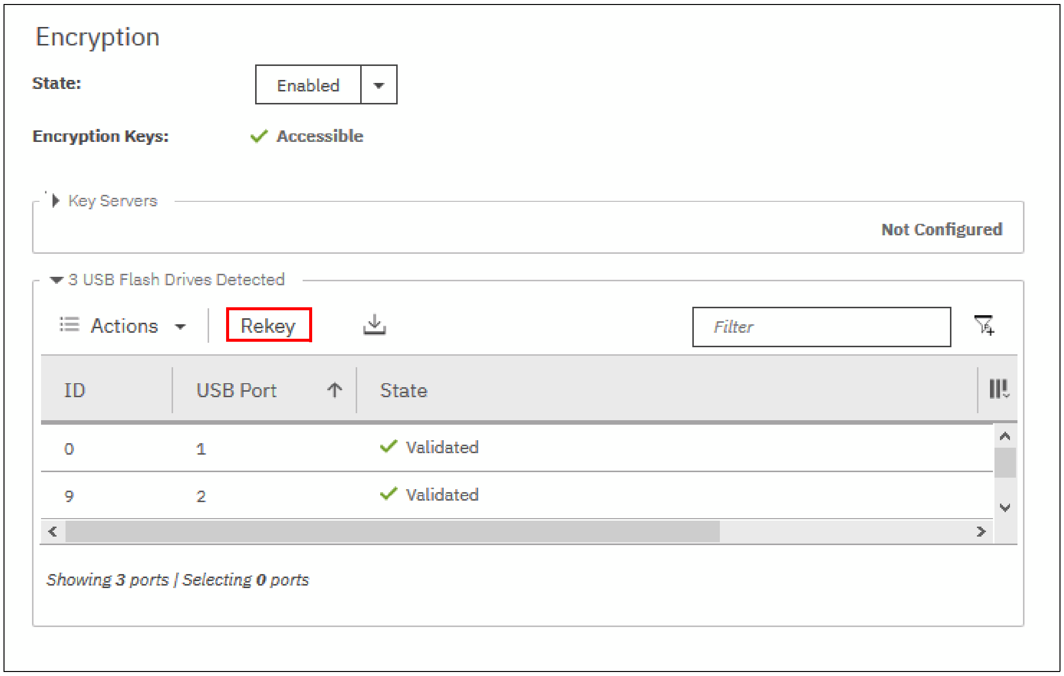


Figure 12-87 Start rekey on USB flash drives provider

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 12-88 on page 670. Click **Commit** to finalize the rekey operation.

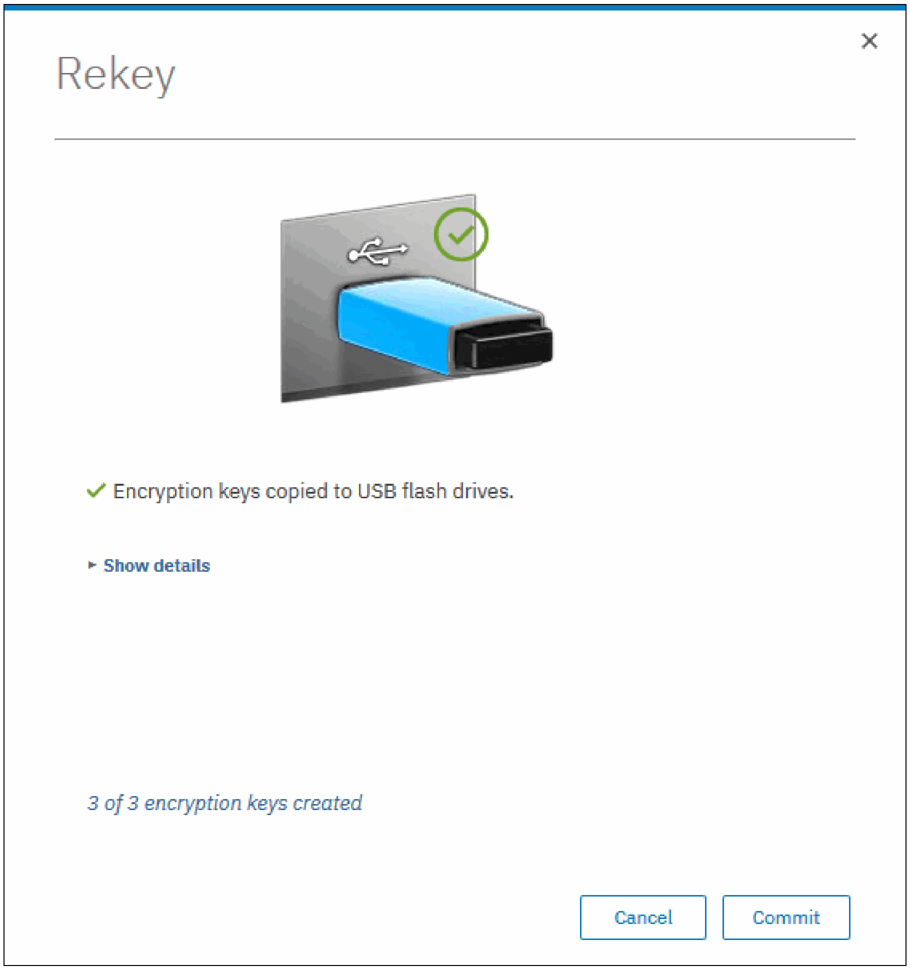


Figure 12-88 Writing new keys to USB flash drives

4. You receive a message confirming the rekey operation was successful, as shown in Figure 12-89. Click **Close**.

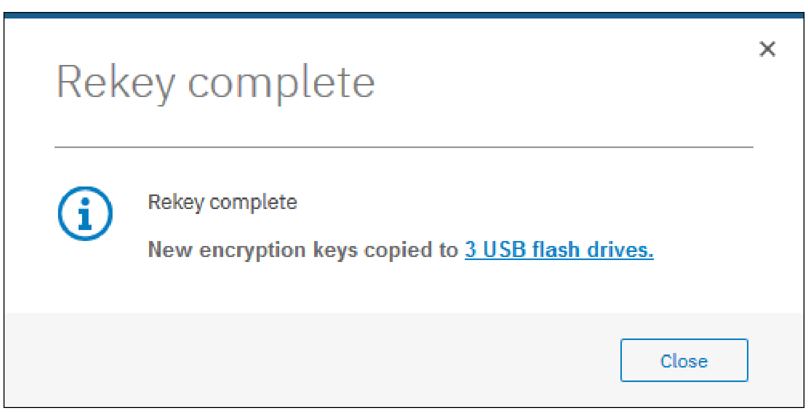


Figure 12-89 Successful rekey operation using USB flash drives

12.10 Disabling encryption

You are prevented from disabling encryption if any encrypted objects are defined apart from self-encrypting MDisks. You can disable encryption in the same way whether you use USB flash drives, key server, or both providers.

To disable encryption, complete the following steps:

1. Click **Settings** → **Security** → **Encryption** and click **Enabled**. If no encrypted objects exist, a menu is displayed. Click **Disabled** to disable encryption on the system. Figure 12-90 shows an example for a system with both encryption key providers configured.

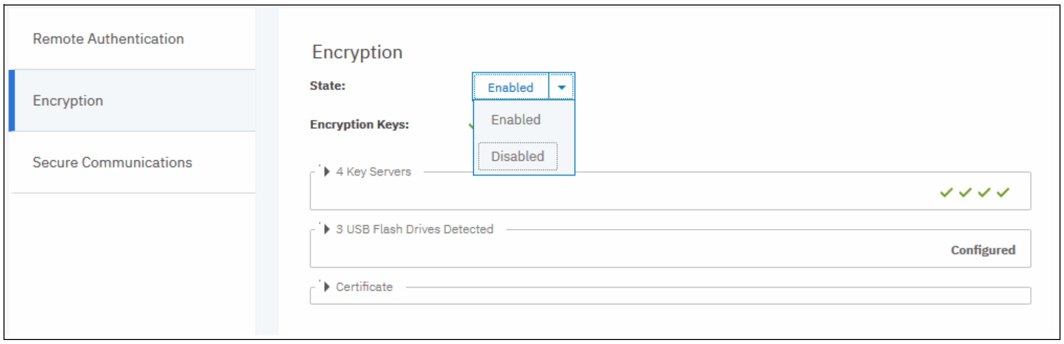


Figure 12-90 Disabling encryption on a system with both providers

2. You receive a message confirming that encryption was disabled. Figure 12-91 shows the message when a key server is used.

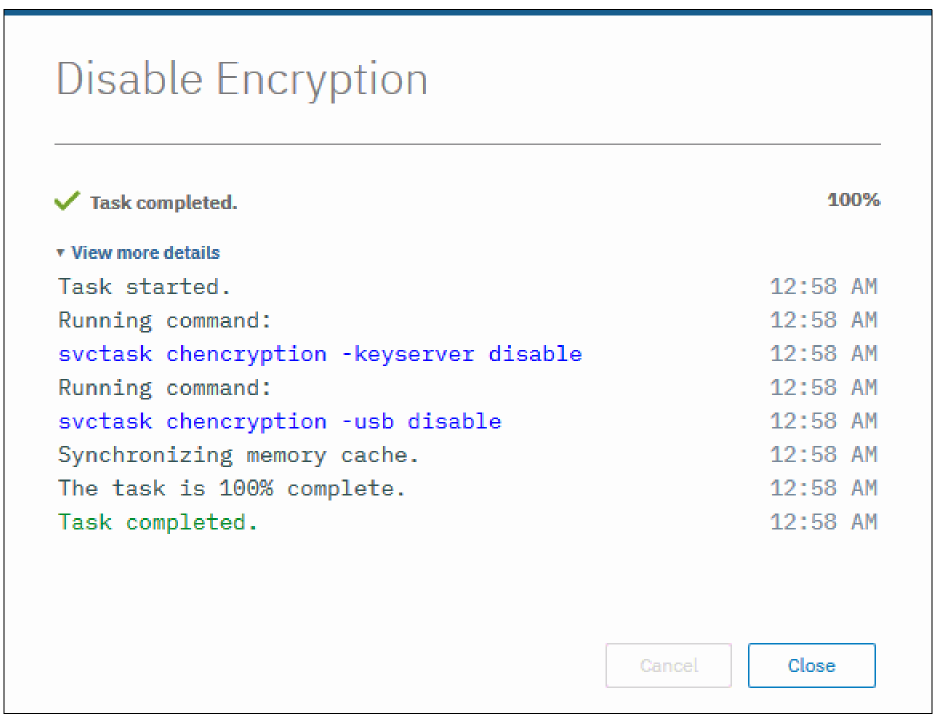


Figure 12-91 Encryption disabled

13

RAS, monitoring, and troubleshooting

This chapter introduces useful and common procedures to maintain the IBM Spectrum Virtualize. It includes the following topics:

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 Storwize V7000 system reliable.

Fault tolerance and high levels of availability are achieved by using the following methods:

- The Redundant Array of Independent Disks (RAID) capabilities of the underlying disks
- IBM Storwize V7000 nodes clustering using a *Compass* architecture
- Auto-restart of hung nodes
- Integrated Battery Backup Units (BBU) to provide memory protection if a site power failure occurs
- Host system failover capabilities using N-Port ID Virtualization (NPIV)

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 (DMP) with guided online replacement process
- Concurrent log analysis and memory dump data recovery tools
- Concurrent maintenance of IBM Storwize V7000 components
- Concurrent upgrade of IBM Storwize V7000 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 LED indicators
- Error and event notification through Simple Network Management Protocol (SNMP), syslog, and email
- Optional Remote Support Assistant

The heart of IBM Storwize V7000 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 IBM Storwize V7000 system, monitoring, and troubleshooting.

Throughout this chapter, the term *IBM Storwize V7000* refers to both models of the product: IBM Storwize V7000 Gen2 and Gen2+.

13.1.1 Node canisters

Two node canisters are contained in the control enclosure that work as a clustered system running the IBM Spectrum Virtualize software. Figure 13-1 shows the ports and indicator lights of a node canister, with the Front Panel LEDs providing a convenient status summary. The second canister is placed next to the first one in a side-by-side configuration.

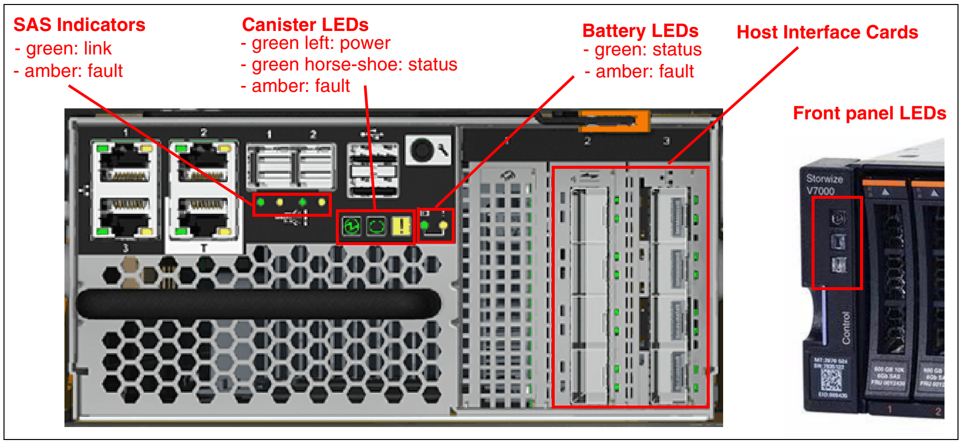


Figure 13-1 Ports and indicators of node canister (524 controller model)

Host interface cards

Two host interface card (HIC) slots are vertically placed on the right side of the canister complemented by an additional slot for a Compression Accelerator card or a compression pass-thru card if the additional compression engine was not purchased. HIC slots can accommodate up to 2 dual-port 16 gigabit per second (Gbps) Fibre Channel (FC) cards, or 1 four-port 10-gigabit Ethernet (GbE) adapter stand-alone or in combination with an FC card.

The 16 Gbps FC adapter is based on the Emulex Lancer multiprotocol chip, and supports FC or Fibre Channel over Ethernet (FCoE) traffic using a single chip (Converged Network Adapter). It can be configured as dual-port 16 Gbps FC adapter or as a four-port 10 GbE with FCoE support. For FC configurations, the meaning of port LEDs is explained in Table 13-1.

Table 13-1 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 Universal Serial Bus (USB) connectors are available in the horizontal position. 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 10/100/1000 megabits per second (Mbps) Ethernet ports are side by side on the canister. They are marked as 1 and 3 on the left and 2 and T on the right. Each port has two LEDs, and their status values are listed in Table 13-2. Although ports 1 - 3 are available for management and internet Small Computer System Interface (iSCSI) purposes, the T port is strictly dedicated to the technician actions (initial and emergency configuration by local support personnel).

Table 13-2 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

Two 12 Gbps serial-attached SCSI (SAS) ports are side by side on the canister with indicator LEDs below them. They are numbered 1 on the left and 2 on the right. Each port is associated with one green and one amber LED indicating its status of the operation, as listed in Table 13-3.

Table 13-3 SAS LED statuses

LED	Meaning
Green	Link is connected and up.
Orange	Fault on the SAS link (disconnected, wrong speed, errors).

Node canister status LEDs

There are three LEDs in a row in the upper middle position of the canister that indicates the status and the functionality of the node (see Table 13-4).

Table 13-4 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, node is in service state or in error, preventing the software to start.
			Flashing (2 Hz)	Canister is started and in standby mode.
			Flashing (4 Hz)	Node is running 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 at this time.
			Off	No power to the canister or canister is in standby mode.

Position	Color	Name	State	Meaning
Right	Amber	Fault	On	The canister is in a service state, or in error, preventing the software from starting.
			Flashing (2 Hz)	Canister is being identified.
			Off	Node is either in candidate or active state.

13.1.2 Expansion canisters

As Figure 13-2 shows, two 12 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. Similar to the controller canisters, expansion canisters are also installed in the enclosure side by side in a vertical position.

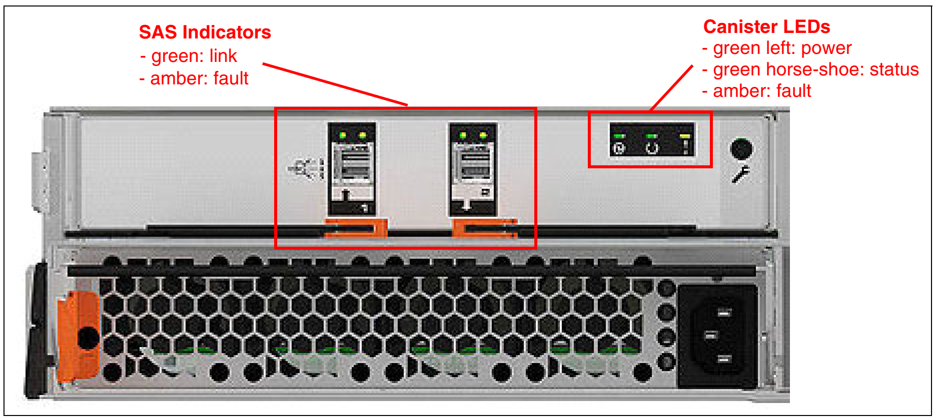


Figure 13-2 Expansion canister status LEDs

The interpretation of the SAS status LED indicators has the same meaning as the LED indicators of SAS ports in the control enclosure (Table 13-3 on page 676).

Table 13-5 lists the LED status values of the expansion canister.

Table 13-5 Expansion canister LEDs statuses

Position	Color	Name	State	Meaning
Left	Green	Power	On	The canister is powered on.
			Off	No power available to the canister.
Middle	Green	Status	On	The canister is operating normally.
			Flashing	There is an error with the vital product date (VPD).
Right	Amber	Fault	On	There is an error 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 13-3 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.

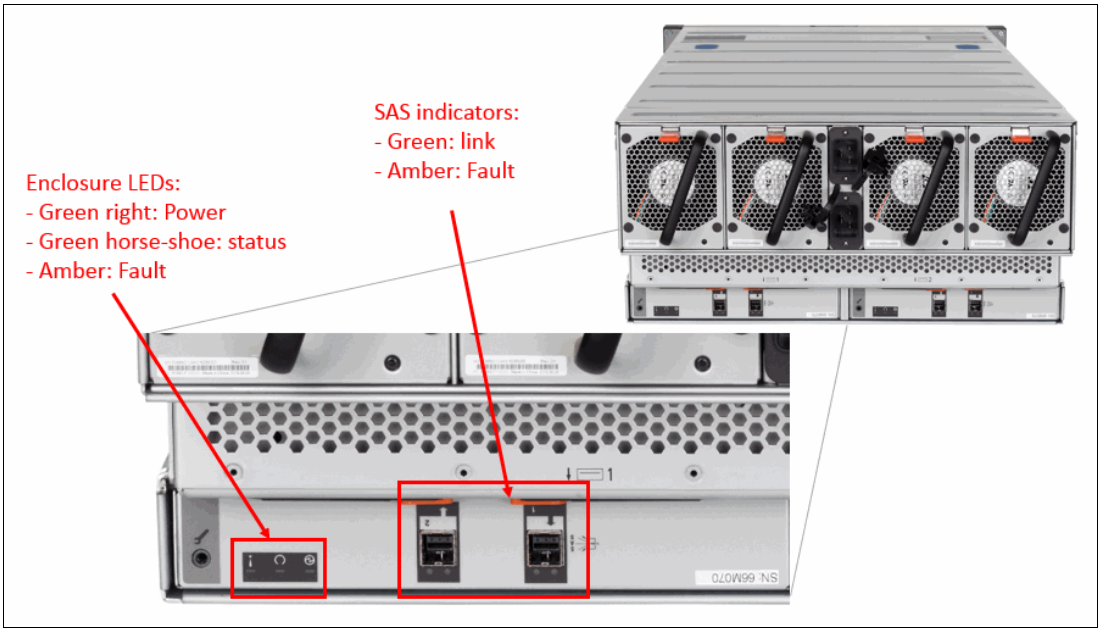


Figure 13-3 Dense Drawer LEDs

The interpretation of SAS status LED indicators has the same meaning as the LED indicators of SAS ports mentioned in the previous section (see Table 13-5 on page 677).

Table 13-6 shows the LED status values of the expansion canister.

Table 13-6 Expansion canister LEDs statuses

Position	Color	Name	State	Meaning
Right	Green	Power	On	The canister is powered on.
			Off	No power 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 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. IBM Storwize V7000 Gen2 control enclosure attaches up to 20 expansion enclosures or up to four Dense Drawer enclosures.

A strand starts with an SAS initiator chip inside an IBM Storwize V7000 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 is no single point of failure exists.

Figure 13-4 shows how the SAS connectivity works inside the node and expansion canisters.

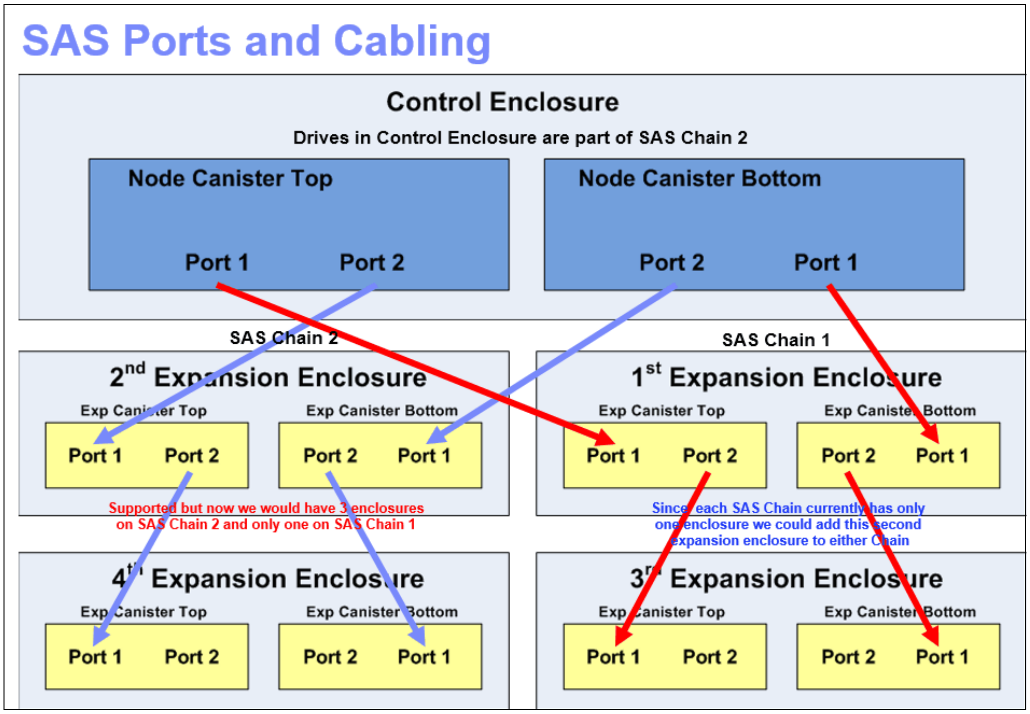


Figure 13-4 Concept of SAS chaining

A chain consists of a set of enclosures that are correctly interconnected (Figure 13-5). Chain 1 of an I/O group is connected to SAS port 1 of both node canisters. Chain 2 is connected to SAS port 2. This configuration means that chain 2 includes the SAS expander and drives of the control enclosure.

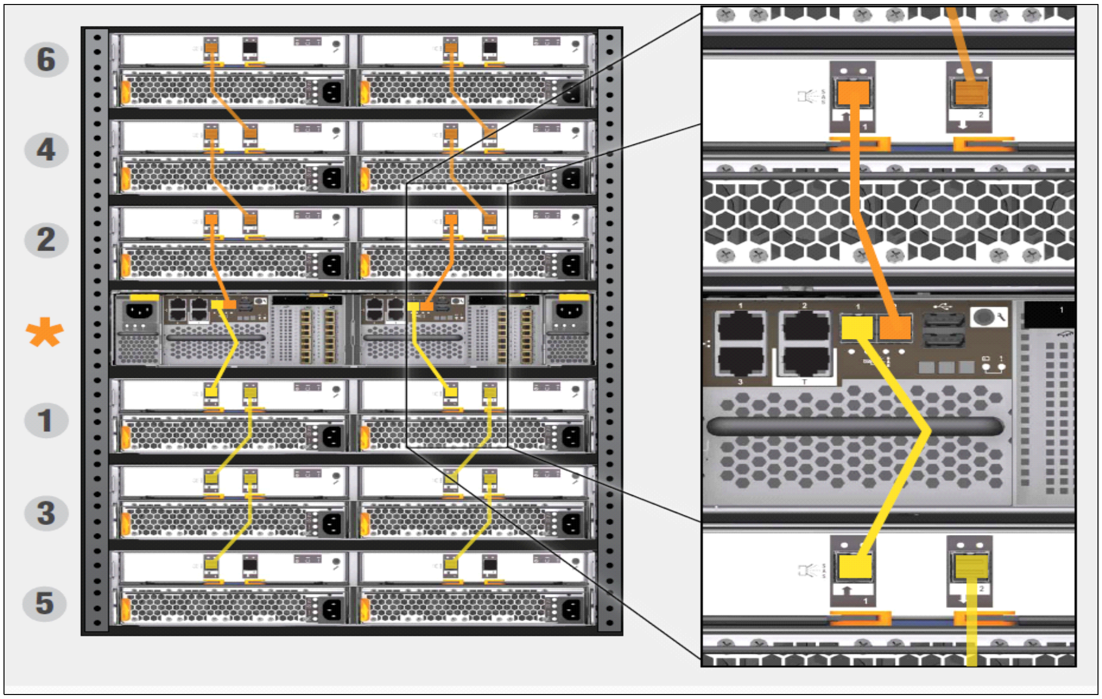


Figure 13-5 SAS cabling with numbering of enclosures

At system initialization, when devices are added to or removed from strands, IBM Storwize V7000 performs a discovery process to update the state of the drive and enclosure objects.

13.1.5 Power

All enclosures accommodate two power supply units (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 integrated in the node canister continues to supply power to the node. A fully charged battery is able to perform two fire hose dumps. It supports the power outage for 5 seconds before initiating safety procedures.

Figure 13-6 shows two PSUs present in the control and expansion enclosure. The controller PSU has two green and one amber indication LEDs that report the status of the PSU.

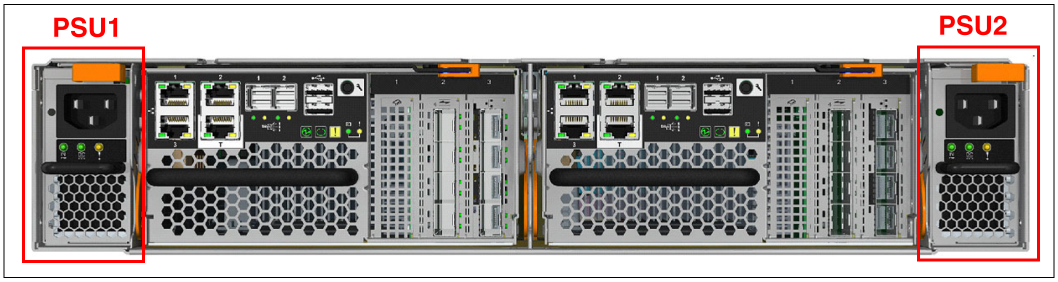


Figure 13-6 Controller and expansion enclosure LED status indicator

Figure 13-7 presents the rear overview of the enclosure canister with a PSU. In contrast to the control enclosure, these PSUs do not have a power switch. The enclosure is powered on by the direct attachment of a power cable.

Figure 13-7 Expansion enclosure power supply unit

Power supplies in both control and expansion enclosures are hot-swappable and replaceable without a need 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 IBM Storwize V7000

You can safely shut down an IBM Storwize V7000 system by using the GUI or CLI.

Important: Never shut down your IBM Storwize V7000 system by powering off the PSUs, removing both PSUs, or removing both power cables from a running system. It can lead to inconsistency or loss of the data that is staged in the cache.

Before shutting down IBM Storwize V7000, stop all hosts that have allocated volumes from the device. This step can be skipped for hosts that have volumes that are also provisioned with mirroring (host-based mirror) from different storage devices. However, doing so incurs errors that are related to lost storage paths/disks on the host error log.

You can shut down only one 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 the entire cluster, you need to power on locally to start the system.

13.2.1 Shutting down and powering on a complete infrastructure

When you shut down or power on the entire infrastructure (storage, servers, 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 IBM Storwize V7000 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 Storwize V7000 systems:
   a. Shut down the Storwize V7000 by using the GUI or CLI. For more information, see 4.7.2, “Shutting down the system” on page 124.
   b. Power off both switches of the controller enclosure.
   c. Power off both switches of all the expansion enclosures.
3. Shut down your SAN switches.

Powering on

To power on your infrastructure, complete the following steps:

1. Power on your SAN switches and wait until the boot 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.
3. When the storage systems are up, power on your servers and start your applications.

13.3 Configuration backup

You can download and save the configuration backup file by using the IBM Storwize V7000 graphical user interface (GUI) or command-line interface (CLI). On an ad hoc basis, manually perform this procedure because it can save the file directly to your workstation. The command-line option requires login to the system and downloading the dumped file by using specific Secure Copy Protocol (SCP). The command-line option is a good practice for an automated backup of the configuration.

Important: Generally, perform a daily backup of the Storwize V7000 configuration backup file, which is 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 also important. It contains configuration data of arrays, pools, volumes, and so on. 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 **svcconfig backup** command. Then, the output of the command can be downloaded by using SCP or GUI.

13.3.1 Backup using CLI

You can use CLI to trigger configuration backup 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 1AM cron backup if you have not made any configuration changes,

Example 13-1 shows how to use the **svcconfig backup** command to generate an ad hoc backup of the current configuration.

Example 13-1 Saving the configuration using CLI

```
IBM_Storwize:ITSO-V7k:superuser>svcconfig backup
..................................................................................
..............................................
CMMVC6155I SVCCONFIG processing completed successfully
IBM_Storwize:ITSO-V7k: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. Use the **lsdumps** command to list them (see Example 13-2).

Example 13-2 Listing the backup files in CLI

```
IBM_Storwize:ITSO-V7k:superuser>lsdumps |grep backup
187 svc.config.backup.bak_7822DFF-1
208 svc.config.backup.xml_7822DFF-1
209 svc.config.backup.sh_7822DFF-1
210 svc.config.backup.log_7822DFF-1
IBM_Storwize:ITSO-V7k:superuser>
```

> **Note:** The `svc.config.backup.bak` file is a previous copy of the configuration, and not part of the current backup.

Table 13-7 lists the three files that are created by the backup process.

Table 13-7 Files created by the backup process

File name	Description
`svc.config.backup.xml`	This file contains your cluster configuration data.
`svc.config.backup.sh`	This file contains the names of the commands that were issued to create the backup of the cluster.
`svc.config.backup.log`	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 using UNIX **scp** or **pscp** for Microsoft Windows, as shown in Example 13-3. Replace the IP address with the cluster IP address of your Storwize V7000 and specify a local folder on your workstation. In this example, we are saving to `C:\V7000Backup`.

Example 13-3 Saving config backup files to your workstation

```
C:\putty>pscp -unsafe
superuser@9.174.157.250:/dumps/svc.config.backup.* c:\V7000backup
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 V7000backup
 Volume in drive C has no label.
 Volume Serial Number is 0608-239A

 Directory of C:\V7000backup

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)  79.028.662.272 bytes free

C:\>
```

The use of the **-unsafe** option enables you to use the wildcard for downloading all the `svc.config.backup` files in a single command.

> **Tip:** If you encounter `Fatal: Received unexpected end-of-file from server` when using the **pscp** command, consider upgrading your version of PuTTY.

13.3.2 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. Navigate to **Settings** → **Support** → **Support Package**.
2. Click the **Manual Upload Instructions** twistie to expand it.
3. Click **Download Support Package**, as shown in Figure 13-8 on page 685.

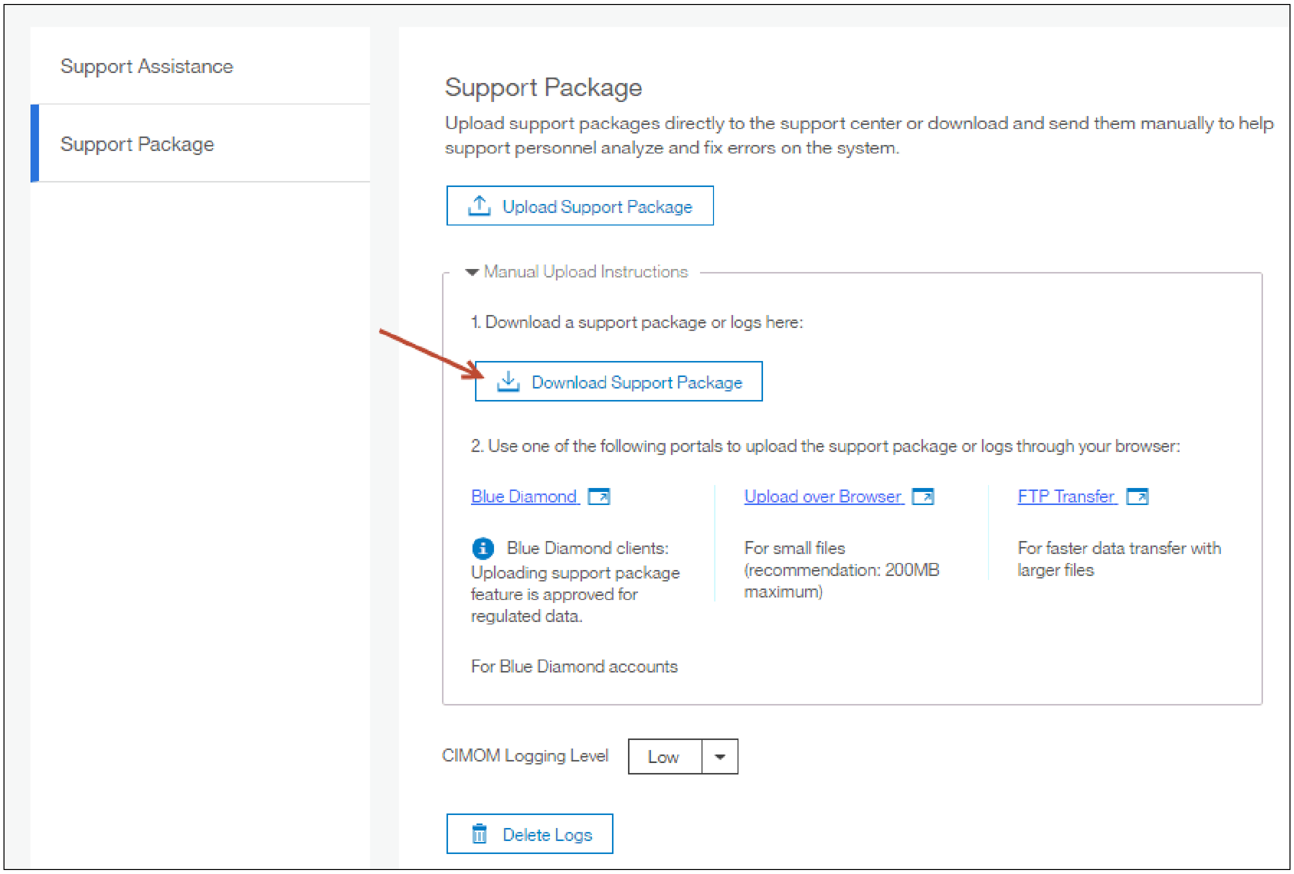


Figure 13-8 Download Support Package

The Download New Support Package or Log File window opens, as shown in Figure 13-9 on page 686.

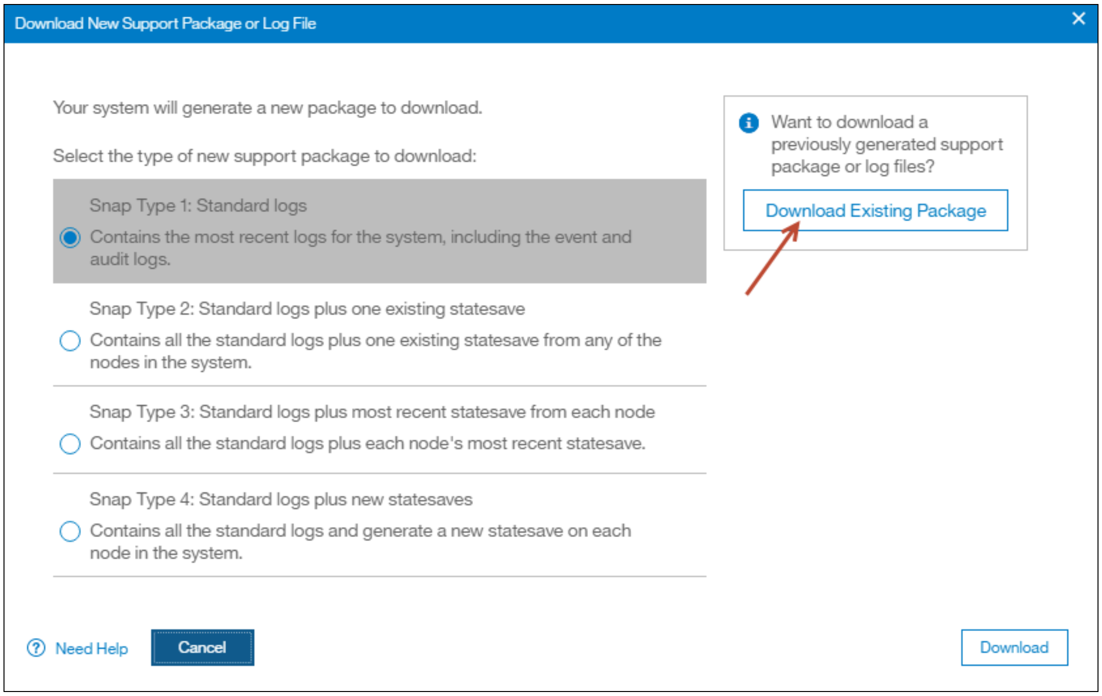


Figure 13-9   Download Existing Package

4. Click **Download Existing Package** to show a list of files found on the config node. We filtered the view by clicking in the **Filter** box, entering `backup`, and pressing **Enter**, as shown in Figure 13-10.

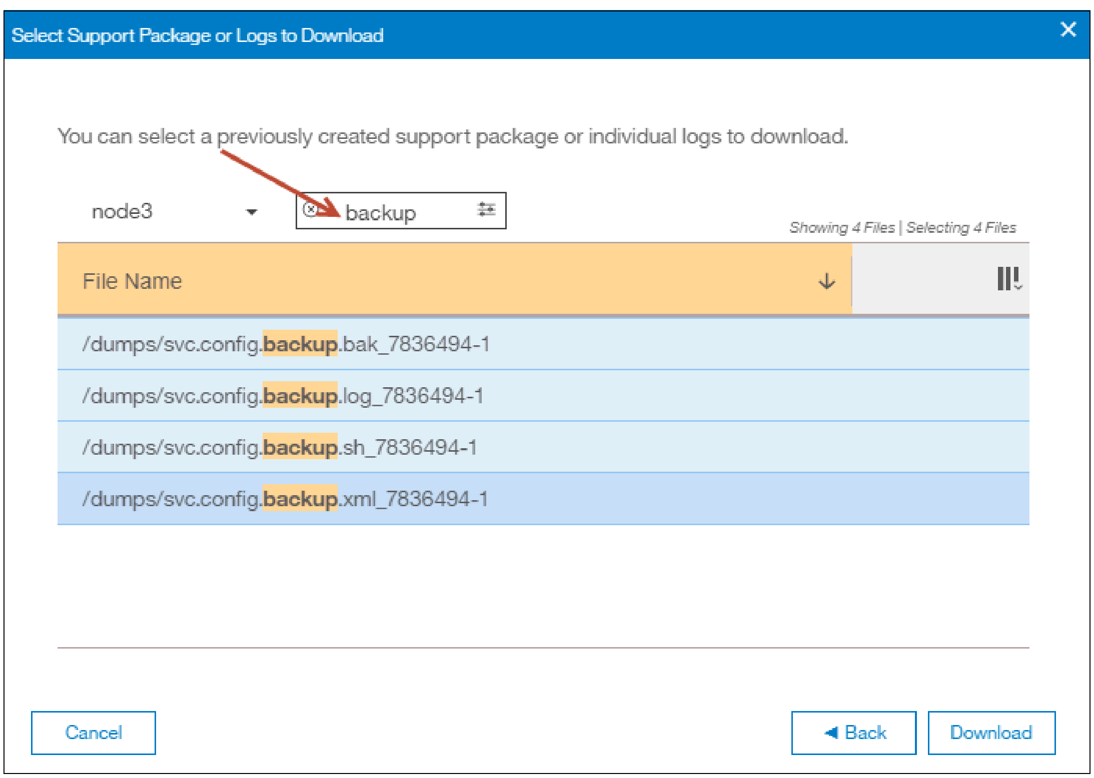


Figure 13-10   Filtering specific files for download

5. Select all of the files to include in the compressed file; then, click **Download**. Depending on your browser preferences, you might be prompted where to save the file or it downloads to your defined download directory.

13.4 Software update

This section describes the operations to update your Storwize V7000 software to V8.1.

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_2145_INSTALL_8.2.1.0
```

13.4.1 Precautions before the update

This section describes the precautions that you should take before you attempt an update.

> **Important:** Before you attempt any IBM Storwize V7000 code update, read and understand the Storwize V7000 concurrent compatibility and code cross-reference matrix. For more information, see this website and click **Latest Storwize V7000 code**.

During the update, each node in your Storwize V7000 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 storage area networks (SANs) work.

If you do not perform this check, certain hosts might lose connectivity to their volumes and experience I/O errors when the Storwize V7000 node that provides that access is shut down during the update process. You can check the I/O paths by using **`datapath query`** SDD commands.

13.4.2 IBM Storwize V7000 update test utility

The software update test utility is a Storwize V7000 software utility that checks for known issues that can cause problems during a Storwize V7000 software update. For more information about the utility, see this website.

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 IBM Storwize V7000 system to perform a readiness check-in preparation for a software update. Run this utility for a final time immediately before you apply the software update to ensure that there were no new releases of the utility since it was originally downloaded.

The installation and use of this utility is nondisruptive, and does not require restart of any IBM Storwize V7000 nodes. Therefore, there is no interruption to host I/O. The utility is only installed on 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 the most current information, see this website.

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).

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.

13.4.3 Updating your Storwize V7000 to V8.2.1

To update the IBM Spectrum Virtualize software to V8.2.1, complete the following steps:

1. Open a supported web browser and navigate to your cluster IP address. A login window opens (see Figure 13-11).

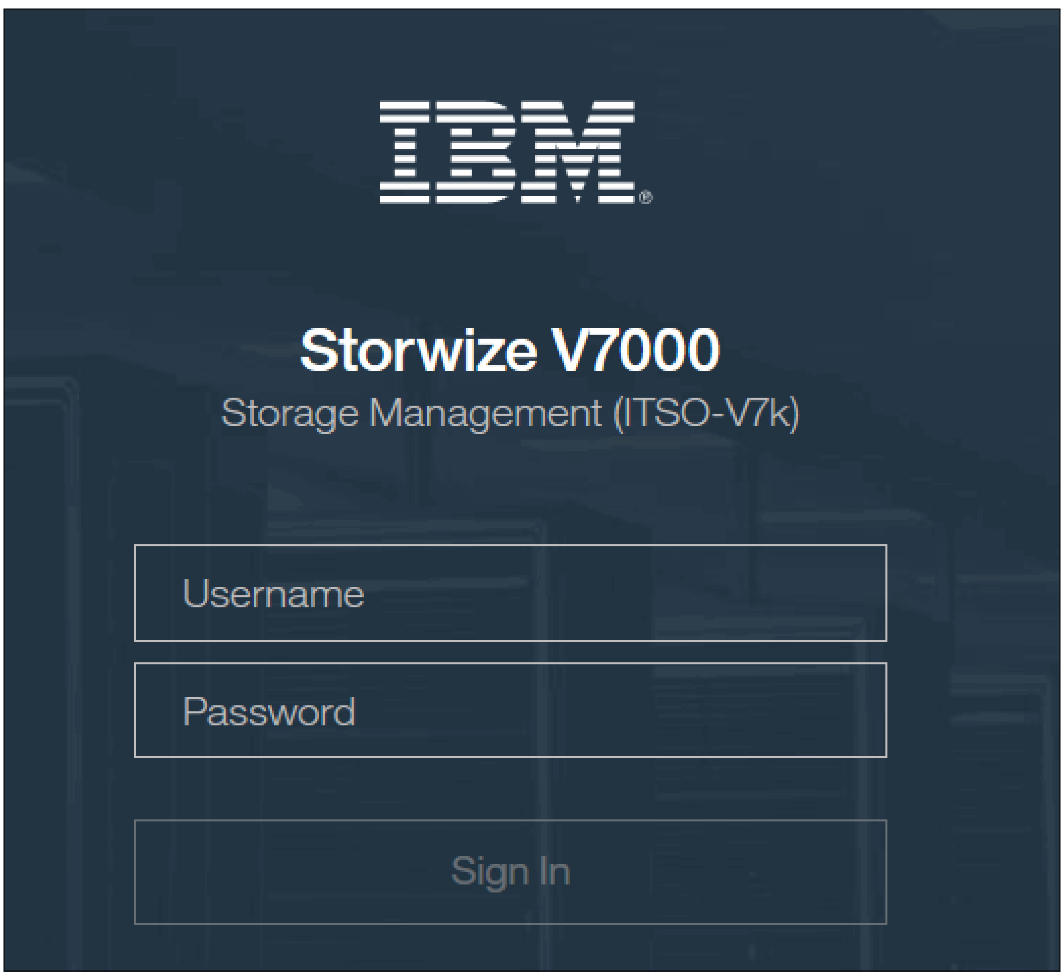


Figure 13-11 IBM Storwize V7000 pre-8.1 GUI login window

2. Log in by using the superuser credentials. The IBM Storwize V7000 management home window opens. Move the cursor over **Settings** and click **System** (see Figure 13-12).

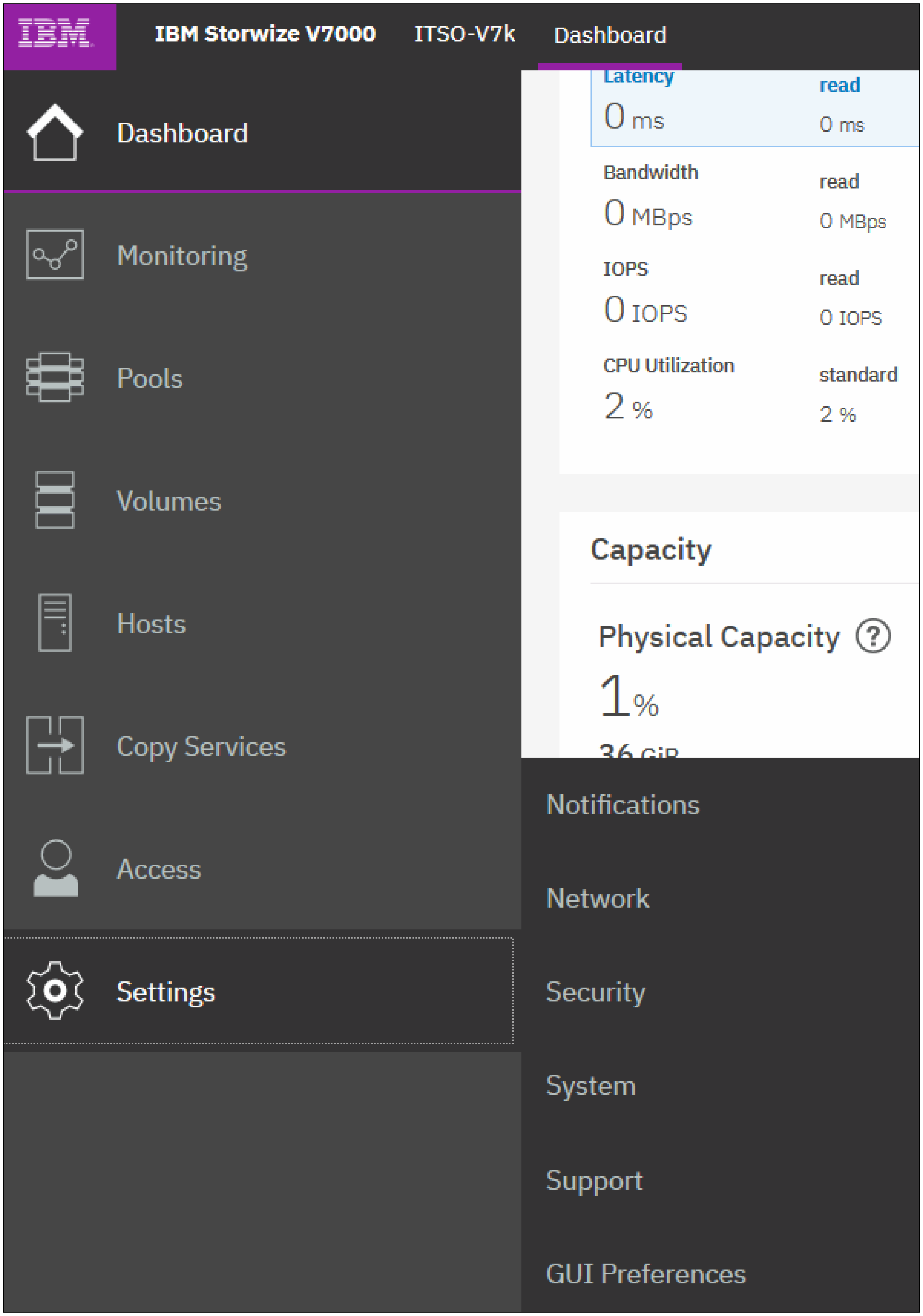


Figure 13-12 Settings menu

3. In the **System** menu, click **Update System**. The Update System window opens (see Figure 13-13).

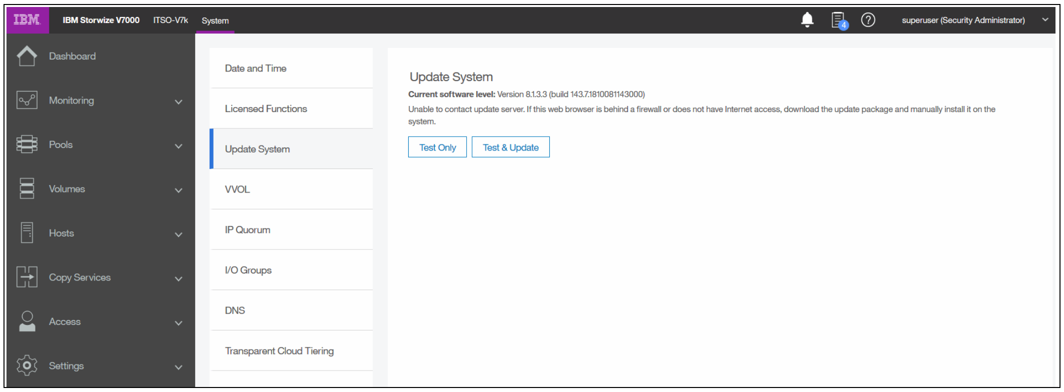


Figure 13-13 Update System window

4. 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 this website (an IBM account is required) to add your IBM Storwize V7000 system to the notifications list to be advised of support information, and to download the current code to your workstation for later upload.

5. Because you downloaded both files from this website, you can click each folder, browse to the location where you saved the files, and upload them to the IBM Storwize V7000. If the files are correct, the GUI detects and updates the target code level, as shown in Figure 13-14.

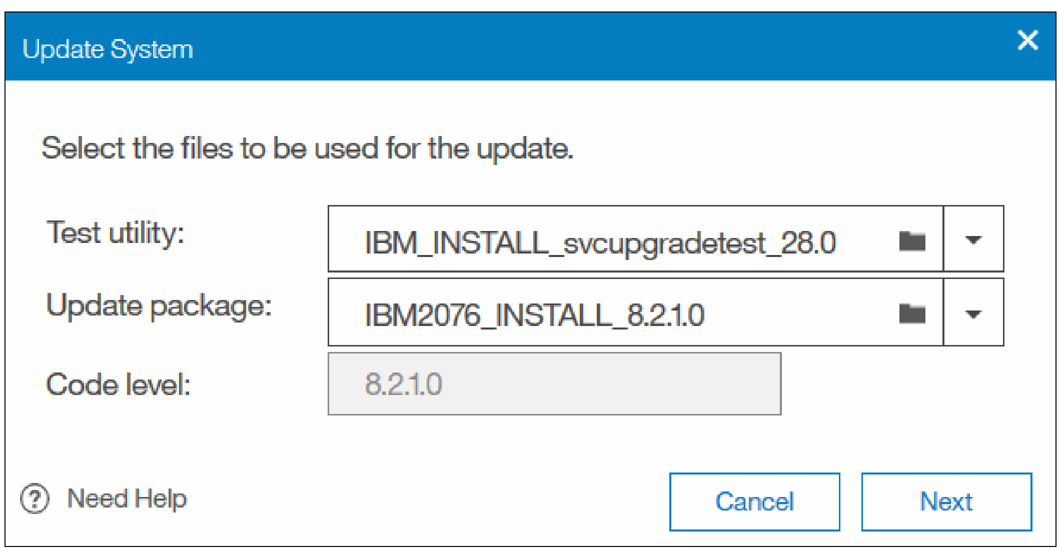


Figure 13-14 Upload option for both Test utility and Update Package

6. Select the type of update you want to perform, as shown in Figure 13-15. 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 **Finish** to begin the update package upload process.

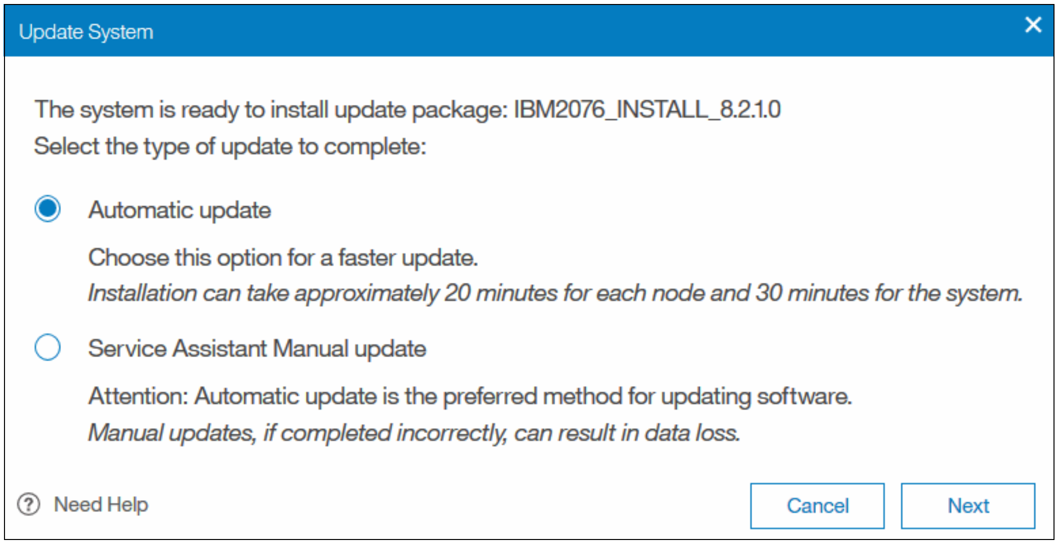


Figure 13-15 The update type selection

When updating from a V8.1 or later level, another window is displayed at this point 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 13-16. The pause option requires that you click **Resume** to continue the update after each pause. Click **Finish**.

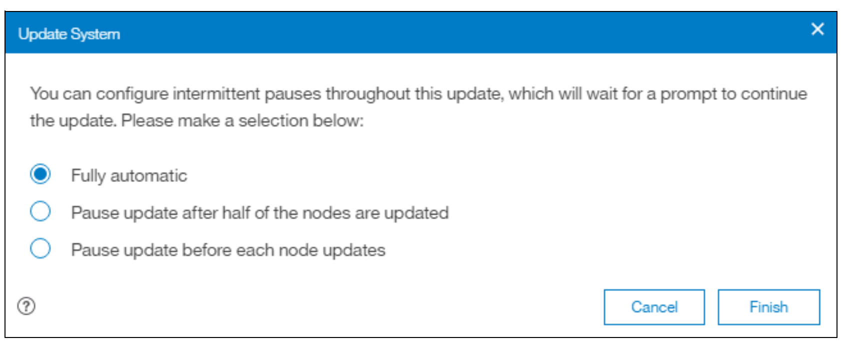


Figure 13-16 New V8.1 update pause options

7. 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 13-17 on page 692).

Figure 13-17 Issues detected by the update test utility

The results pane opens and shows you what issues were detected (see Figure 13-18). 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 assist with resolving more serious issues before continuing.

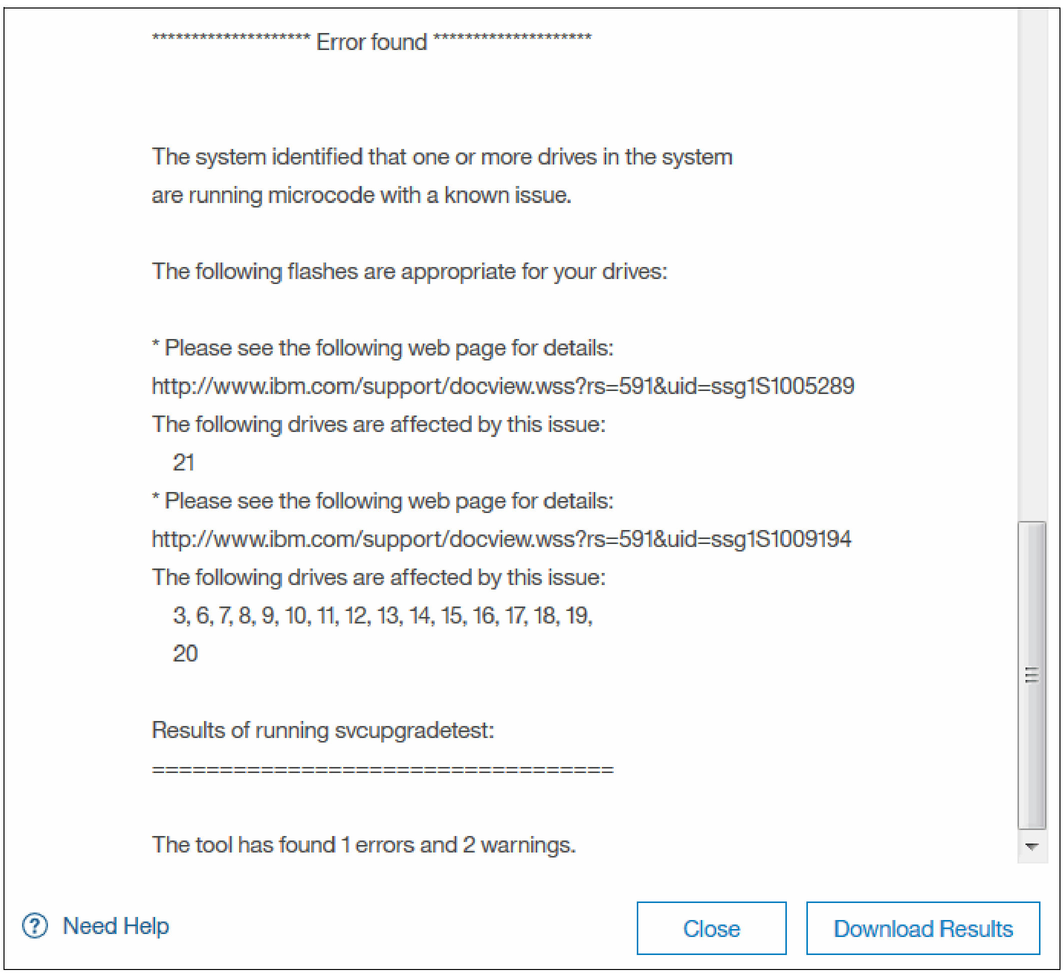


Figure 13-18 Description of the warning from the test utility

8. Click **Resume** in the Update System window and the update proceeds, as shown in Figure 13-19.

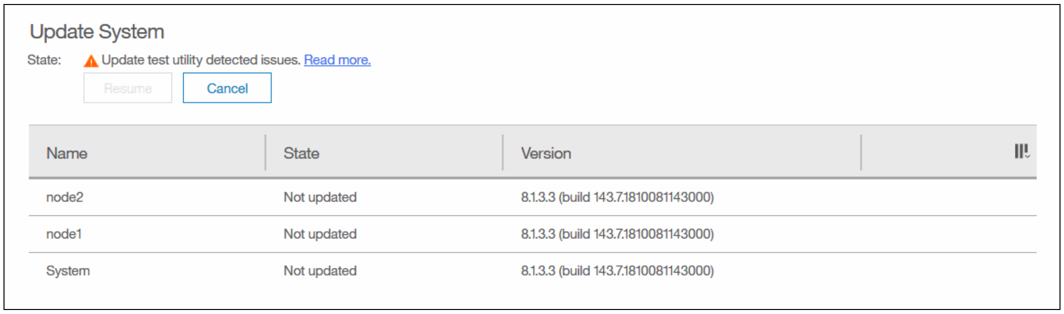


Figure 13-19 Resuming the update

9. Because of the utility detecting issues, another warning comes up to ensure that you investigated them and are certain you want to proceed, as shown in Figure 13-20. When you are ready to proceed, click **Yes**.

Figure 13-20 Warning before you can continue

10. 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 13-21.

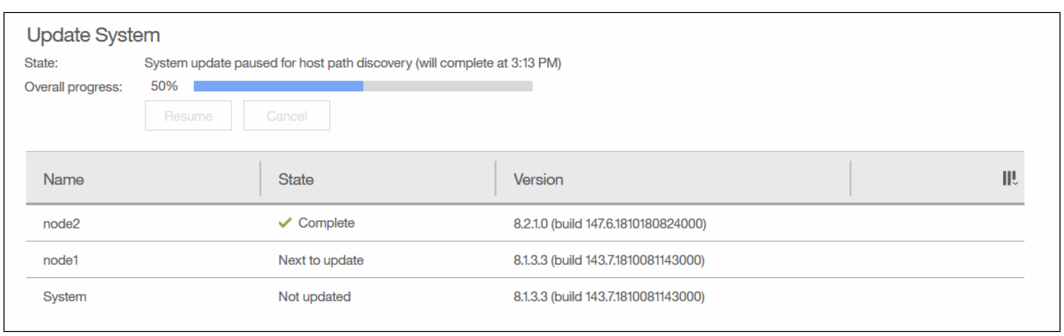


Figure 13-21 Update process paused for host path recovery

11. 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 displays, prompting you to refresh the current session, as shown in Figure 13-22 on page 694.

Tip: If you are updating from V7.8 or later code, the 30-minute wait period can be adjusted by using the **`applysoftware`** CLI command with the **`-delay (mins)`** parameter to begin the update instead of using the GUI.

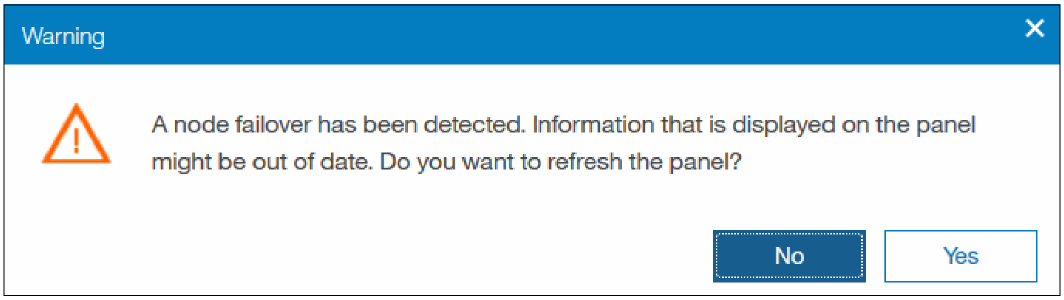


Figure 13-22 Node failover

We now see the new V8.2.1 GUI and the status of the second node updating, as shown in Figure 13-23.

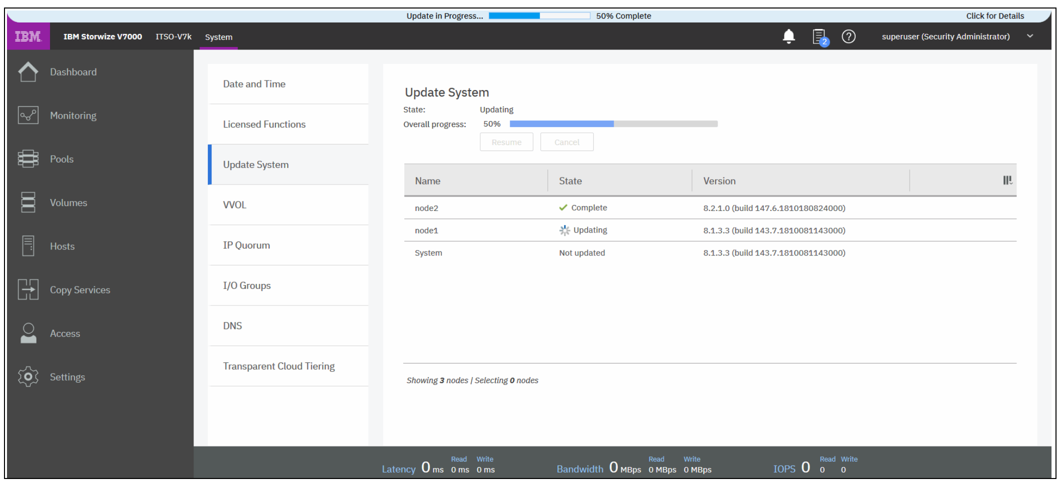


Figure 13-23 New GUI after node failover

After the second node completes, the update is committed to the system, as shown in Figure 13-24.

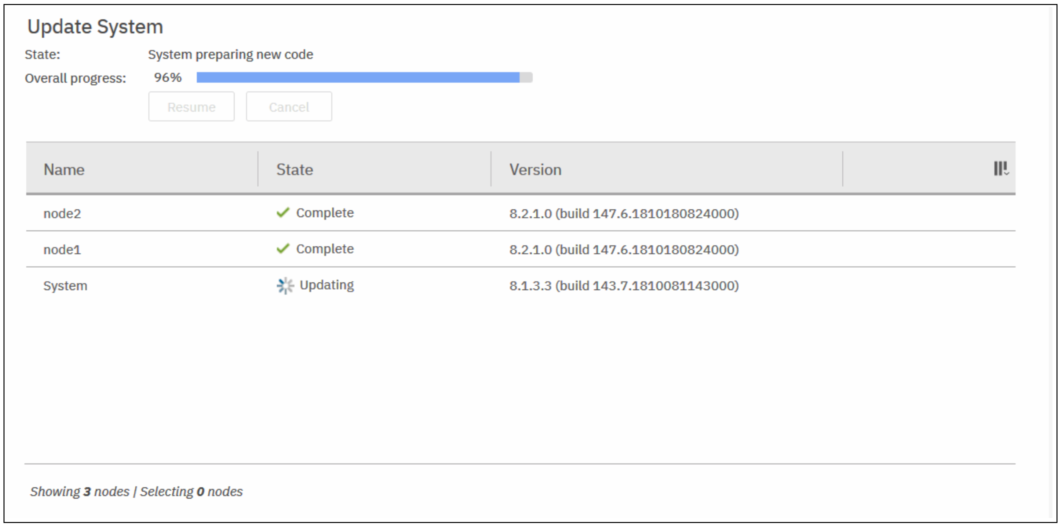


Figure 13-24 Updating system level

The update process completes when all nodes and the system unit are committed. The final status indicates the new level of code installed in the system.

> **Note:** If your nodes have more than 64 GB of memory before updating to V8.1, each node posts an 841 error after the update completes.
>
> Because V8.1 allocates memory differently, the memory must be accepted by running the fix procedure for the event or issue the CLI command **`svctask chnodehw <id>`** for each node. However, only run this command for one node at a time. For more information, see the Storwize V7000 IBM Knowledge Center.

13.4.4 Updating IBM Storwize V7000 drive code

After completing the Storwize V7000 software update as described in 13.4, “Software update” on page 687, the firmware of the Storwize V7000 drives also must be updated. The upgrade test utility identified that downlevel drives are in the system, as shown in Figure 13-25. However, this fact does not stop the system software from being performed.

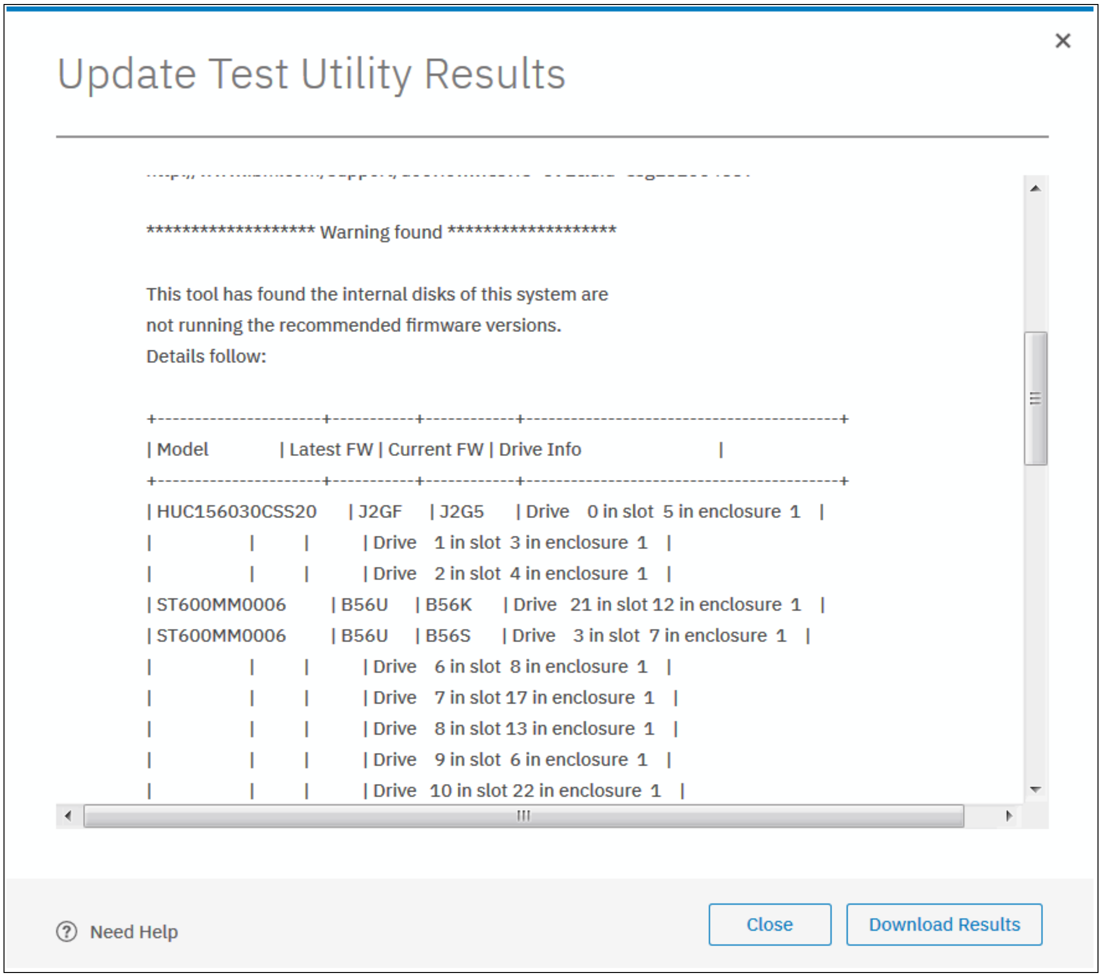


Figure 13-25 Upgrade test utility drive firmware warning

To update the IBM Storwize V7000 drive code, complete the following steps:

1. Download the latest Drive firmware package for IBM Storwize V7000 from Fix Central.
2. On the Storwize V7000 GUI, navigate to **Pools** → **Internal Storage** and select **All Internal.**

3. Click **Actions** and select **Upgrade all**, as shown in Figure 13-26.

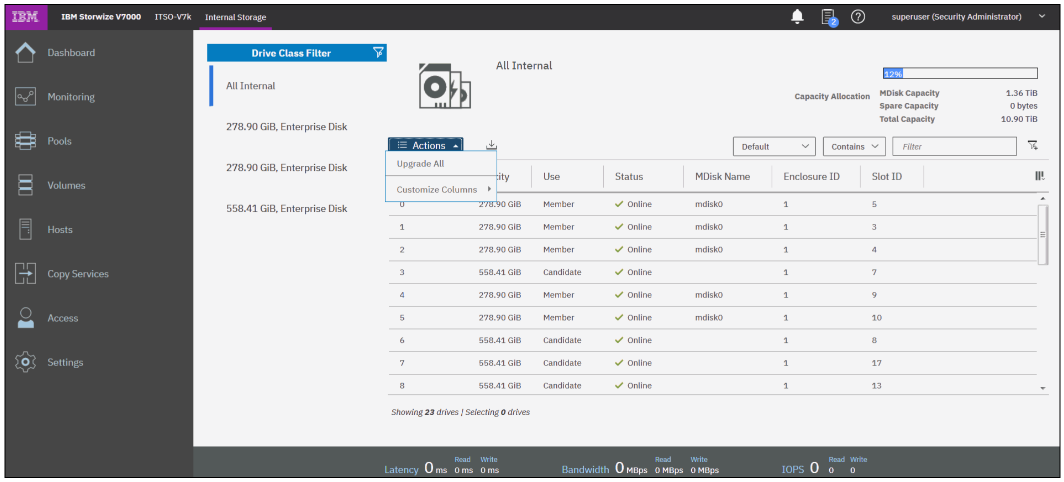


Figure 13-26 Upgrade all internal drives

> **Tip:** The **Upgrade all** drives 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** only upgrade that drives firmware. You can clear an individual drive by pressing **CTRL** and clicking the drive again.

4. The Upgrade all drives panel opens, as shown in Figure 13-27, in which we click the small folder at the right side of the Upgrade package entry box to navigate to where we saved the downloaded file in step 1. Click **Upgrade** to upload the firmware package and begin upgrading any drives that are downlevel. Do *not* select the option to install firmware, even if the drive is running a newer level. Only do this under guidance from IBM Support.

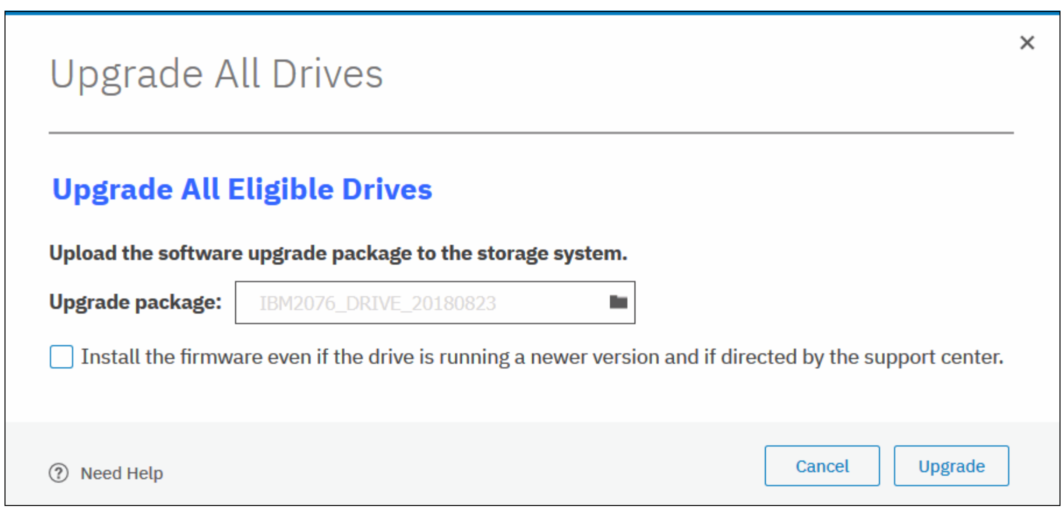


Figure 13-27 Select the Drive Upgrade package

Note: The system upgrades any member drives one at a time. Although the firmware upgrades are concurrent, it does cause a brief reset to the drive. However, the RAID technology allows the Storwize V7000 to ride through this brief interruption. After a drive completes its update, a calculated wait time exists before the next drive update ensures that the previous drive is stable after upgrading and can vary on system load.

5. 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 13-28.

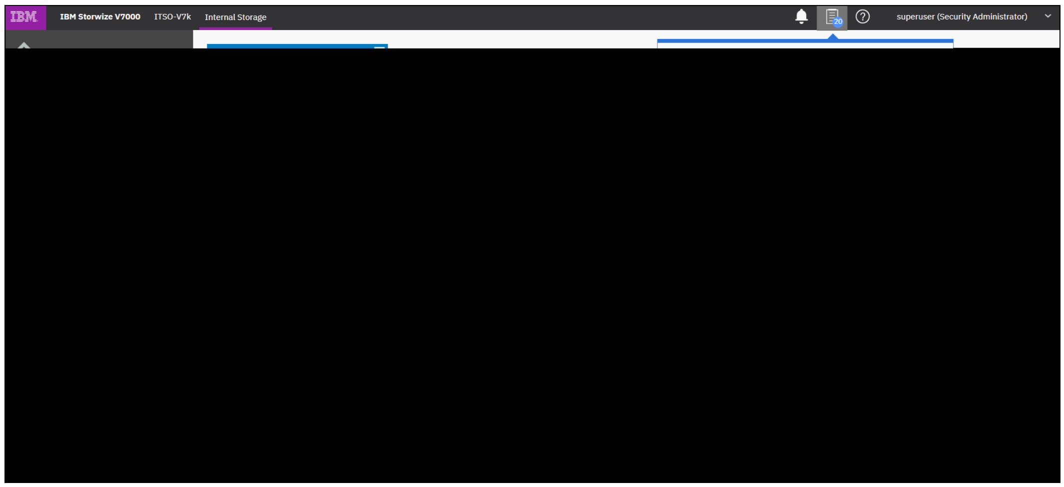


Figure 13-28 Selecting Drive upgrade running Task view

The Drive upgrade running task panel is displayed in which the drives that are pending upgrade and an estimated time of completion are listed, as shown in Figure 13-29.

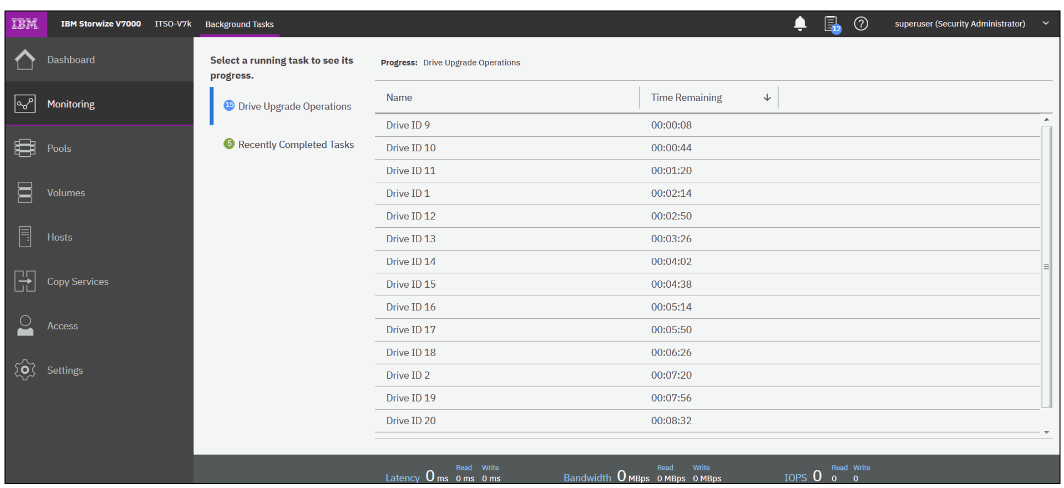


Figure 13-29 Drive upgrade progress

6. We can also view each drives firmware level from the Pools Internal Storage All Internal panel by enabling the drive firmware option after right-clicking in the column header line, as shown in Figure 13-30.

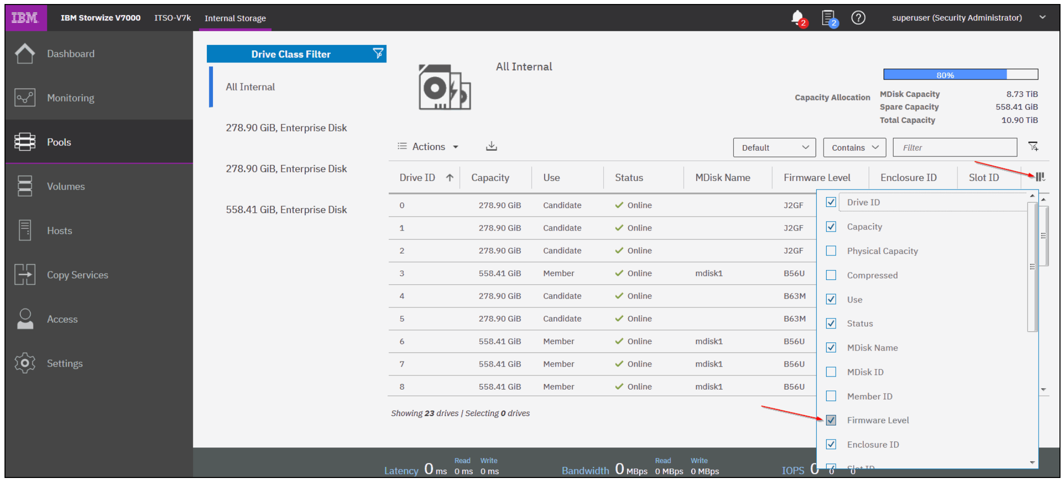


Figure 13-30 Viewing Drive firmware levels

With the Firmware level column enabled, we can see the current level of each drive, as shown in Figure 13-31.

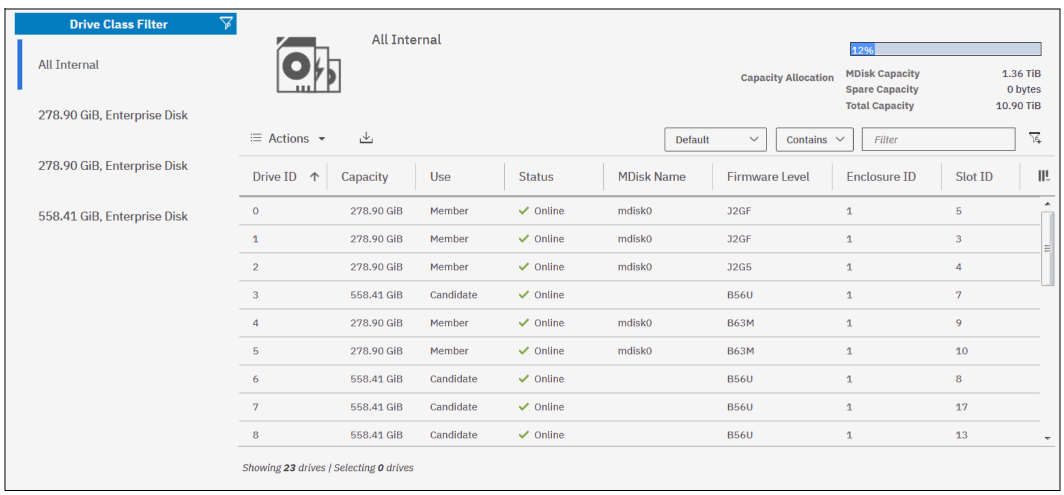


Figure 13-31 Drive firmware display

13.4.5 Manually updating the Storwize V7000

This example assumes that you have an 8-node cluster of the IBM Storwize V7000 cluster, as listed in Table 13-8.

Table 13-8 The iogrp

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 `candidate` status.
3. Select the candidate node and click **Update Manually** from the left pane.
4. Browse and locate the code that you downloaded and saved to your PC.
5. Upload the code and click **Update**.

   When the update is completed, a message caption indicating software update completion displays. The node then reboots, and appears again in the service GUI after approximately 20 - 25 minutes in `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: At this point, the update is 50% completed. You now have one node from each iogrp 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 config node is removed from the cluster, which keeps the cluster manageable.

    The removed configuration node becomes 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.

13.5 Health checker feature

The IBM Spectrum Control health checker feature runs in the IBM Cloud. Based on the weekly call home inventory reporting, it proactively creates recommendations. These recommendations are provided on the IBM Call Home Web, which is found at ibm.com. Click **Support** → **My support** → **Call Home Web** (see Figure 13-32).

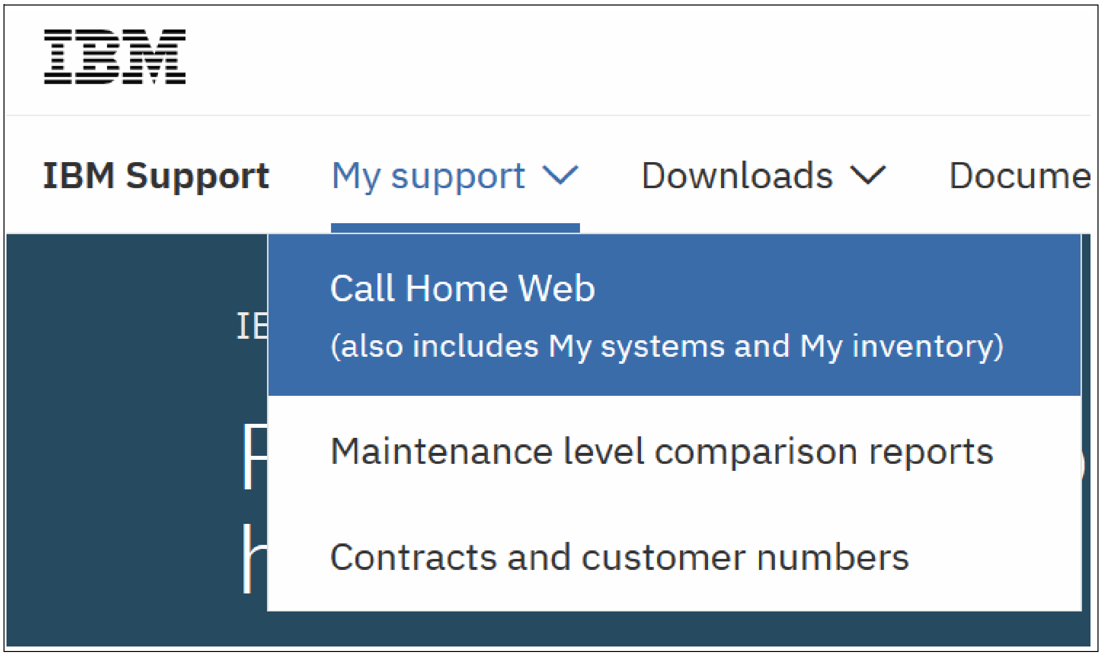


Figure 13-32 Call Home Web on ibm.com

For a video guide on how to set up and use IBM Call Home Web, see this YouTube web page.

Another feature is *Critical Fix Notification* function, which enables IBM to warn Storwize V7000 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 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 IBM Storwize V7000 and IBM SAN Volume Controller systems themselves do not need to be connected to the internet.
- The function cannot be disabled. Each time 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 strongly preferred to enable Call Home and Inventory reporting on your IBM Spectrum Virtualize clusters.

13.6 Troubleshooting and fix procedures

The management GUI of IBM Storwize V7000 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 explain how to effectively use its features to avoid service disruption of your IBM Storwize V7000.

Figure 13-33 shows the menu to start the Monitoring menu for System information, viewing Events, or seeing real-time Performance statistics.

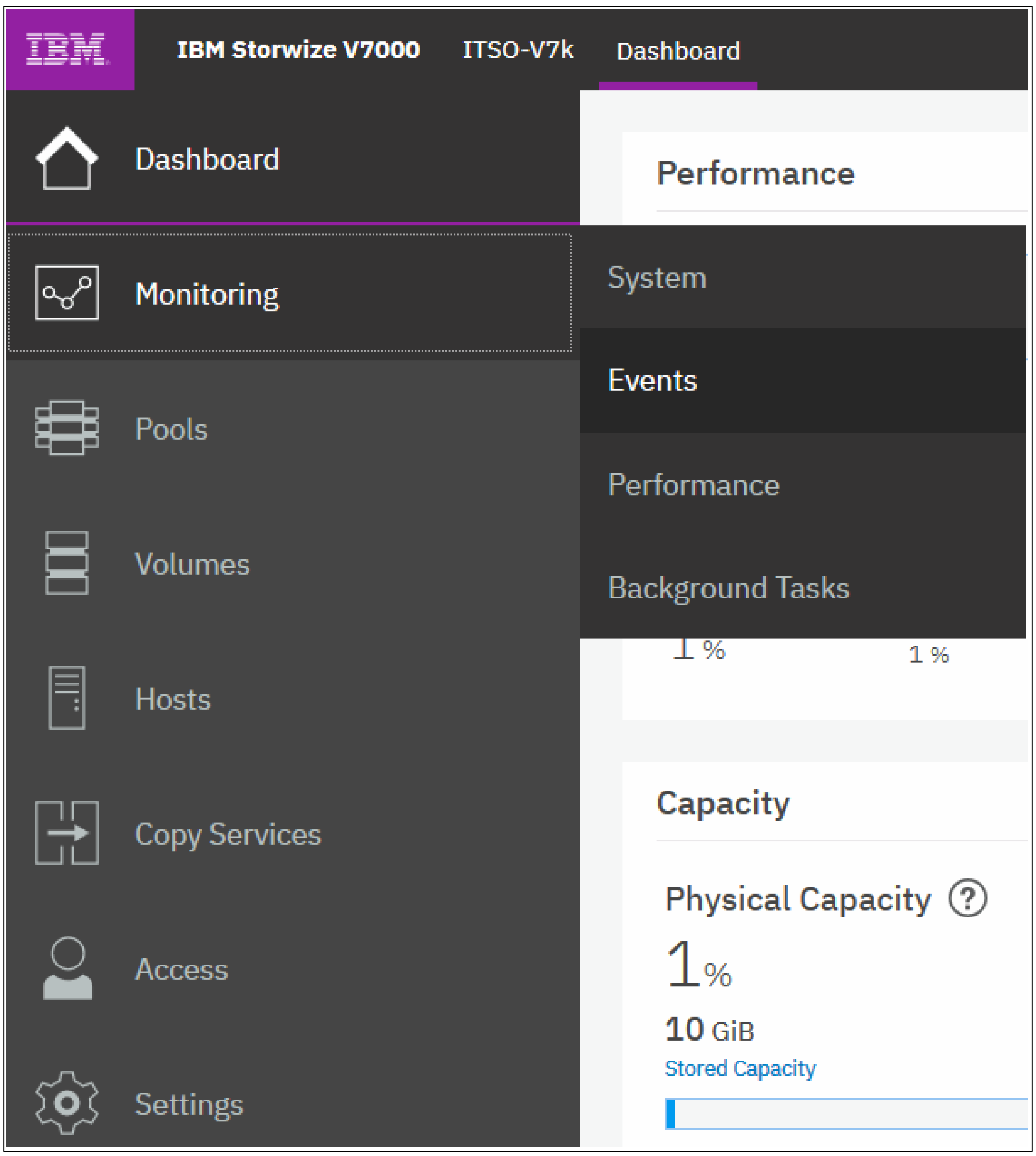


Figure 13-33 Monitoring options

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 in three 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 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 usually 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 on an automatic manner.

13.6.1 Managing event log

Regularly check the status of the system 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. Click **Monitoring** → **Events** to see all problems that exist on the system (see Figure 13-34 on page 704).

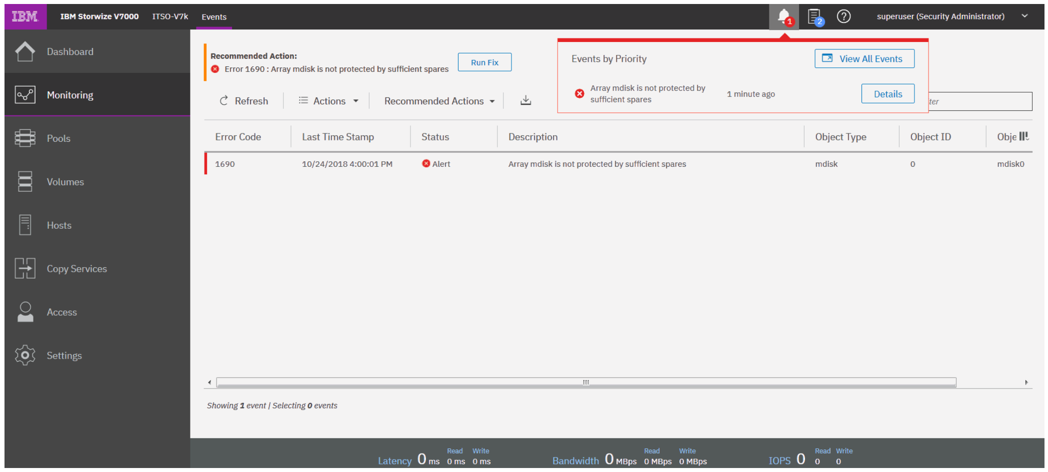


Figure 13-34 Messages in the event log

2. Select **Show All** → **Recommended Actions** to display the most important events to be resolved (see Figure 13-35). The Recommended Actions tab shows the highest priority maintenance procedure that must be run. Use the troubleshooting wizard so that IBM Storwize V7000 system can determine the proper order of maintenance procedures.

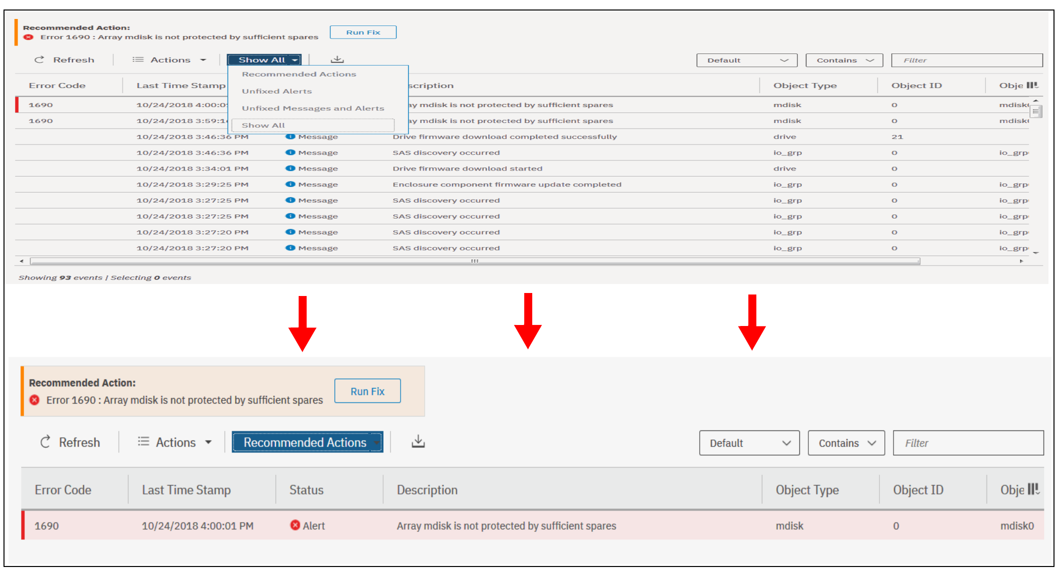


Figure 13-35 Recommended Actions

In this example, the Array MDisk is not protected by sufficient spares and is listed (service error code 1690). Click **Run Fix**. 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 13-36 on page 705).

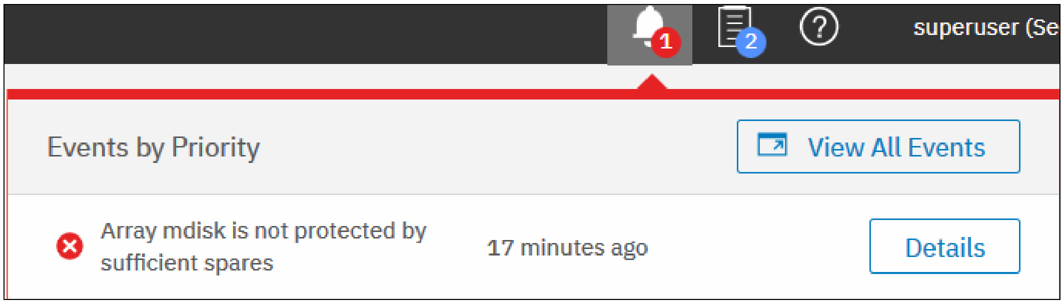


Figure 13-36 Status alerts

13.6.2 Running a fix procedure

If an error code exists for the alert, run the fix procedure to assist you in resolving 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. Always use the fix procedures for both software configuration problems and hardware failures. The fix procedures analyze the system to ensure that the required changes will 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 provides various options to correct the problem. Where possible, the fix procedure runs the commands that are required to reconfigure the system.

Note: After V7.4, you are no longer required to run the fix procedure for a failed Storwize V7000 drive. Hot plugging a replacement drive automatically triggers the validation processes.

The fix procedure also checks that any other existing problem do not result in volume access being lost. For example, if a power supply unit in a node enclosure must be replaced, the fix procedure checks and warns you if the integrated battery in the other power supply unit is not sufficiently charged to protect the system.

Hint: Always use the **Run Fix** button, which resolves the most serious issues first. Often, other alerts are corrected automatically because they were the result of a more serious issue.

The following example demonstrates how to clear the error that is related to a SAN Fabric zoning error to a backend disk controller:

1. From the GUI menu on the left, click **Monitoring** → **Events**, and list only the recommended actions by using the **Actions** menu (see Figure 13-37). Click **Run Fix**.

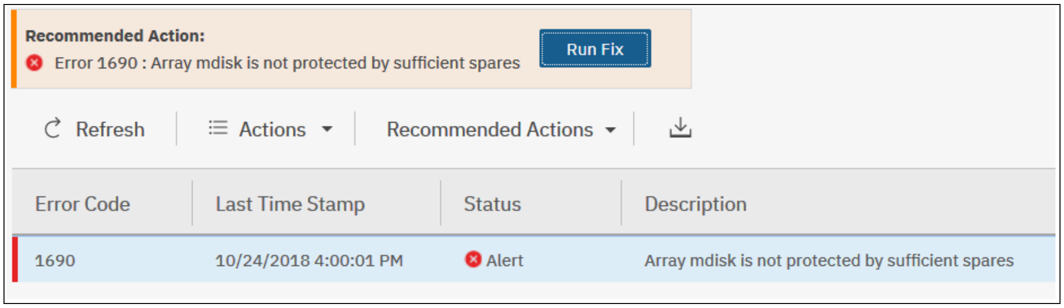


Figure 13-37 Initiate Run Fix procedure from the management GUI

2. The next window shows you which MDisk requires 1 spare to be available (see Figure 13-38). Click **Next**.

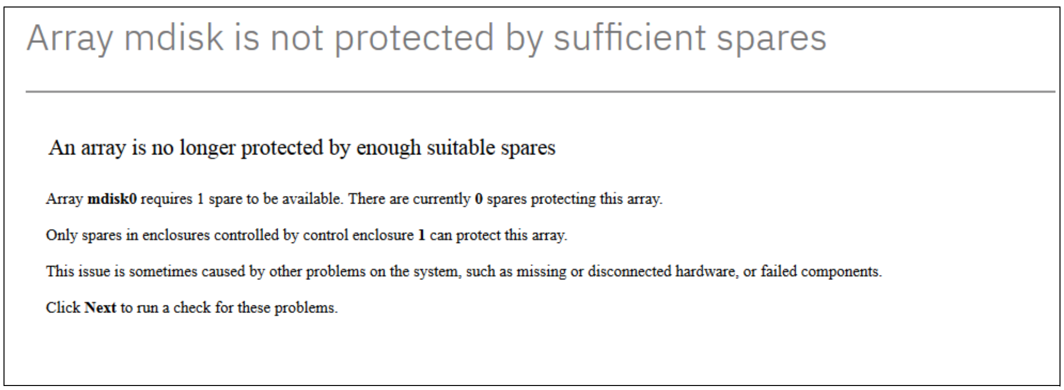


Figure 13-38 Determination of problem

3. In the next window (see Figure 13-39), indicate which of these drives you want to convert to spares by selecting them. Select one drive that has the same capacity and RPM of all other drives within the same mdisk0. Then, click **Next**.

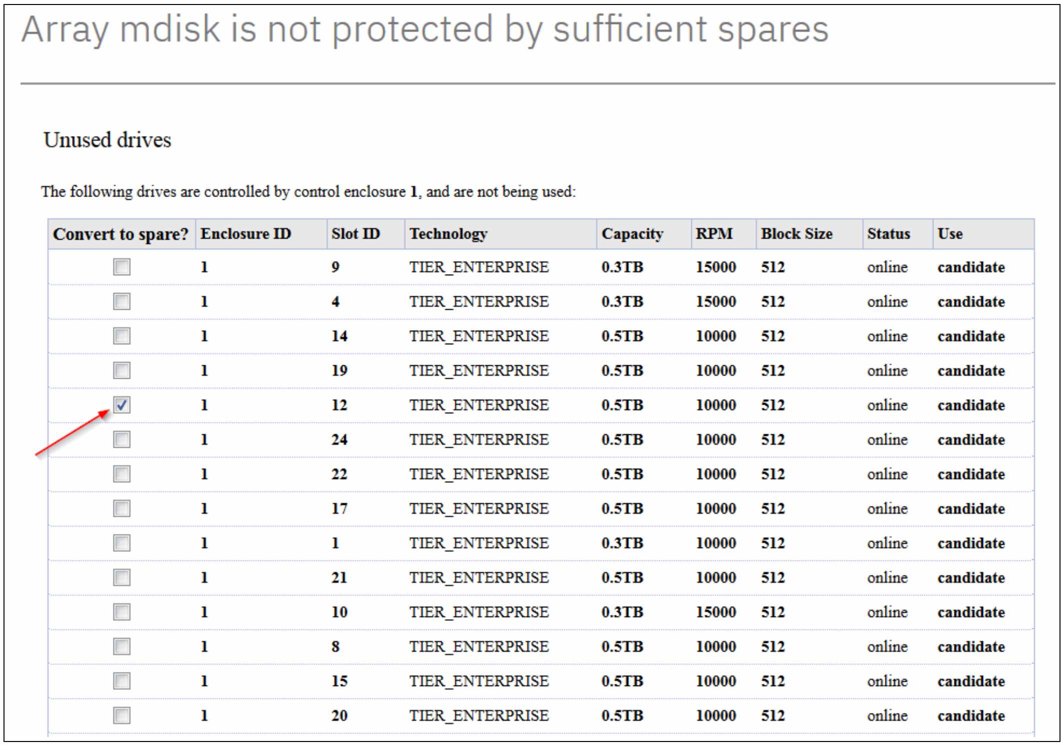


Figure 13-39 Selecting a drive conversion

4. The next window (see Figure 13-40) shows that all requested drives were converted to spares. Click **Next** to check whether the problem is fixed

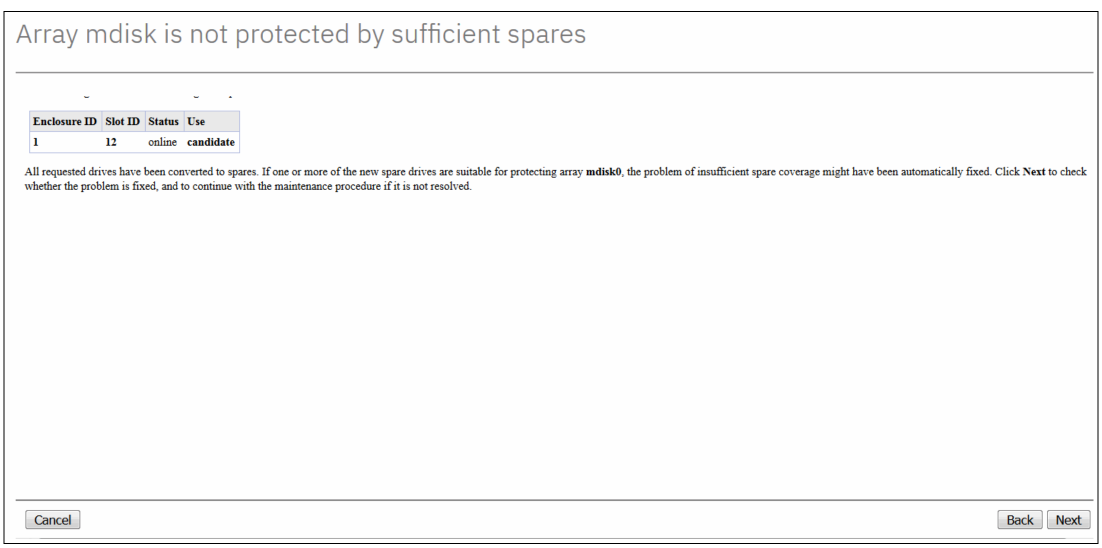


Figure 13-40 Status message of converted drives

5. The event is marked as fixed, and you can safely finish the fix procedure. Click **Close** and the event is removed from the list of events (see Figure 13-41).

Figure 13-41 Correctly finished fix procedure

Resolve 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 designed to be 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.

13.6.3 Event log details

Multiple views of the events and recommended actions are available. The GUI works like a typical Microsoft Windows pop-up menu, so the event log grid is manipulated through the row that contains the column headings (see Figure 13-42). When you click the column icon at the right end of the table heading, a menu for the column choices opens.

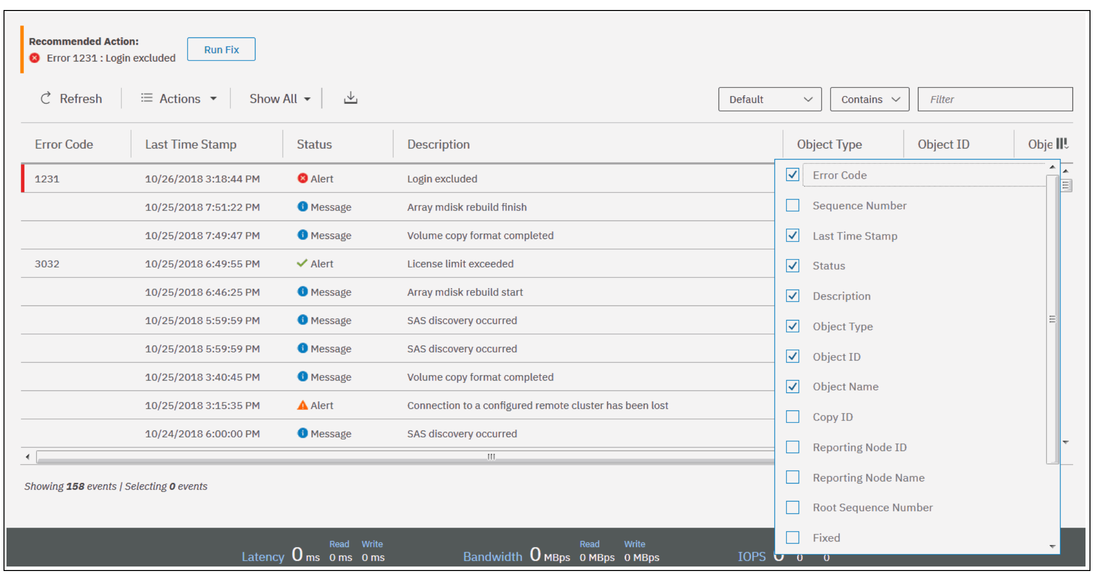


Figure 13-42 Grid options of the event log

Select or remove columns as needed. You can then also extend or shrink the width of the column to fit your window resolution and size. Use this method to manipulate it for most grids in the management GUI of IBM Storwize V7000, not only the events pane.

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 time stamp. All fields have the sequence number, event count, and the fixed state. Using 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 properties and sense data of a specific event, double-click the specific event anywhere in its row.

The properties window opens (see Figure 13-43) with all the relevant sense data. This data includes the first and last time of an event occurrence, worldwide port name (WWPN), and worldwide node name (WWNN), enabled or disabled automatic fix, and so on.

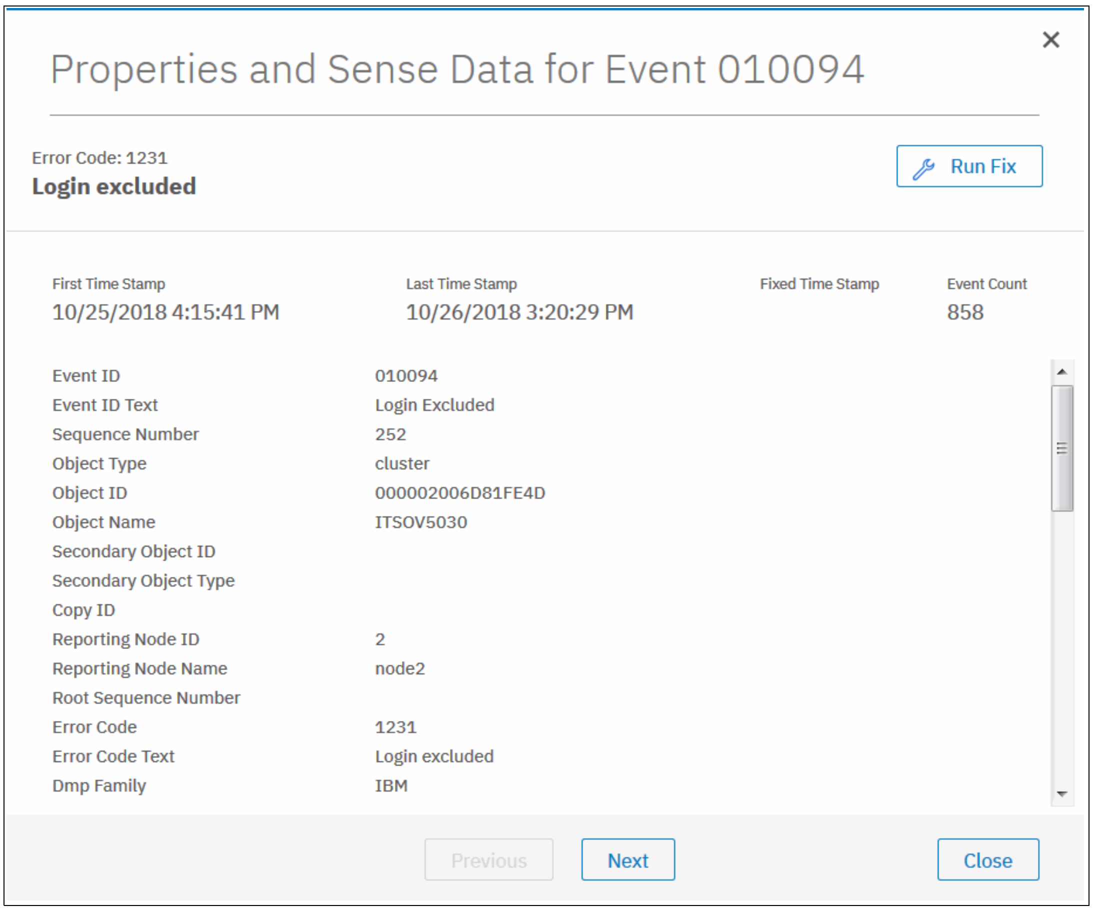


Figure 13-43 Event sense data and properties

For more information about troubleshooting options, search for "Troubleshooting" at this IBM Knowledge Center web page.

13.7 Monitoring

An important step is to correct any issues that are reported by your IBM Storwize V7000 system as soon as possible. Configure your system to send automatic notifications to 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, 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 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, that 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 Storwize V7000 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 specifically 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.

13.7.1 Email notifications and the Call Home function

he Call Home function of IBM Storwize V7000 uses the email notification that is sent to the specific IBM support center. Therefore, the configuration is similar to 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, click **Settings** → **Support** (see Figure 13-44 on page 711).

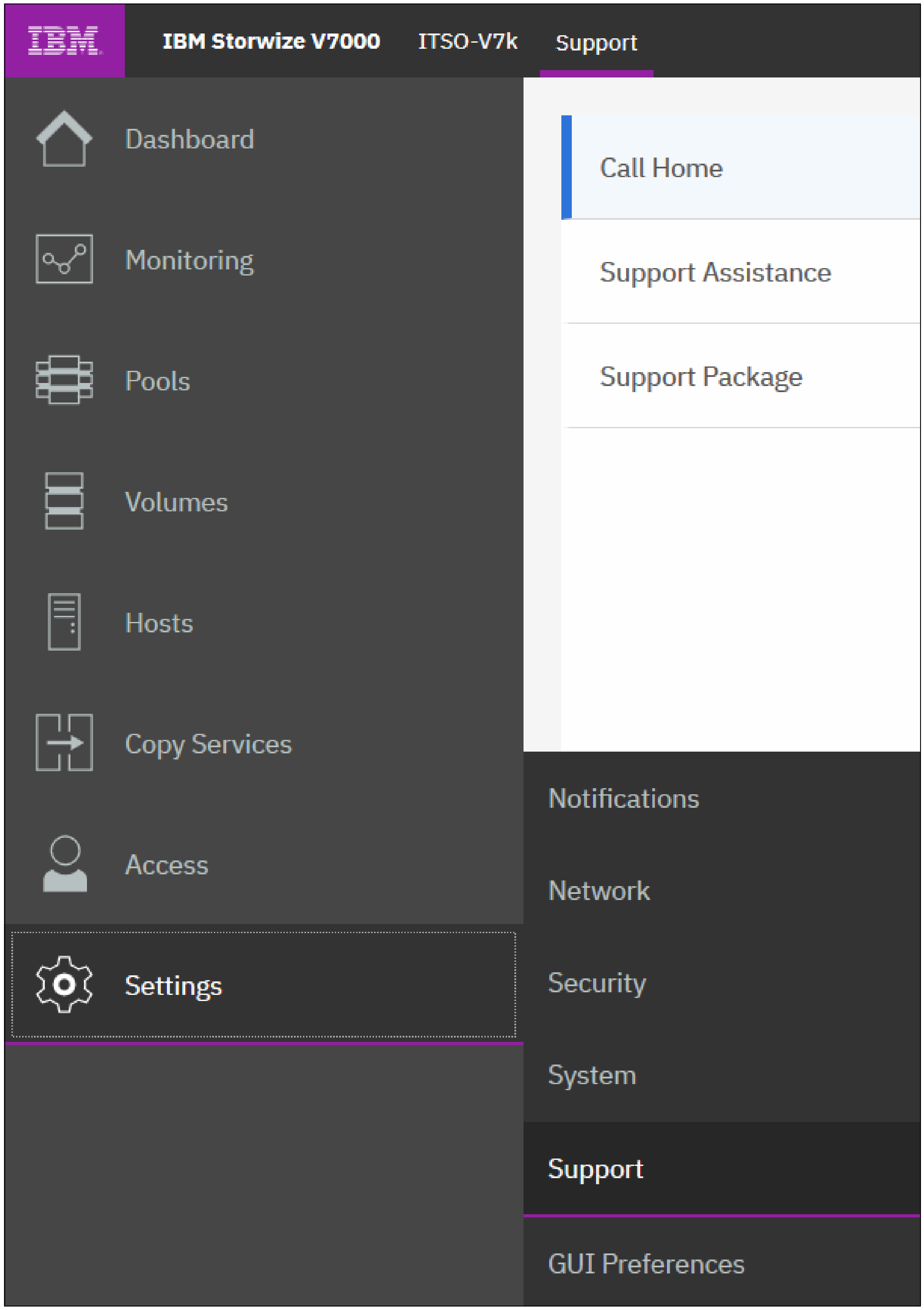


Figure 13-44 Support menu

2. Select **Call Home** and then, click **Enable Notifications** (see Figure 13-45). For more information, see IBM Knowledge Center.

   For the correct functionality 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) using any server in the same network segment.

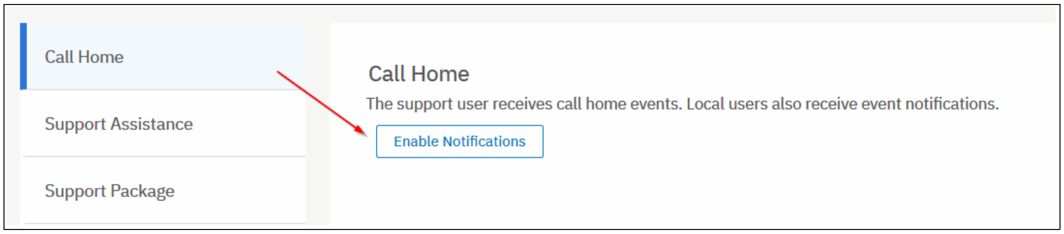


Figure 13-45 Configuration of Call Home notifications

   Figure 13-46 shows the option to enable Cloud Call Home.

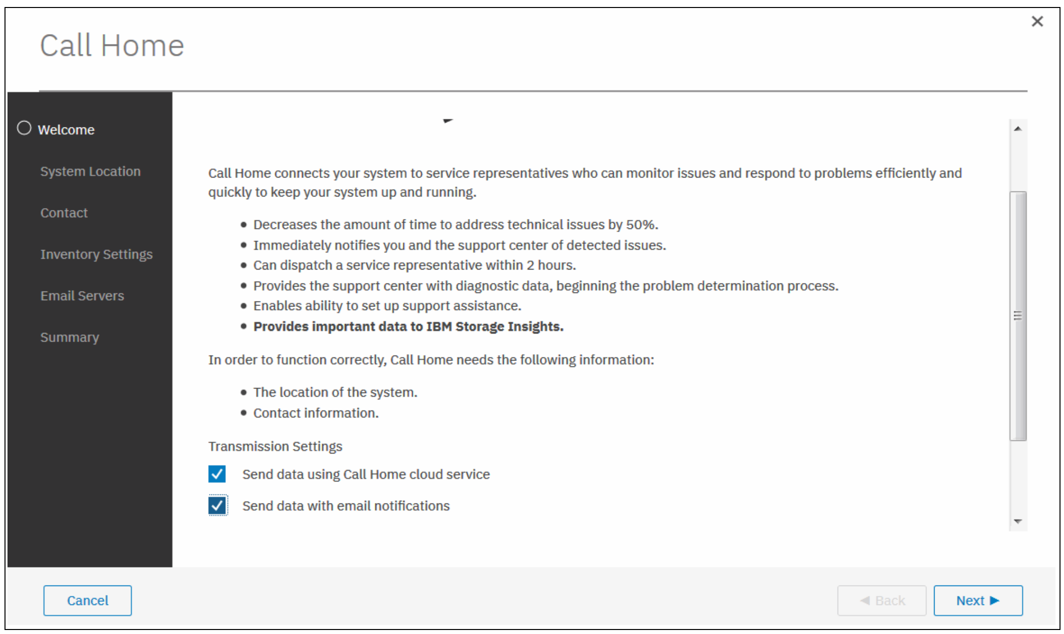


Figure 13-46 Cloud Home service

3. After clicking **Next** on the Welcome window, enter the information about the location of the system (see Figure 13-47 on page 713) and contact information of IBM Storwize V7000 administrator (see Figure 13-48 on page 713) to be contactable by IBM Support. *Always* keep this information current.

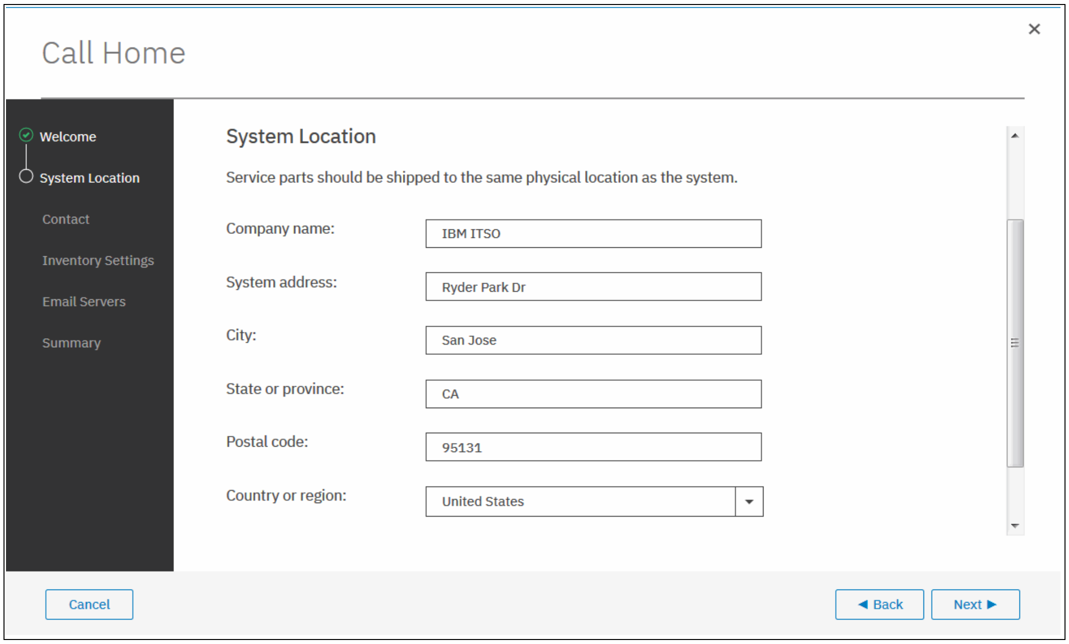


Figure 13-47 Location of the device

Figure 13-48 shows the contact information of the owner.

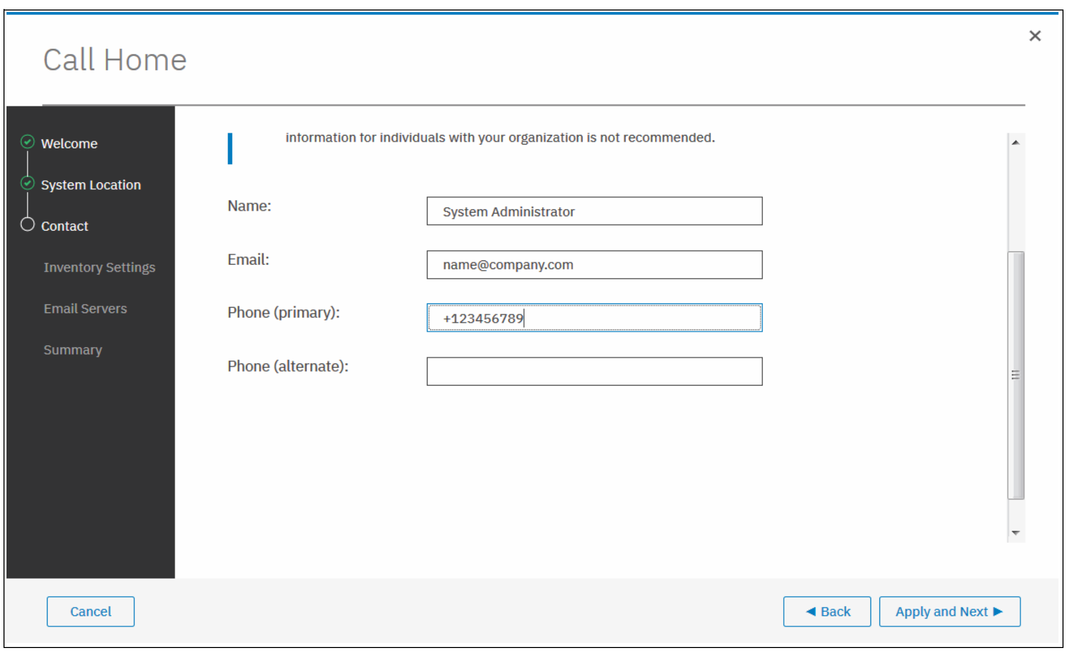


Figure 13-48 Contact information

The next window allows you to enable Inventory Reporting and Configuration Reporting, as shown in Figure 13-49 on page 714.

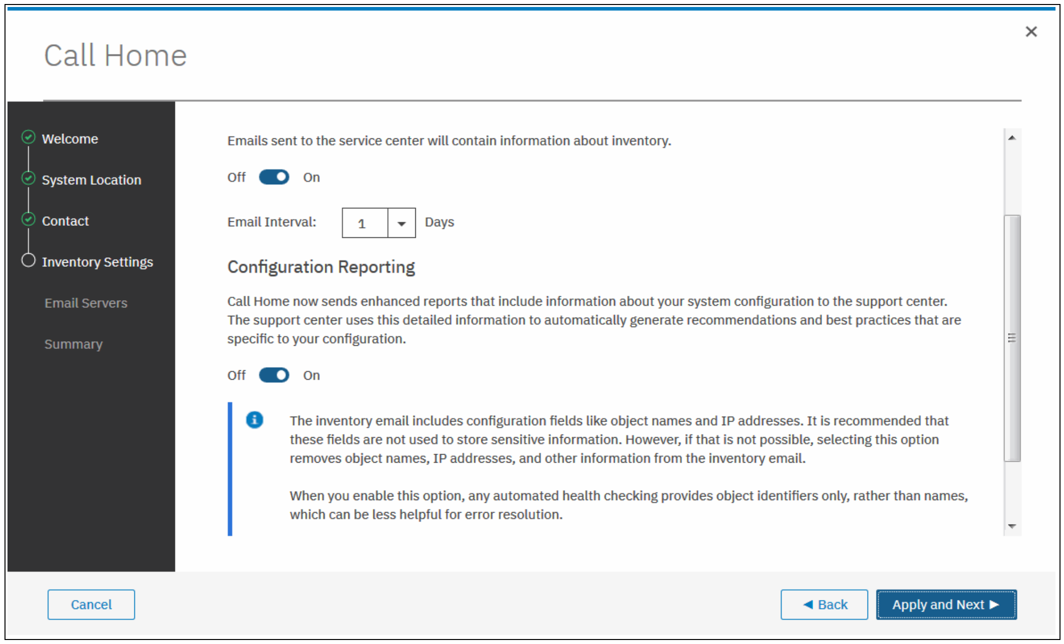


Figure 13-49 Inventory Reporting and Configuration Reporting

4. Configure the SMTP server according to the instructions that are shown in Figure 13-50. When the correct SMTP server is provided, you can test the connectivity by clicking **Ping** to verify it can be contacted. Then, click **Apply and Next**.

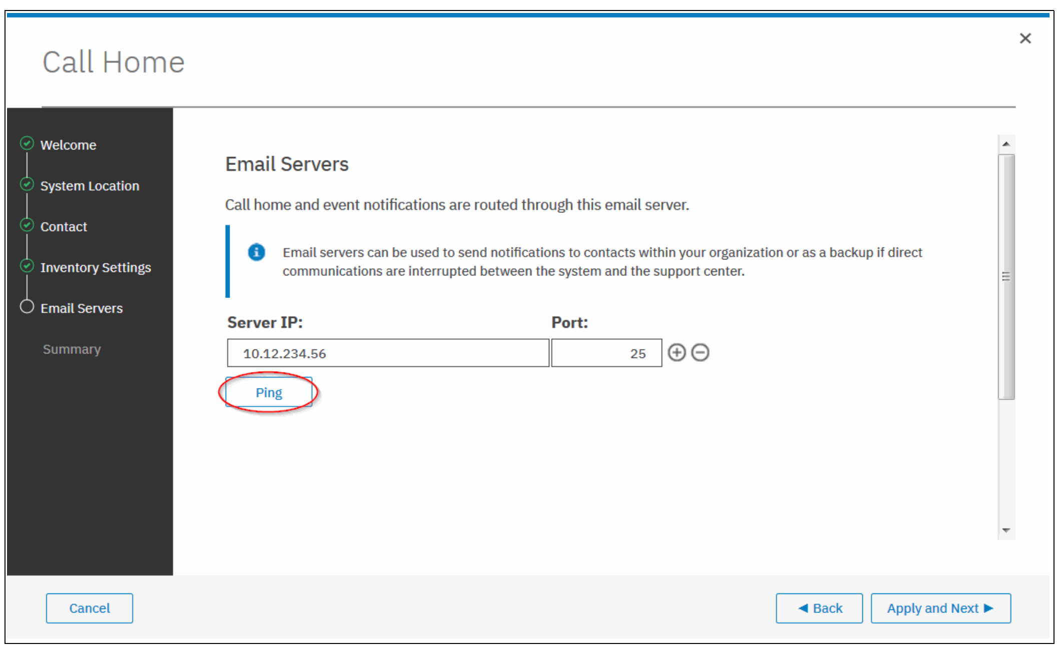


Figure 13-50 Configure email servers and inventory reporting

5. A summary window opens. Verify all of the information and then, click **Finish**. You are returned to the Email Settings window where you can verify email addresses of IBM Support (`callhome0@de.ibm.com`) and optionally add local users who also need to receive notifications (see Figure 13-51.

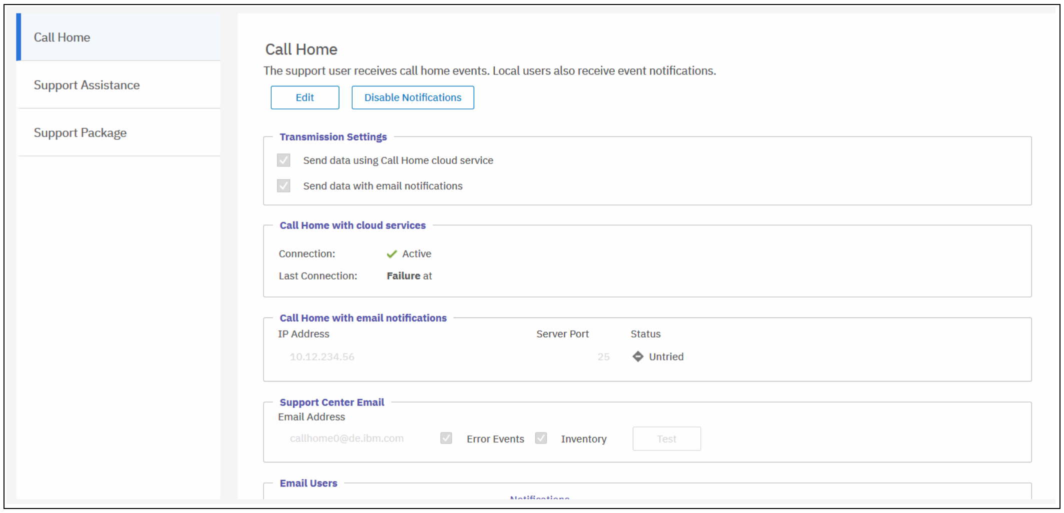


Figure 13-51 Setting email recipients and alert types

The default support email address `callhome0@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 if the hardware or software that you are using requires an update because of any known issue, as described in 13.5, “Health checker feature” on page 701.

Figure 13-51 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 13-52. In the same window, you can define additional email recipients or alter any contact and location details as needed.

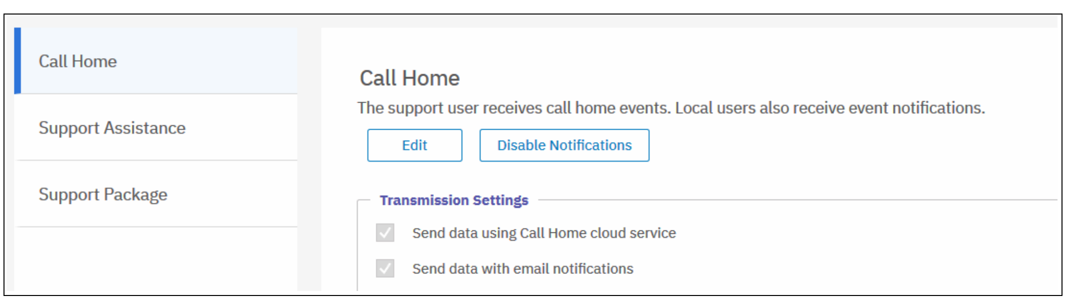


Figure 13-52 Entering edit mode

We strongly suggest that you keep the sending inventory option enabled to IBM Support. However, it might not be of interest to local users, although inventory content can serve as a basis for inventory and asset management.

7. 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 13-53).

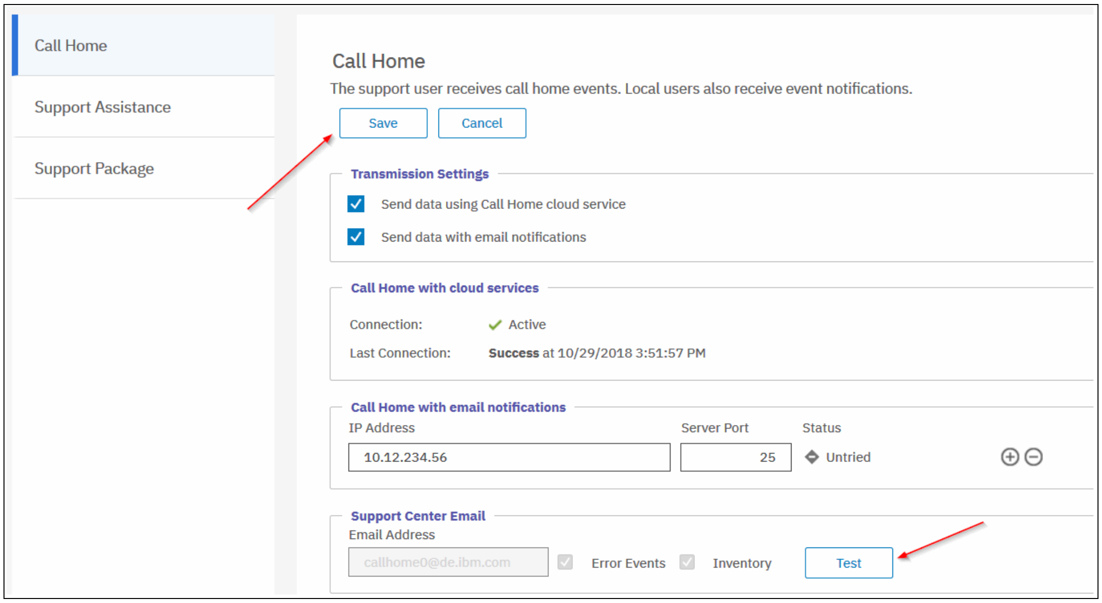


Figure 13-53 Saving modified configuration

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 13-54. This is good practice when performing activities in your environment that might generate errors on your 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 using the CLI **`svctask stopmail`** and **`svctask startmail`** commands.

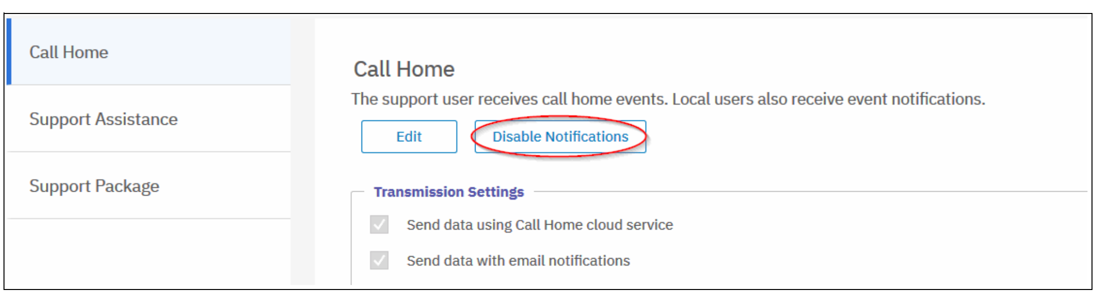


Figure 13-54 Disabling or enabling email notifications

13.7.2 Remote Support Assistance

Introduced with V8.1, Remote Support Assistance allows IBM support to remotely connect to the Storwize V7000 through a secure tunnel to perform analysis, log collection, and software updates. The tunnel can be enabled ad hoc by the client or enable a permanent connection if wanted.

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:

- Call Home is configured with a valid email server.
- A valid service IP address is configured on each node on the Storwize V7000.
- If your Storwize V7000 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 Storwize V7000 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 server must be defined on your Storwize V7000 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 13-55 shows a pop-up window that appears in the GUI after updating to V8.2.1. It prompts you to configure your Storwize V7000 for Remote Support. You can select to not enable it, open a tunnel when needed, or open a permanent tunnel to IBM.

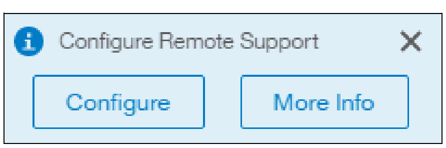


Figure 13-55 Prompt to configure Remote Support Assistance

You can choose to configure SAN Volume Controller, learn more information about the feature, or close the window by clicking **X**. Figure 13-56 shows how you can find the Setup Remote Support Assistance if you closed the window.

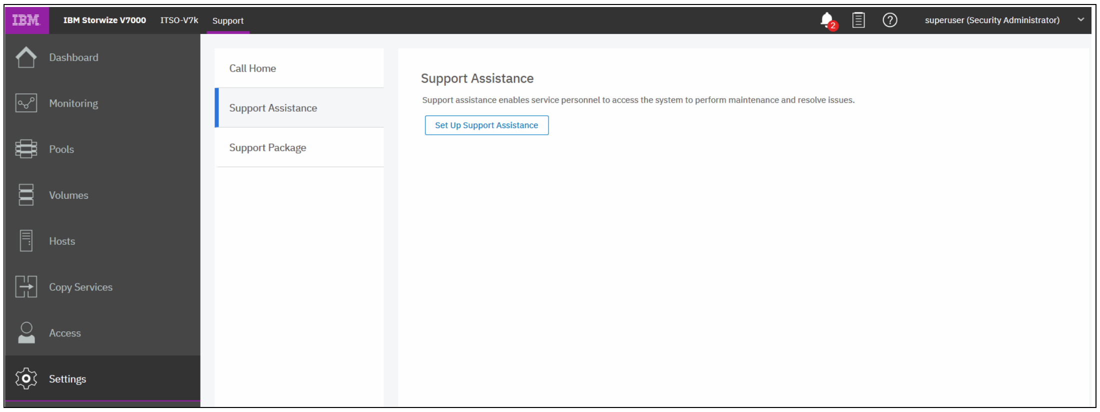


Figure 13-56 Support Assistance menu

Choosing to set up support assistance opens a wizard to guide you through the following configuration process:

1. Figure 13-57 on page 719 shows the first wizard window. Choose not to enable remote assistance by selecting **I want support personnel to work on-site only** or enable remote assistance by choosing **I want support personnel to access my system both on-site and remotely**. Click **Next**.

> **Note:** Selecting **I want support personnel to work on-site only** does not entitle you to expect IBM support to attend on-site for all issues. Most maintenance contracts are for customer-replaceable units (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.

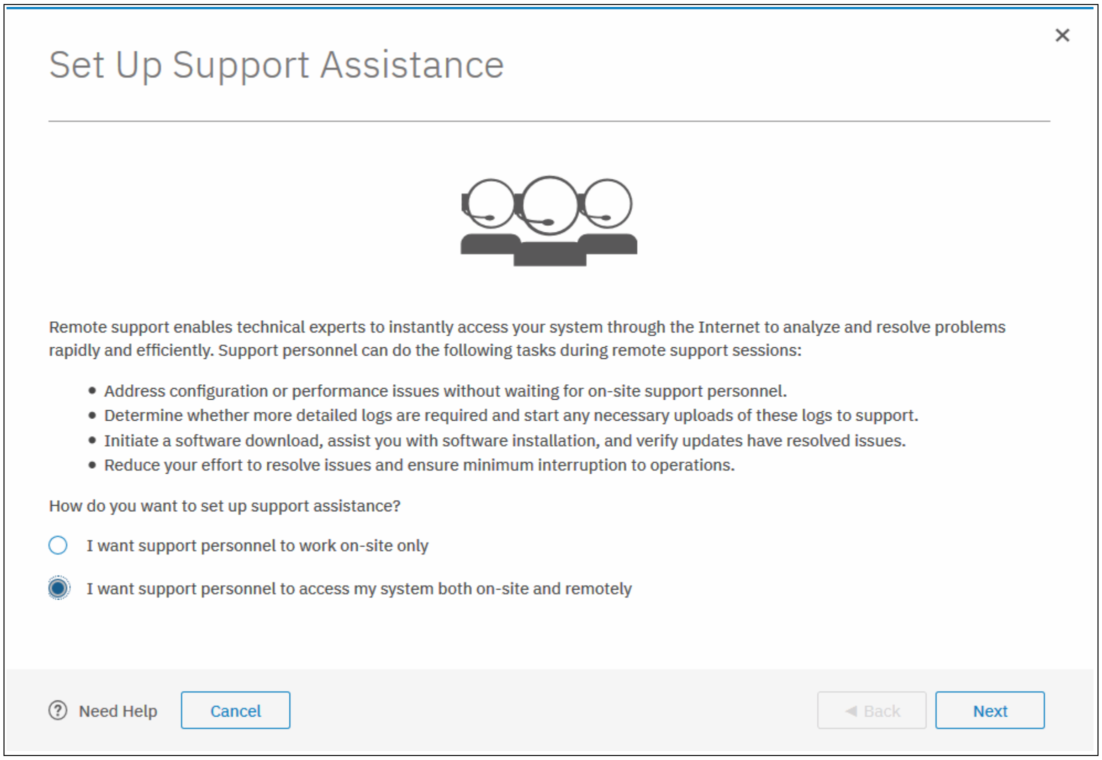


Figure 13-57 Support wizard enable or disable

2. The next window, shown in Figure 13-58, lists the IBM Support center's IP addresses and SSH port that need to be open in your firewall. You can also define a Remote Support Assistance Proxy if you have multiple Storwize V7000 or SAN Volume Controllers in the data center, which allows for firewall configuration only being required for the Proxy Server rather than every storage system. We do not have a proxy server and leave the field blank. Then, click **Next**.

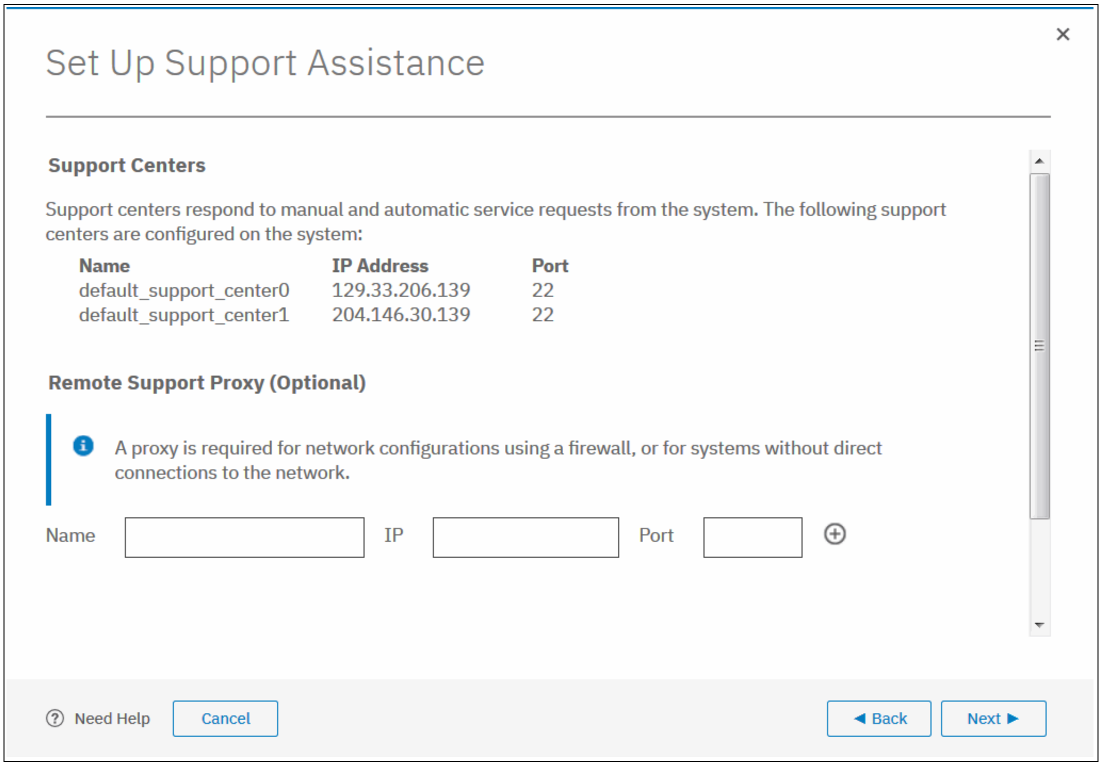


Figure 13-58 Support wizard proxy setup

3. The next window prompts if you want to open a tunnel to IBM permanently, which allows IBM to connect to your Storwize V7000 **At Any Time**, or **On Permission Only**, as shown in Figure 13-59 on page 721. **On Permission Only** requires a storage administrator to log on to the GUI and enable the tunnel when required. Click **Finish**.

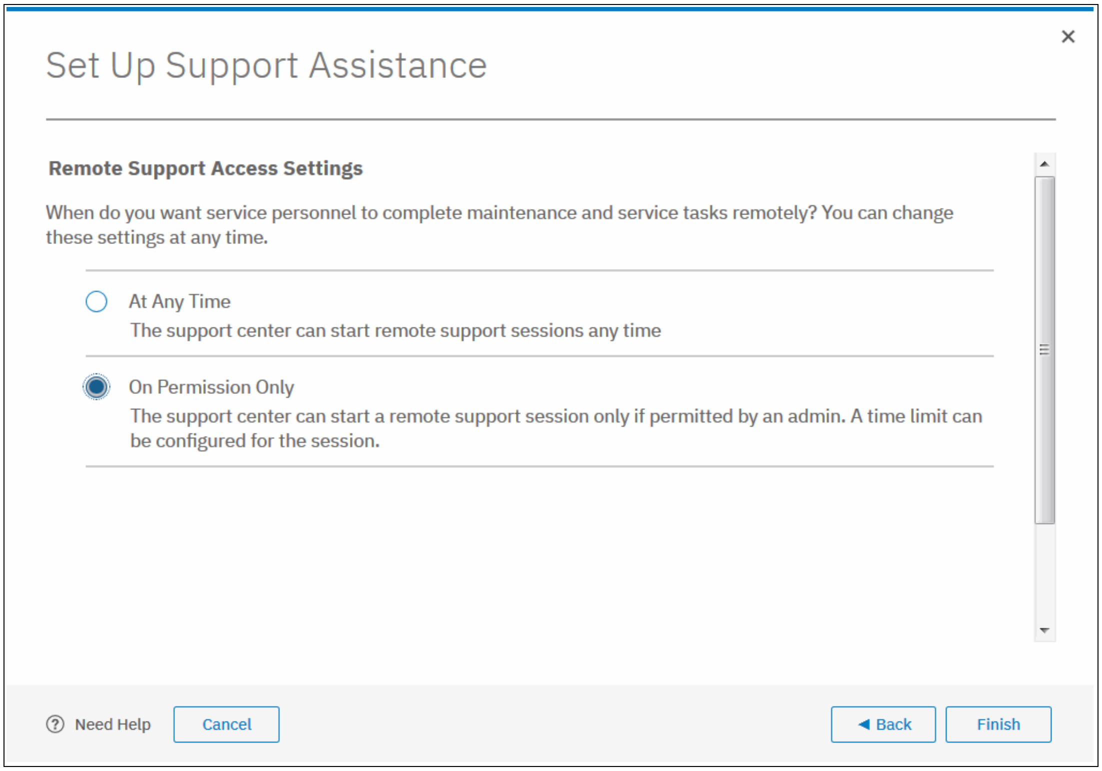


Figure 13-59 Support wizard access choice

4. After completing the remote support setup, you can view the status of any remote connection, start a new session, test the connection to IBM, and reconfigure the setup. As shown in Figure 13-60, we successfully tested the connection. Click **Start New Session** to open a tunnel through which IBM Support can connect.

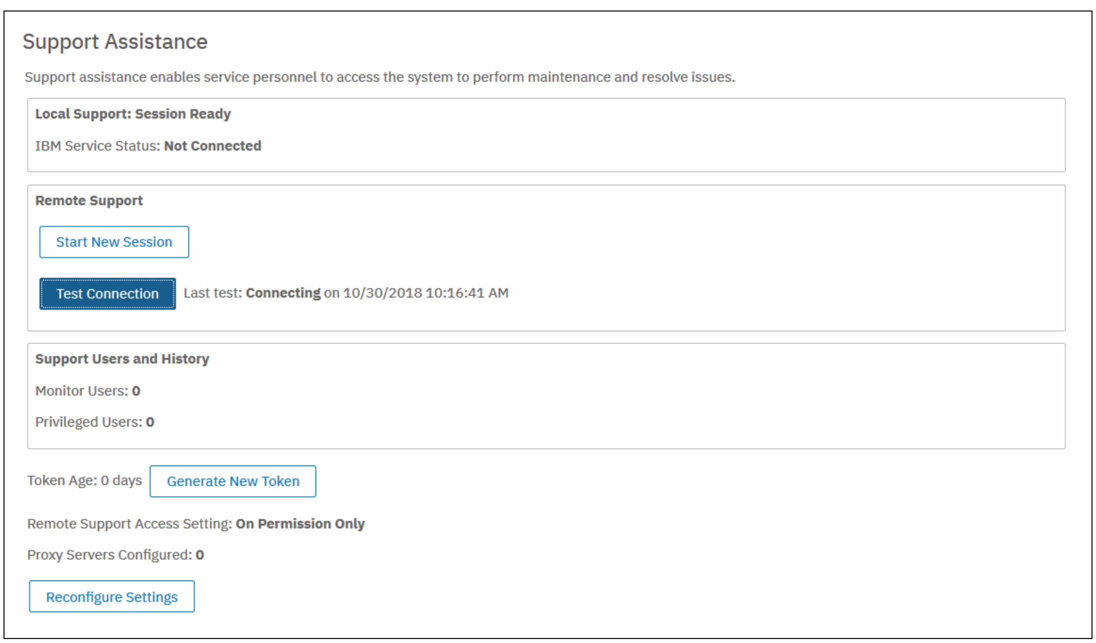


Figure 13-60 Support Status and session management

5. A pop-up window prompts for how long you want the tunnel to remain open if no activity occurs by setting a timeout value. As shown in Figure 13-61, the connection is established and waits for IBM Support to connect.

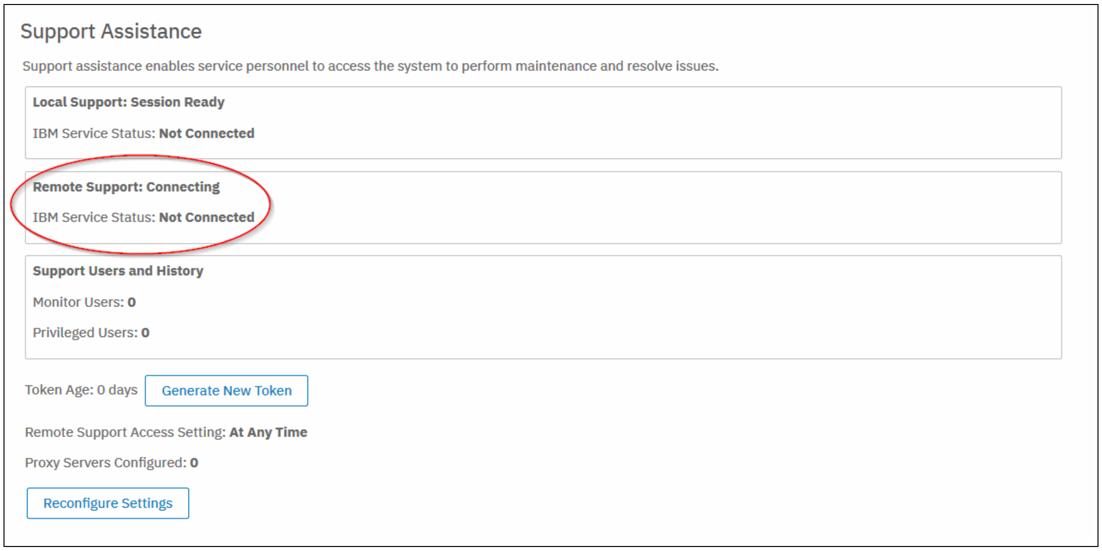


Figure 13-61 Remote Assistance tunnel connection

13.7.3 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 Storwize V7000.

You can configure an SNMP server to receive various informational, error, or warning notifications by entering the following information (see Figure 13-62 on page 723):

- IP Address

  The address for the SNMP server.

- Server Port

  The remote port number for the SNMP server. The remote port 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:** 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.

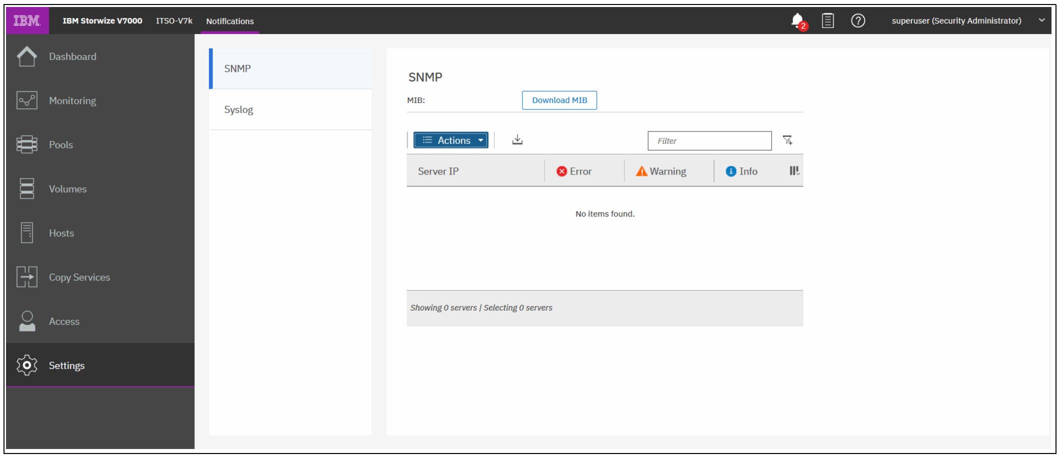


Figure 13-62 SNMP configuration

To add an SNMP server, click **Actions** → **Add** and complete the Add SNMP Server window, as shown in Figure 13-63 on page 724. To remove an SNMP server, click the line with the server you want to remove, and select **Actions** → **Remove**.

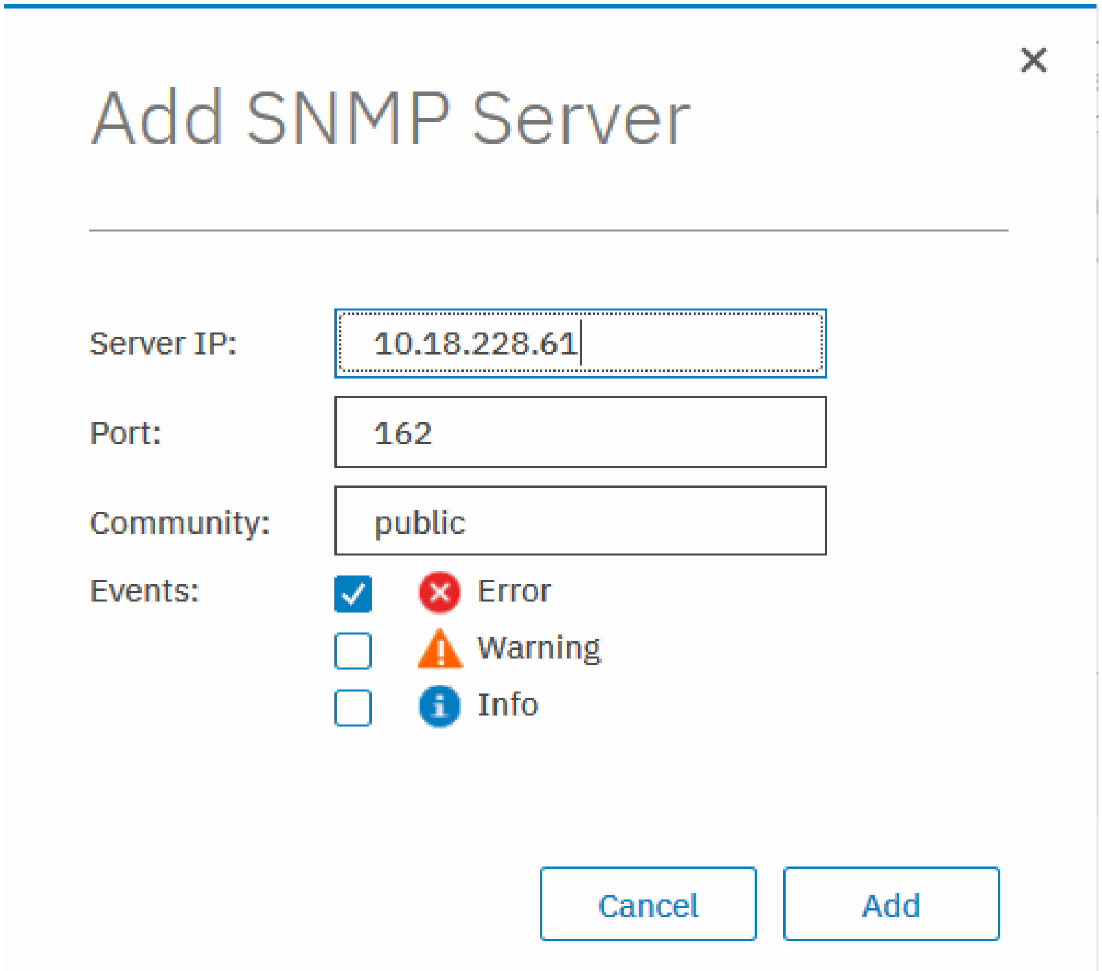


Figure 13-63 Add SNMP Server

13.7.4 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 entering the following information (see Figure 13-64 on page 725):

- 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.

- 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 detail about the event.

– The expanded format provides more details about the event.

- 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 whether any corrective action is necessary.

  > **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.

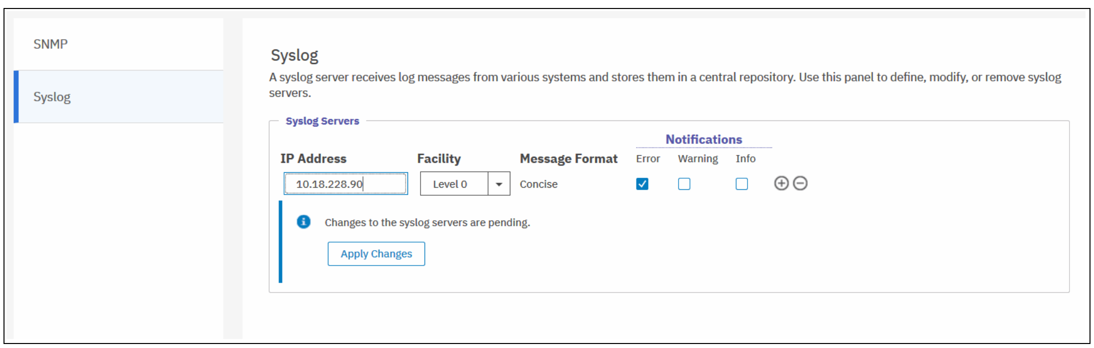


Figure 13-64 Syslog configuration

To remove a syslog server, click the Minus sign (**-**).

To add another syslog server, click the Plus sign (**+**).

The syslog messages can be sent in concise message format or expanded message format.

Example 13-4 shows a compact format syslog message.

Example 13-4 Compact syslog message example

```
IBM2076 #NotificationType=Error #ErrorID=077102 #ErrorCode=1091 #Description=Node
Double fan failed #ClusterName=V7000G2_1 #Timestamp=Wed Jul 02 08:00:00 2017 BST
#ObjectType=Node #ObjectName=Node1 #CopyID=0 #ErrorSequenceNumber=120
```

Example 13-5 shows an expanded format syslog message.

Example 13-5 Full format syslog message example

```
IBM2076 #NotificationType=Error #ErrorID=077102 #ErrorCode=1091 #Description=Node
Double fan failed #ClusterName=V7000G2_1 #Timestamp=Wed Jul 02 08:00:00 2017 BST
#ObjectType=Node #ObjectName=Node1 #CopyID=0 #ErrorSequenceNumber=120 #ObjectID=2
#NodeID=2 #MachineType=2076624#SerialNumber=1234567 #SoftwareVersion=8.1.0.0
```

```
(build 13.4.1709291021000)#FRU=fan 31P1847
#AdditionalData(0->63)=000000004600000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000#Additional
Data(64-127)=0000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000
```

13.8 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 a Secure Shell (SSH) session, management GUI, or Remote Support Assistance. It provides the following entries:

- Identity of the user who issued the action command
- Name of the actionable command
- Time stamp of when the actionable command was issued on the configuration node
- Parameters that were issued 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 13-65 shows the access to the audit log. Click **Audit Log** in the left menu to see which configuration CLI commands were run on IBM Storwize V7000 system.

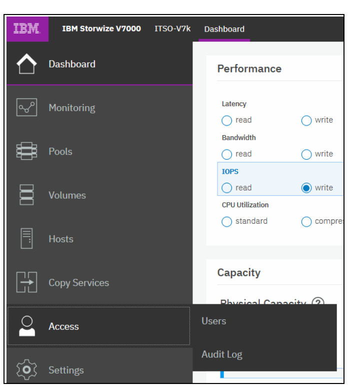


Figure 13-65 Audit Log from Access menu

Figure 13-66 shows an example of the audit log after a volume is created, with a command highlighted. The **Running Tasks** button is available at the top of the window in the status pane that shows the format of this volume. If you click **Running Tasks**, the progress of the running tasks can be displayed by clicking the associated **View** button.

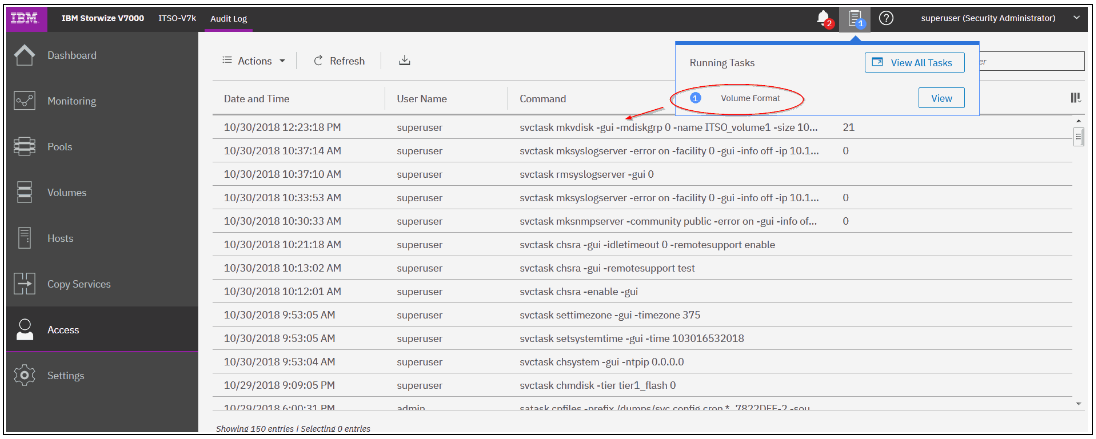


Figure 13-66 Audit log

Changing the view of the Audit Log grid is possible by right-clicking column headings or clicking the sign in the upper right corner (see Figure 13-67). 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.

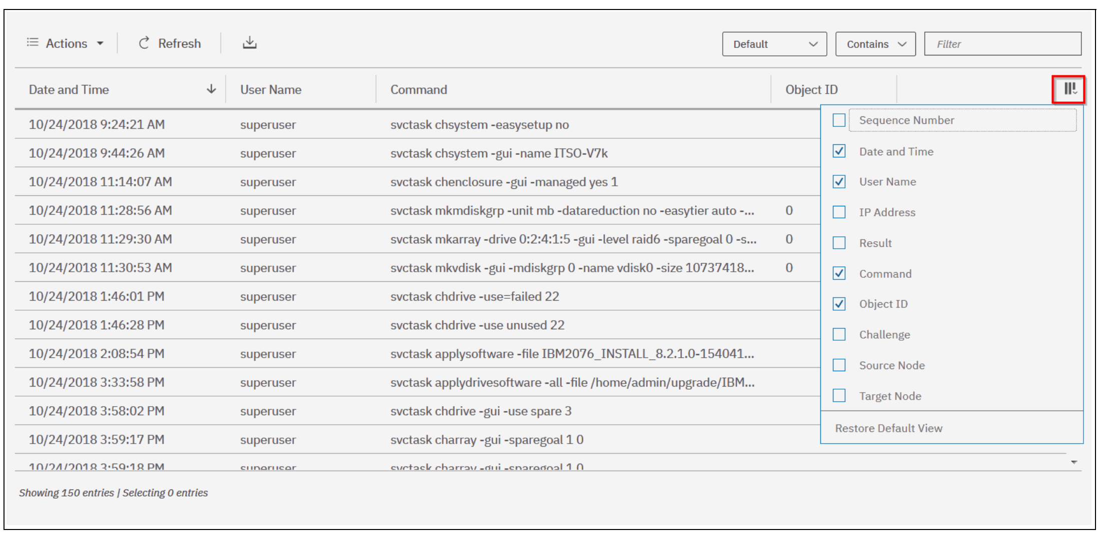


Figure 13-67 How to change audit log column headings

13.9 Collecting support information by using the GUI and CLI

If you encounter a problem and contact the IBM Support Center, you are most likely asked to provide a support package. You can collect and upload this package from the **Settings** → **Support** menu.

13.9.1 Collecting information by using the GUI

To collect information using the GUI, complete the following steps:

1. Click **Settings** → **Support** and the Support Package tab (see Figure 13-68).

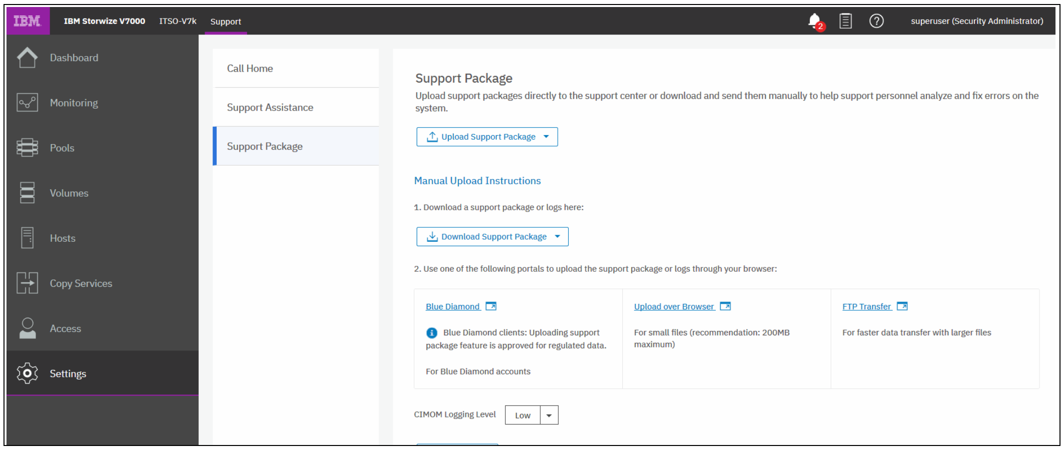


Figure 13-68 Support Package option

2. Click the **Upload Support Package** button.

   Assuming that the problem encountered was an unexpected node restart that logged a 2030 error, we 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 reboots, 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.

3. The Upload Support Package window provides four options for data collection. If you were contacted by IBM Support because your system called home or you manually opened a call with IBM Support, you receive a *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 13-69). In our example, we entered our PMR number, selected snap type 3 (option 3) because this automatically collects the statesave that were created at the time the node restarted, and clicked **Upload**.

> **Tip:** To open a service request online, see the IBM Support Service requests and PMRs web page.

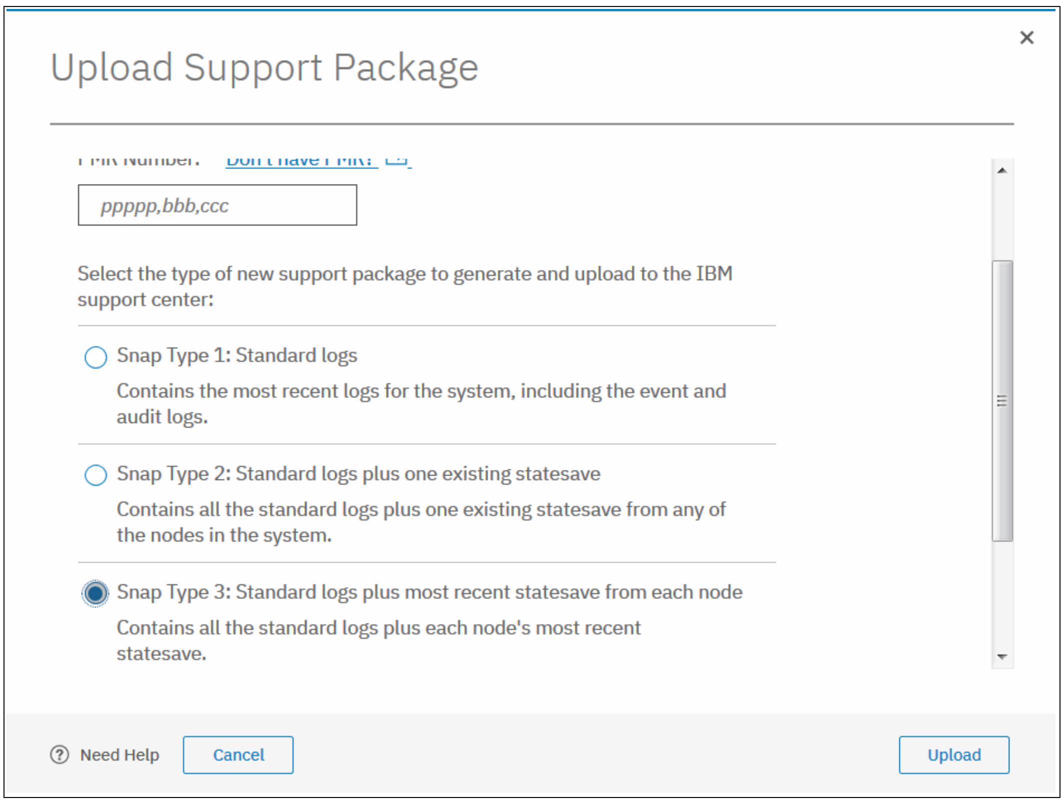


Figure 13-69 Upload Support Package window

4. The procedure to generate the snap on a Storwize V7000 system, including the most recent statesave from each node canister, starts. This process might take a few minutes (see Figure 13-70).

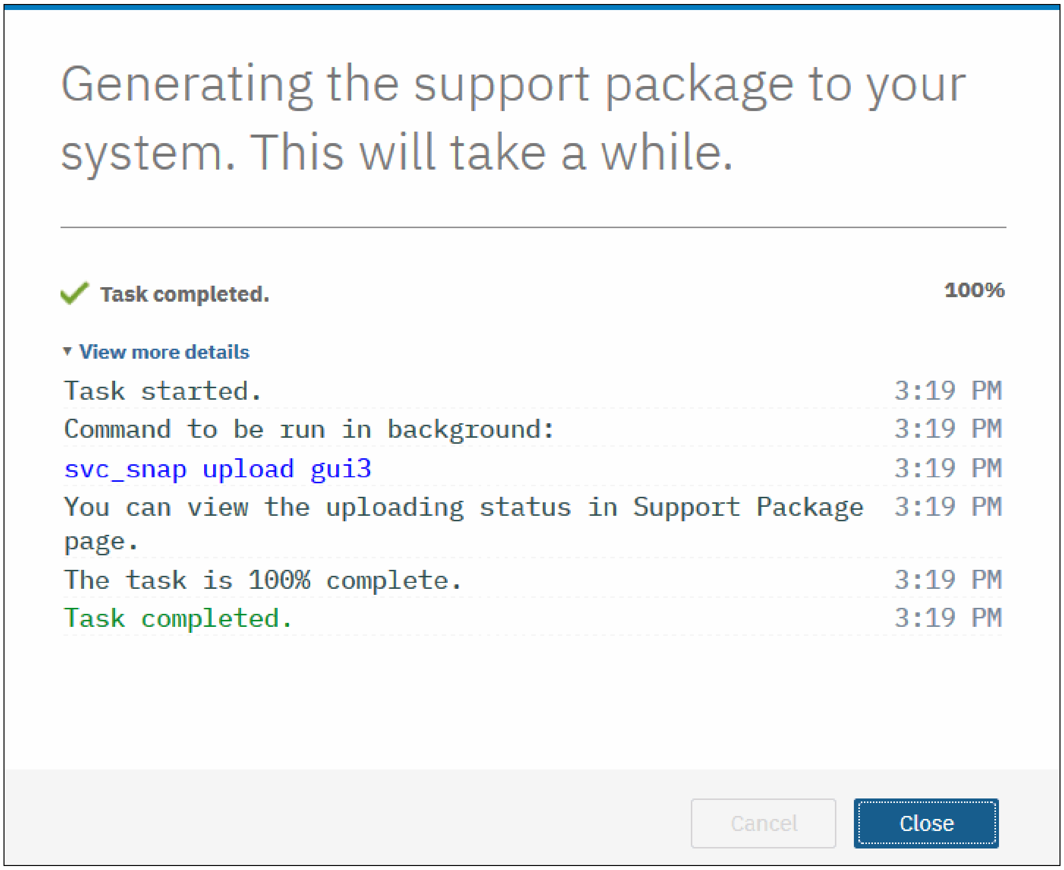


Figure 13-70 Task detail window

13.9.2 Collecting logs 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 issue the **svc_snap** command that matches the type of snap 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
   ```

2. 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 13-6.

Example 13-6 The svc_snap command

```
ssh superuser@9.173.156.250
Password:
IBM_Storwize:ITSO-V7k:superuser>svc_snap upload pmr=12345,000,866 gui3
```

3. 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 named 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).

4. The generated file can then be retrieved from the GUI clicking **Settings** → **Support** → **Manual Upload Instructions** twisty → **Download Support Package** and then, clicking **Download Existing Package**, as shown in Figure 13-71.

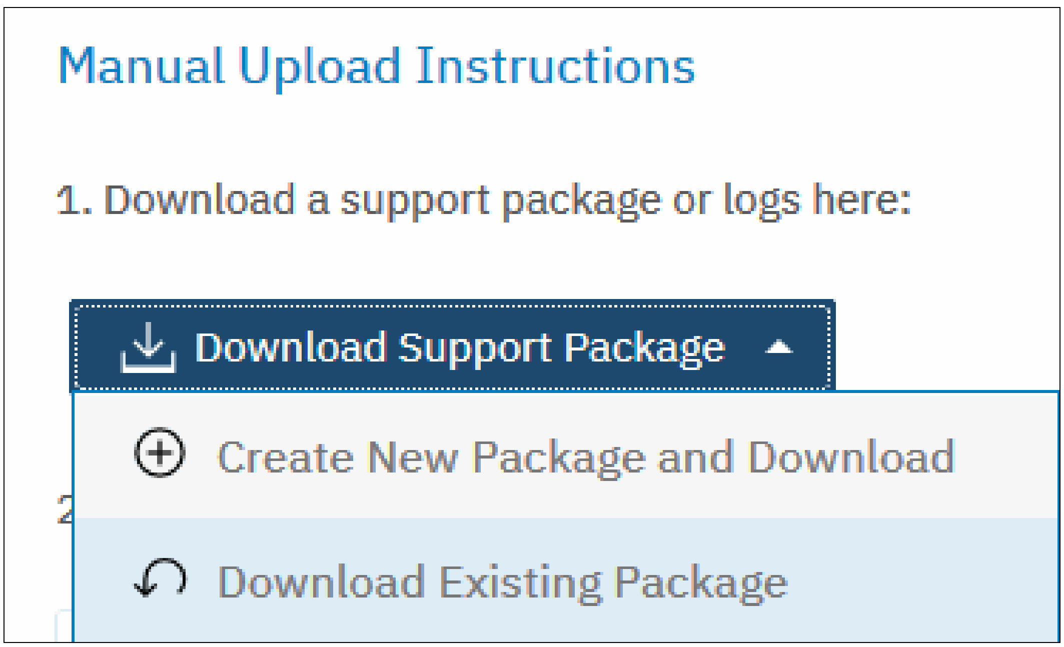


Figure 13-71 Downloaded Existing Package

5. Click in the **Filter** box and enter `snap` to see a list of snap files, as shown in Figure 13-72. Locate the exact name of the snap that was generated by using the **`svc_snap`** command that was issued earlier. Select that file, and click **Download**.

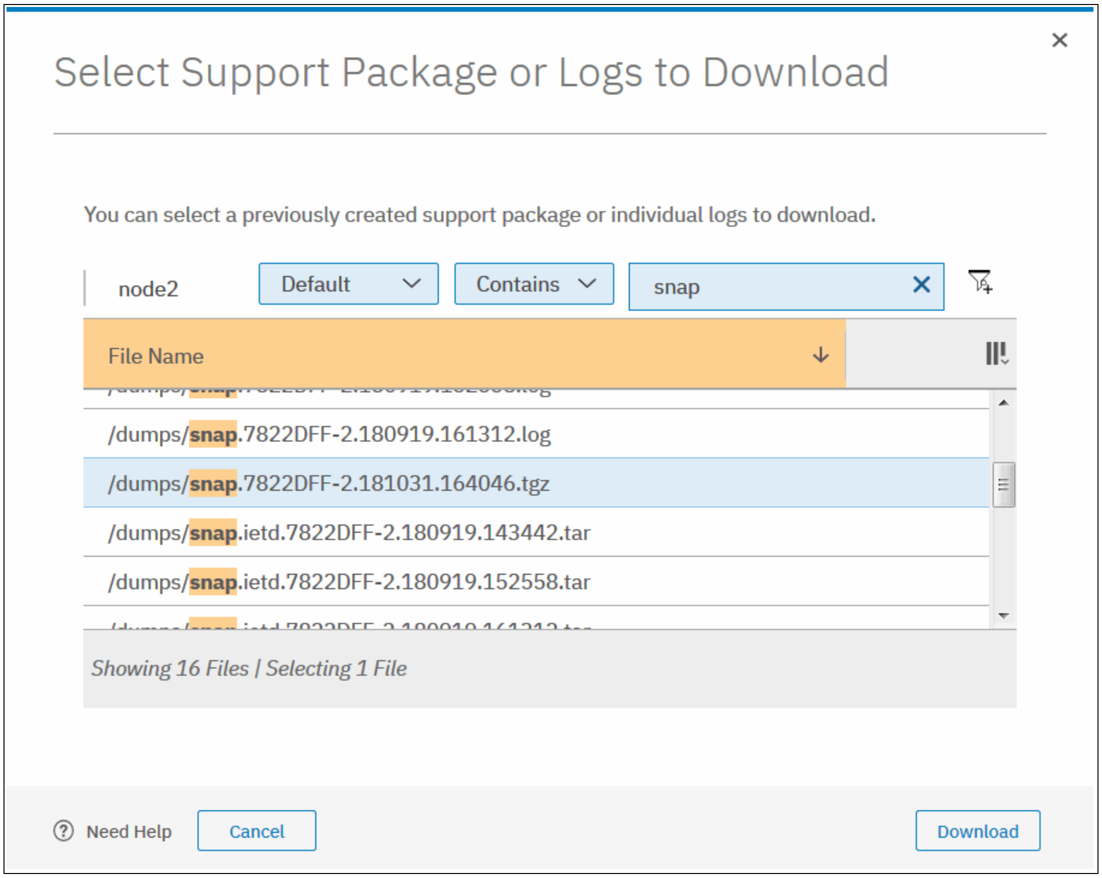


Figure 13-72 Filtering on snap to download

6. Save the file to a folder of your choice on your workstation.

13.9.3 Uploading files to the Support Center

If you chose to not have the Storwize V7000 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 problem management report (PMR). The PMR is also known as a *service request* and is a mandatory requirement when uploading.

To upload information, complete the following steps:

1. Using a web browser, navigate to ECuRep:

   https://www.secure.ecurep.ibm.com/app/upload

   This link takes you to the Secure Upload page (see Figure 13-73).

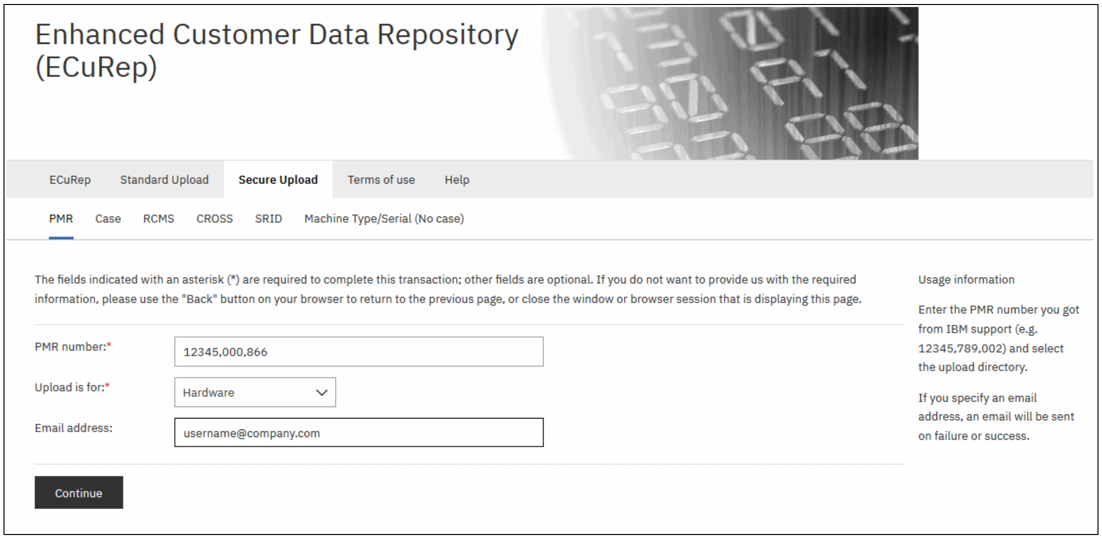


Figure 13-73 ECuRep details

2. Complete the required fields:

   – **PMR number** as 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, we recommend entering 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 13-74).

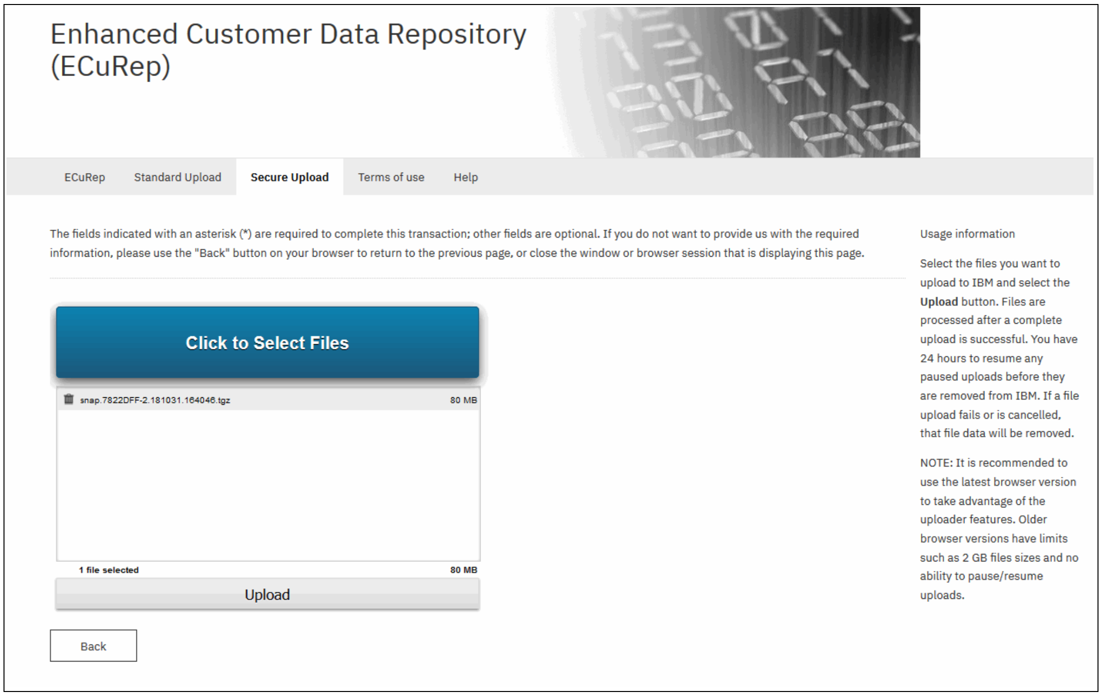


Figure 13-74 ECuRep File upload

4. Select one or more files, click **Upload** to continue, and follow the directions.

13.10 Service Assistant Tool

The Service Assistant Tool (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 IBM Storwize V7000 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 (CIM) object manager (CIMOM) and updating the worldwide node name (WWNN)
- Details about node error codes
- Details about the hardware, such as IP address, Media Access Control (MAC) addresses

The SAT GUI is available by using a service assistant IP address that is configured on each Storwize V7000 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 URL of the service IP of the config node into any web browser and click **Service Assistant Tool** (see Figure 13-75).

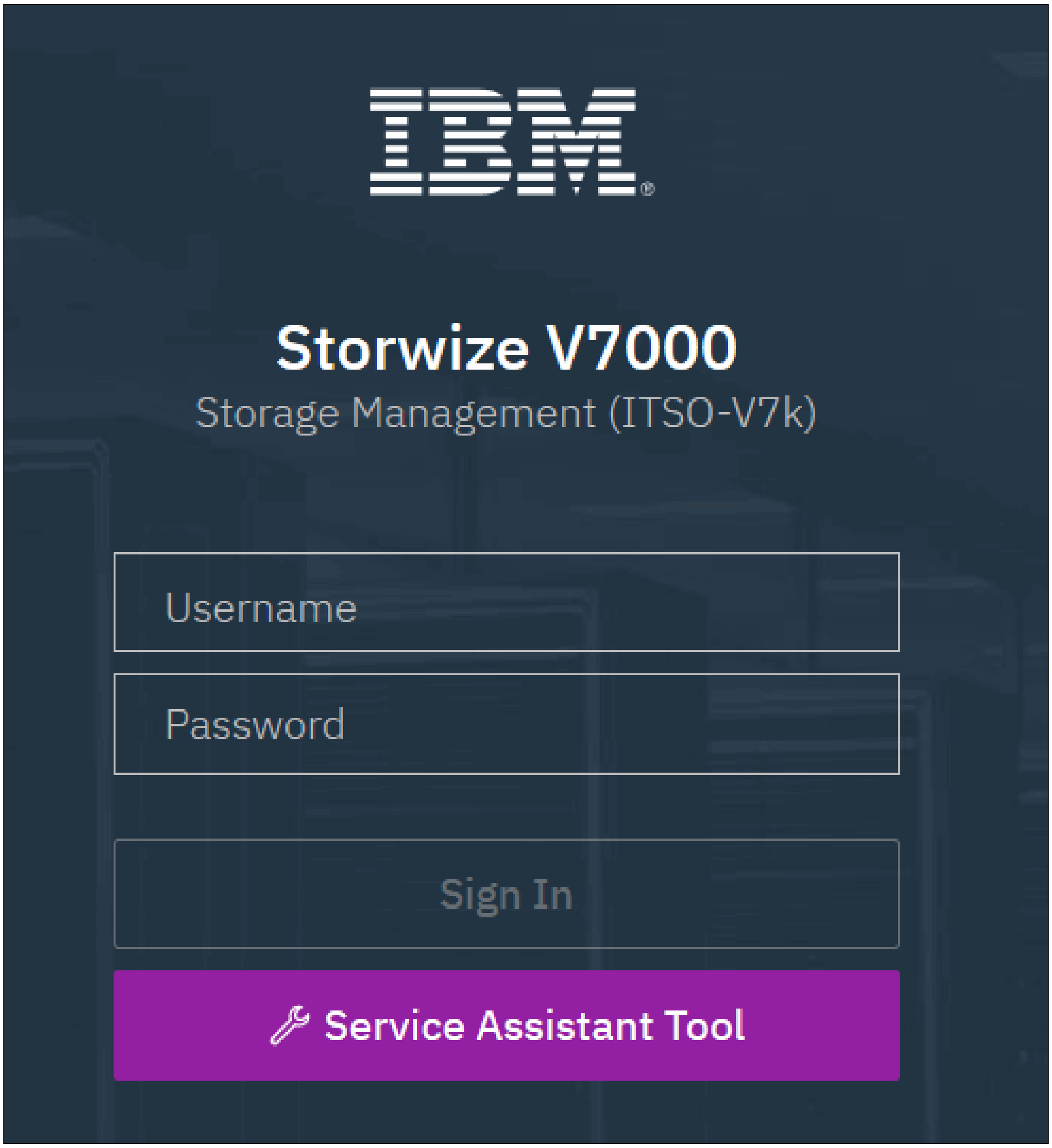


Figure 13-75

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 and should be configured on all nodes of the cluster.

To open the SAT GUI, enter one of the following URLs into a web browser:

- `http(s)://`*`<cluster IP address of your cluster>`*`/service`
- `http(s)://`*`<service IP address of a node>`*`/service`
- `http(s)://`*`<service IP address of config node>`* **`click Service Assistant Tool`**

Complete the following steps to access the SAT:

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 13-76.

Figure 13-76 Service Assistant Tool Login GUI

2. After you are logged in, you see the Service Assistant Home window, as shown in Figure 13-77. 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.

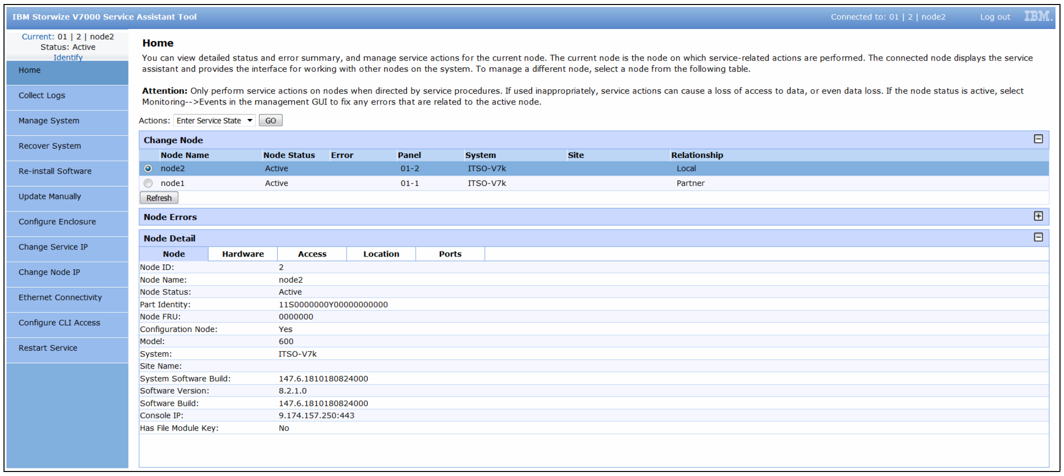


Figure 13-77 Service Assistant Tool GUI

3. The current node canister is displayed in the upper left corner of the GUI. As shown in Figure 13-77, 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 only be carried out if directed to do so by IBM Support.

For more information about how to use the SAT, see this website.

A

Performance data and statistics gathering

This appendix provides a brief overview of the performance analysis capabilities of the IBM Storwize V7000 and IBM Spectrum Virtualize V8.2. 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 Storwize V7000, see *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

This appendix includes the following topics:

Storwize V7000 performance overview

Storage virtualization with the IBM Spectrum Virtualize provides many administrative benefits. In addition, it can provide an increase in performance for some workloads. The caching capability of the 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 to 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. The recommendation 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 suggestion also assumes a core-edge design, where the hosts are on the edges and the Storwize V7000 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 suggestion assumes a multiple-switch SAN fabric design.

- Node-to-node ISL oversubscription

  This area does not apply to Storwize V7000 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 split-cluster capability that was made available since V6.3 (HyperSwap).

  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 requires 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.

- ISL trunking/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. Although the IBM Subsystem Device Driver Path Control Module (SDDPCM), related products, and most vendor multipathing software can support more paths, the Storwize V7000 expects a maximum of eight paths. In general, you see only an effect on performance from more paths than eight. Although the IBM Spectrum Virtualize can work with more than eight paths, this design is technically unsupported.

- Do not intermix dissimilar array types or sizes

  Although the IBM Spectrum Virtualize supports an intermix of differing storage within storage pools, it is best 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 a 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 is designed to run on commodity hardware (mass-produced Intel-based processors [CPUs] with mass-produced expansion cards) and to provide distributed cache and a scalable cluster architecture. One of the main goals of this design was to use refreshes in hardware. Currently, the Storwize V7000 cluster is scalable up to eight nodes (four control enclosures).

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 Storwize V7000 ports. The Storwize V7000 is one of the few devices that can drive ports to their limits on average; therefore, 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 Storwize V7000 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 using the **startstats** command.

The statistics files for volumes, managed disk (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 then 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.

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`.

Use the **startstats** command to start the collection of statistics, as shown in Example A-1.

Example A-1 The startstats command

```
IBM_Storwize:ITSO-V7k:superuser>startstats -interval 2
```

This command starts statistics collection and gathers data at 2-minute intervals.

To verify that the statistics collection interval, display the system properties again, as shown in Example A-2.

Example A-2 Statistics collection status and frequency

```
IBM_Storwize:ITSO-V7k:superuser>lssystem
statistics_status on
statistics_frequency 2
-- The output has been shortened for easier reading. --
```

It is not possible to stop statistics collection with the command **stopstats** starting with V8.1.

> **Collection intervals:** Although more frequent collection intervals provide a more detailed view of what happens within IBM Spectrum Virtualize and Storwize V7000, they shorten the amount of time that the historical data is available on the 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 at the same time. It is important to collect all files from all nodes for a complete analysis. Tools, such as IBM Spectrum Control, 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 is in 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 statistic
- `Nd_stats_`<*node_id*>`_`<*date*>`_`<*time*> for drives statistic

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_161024_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_Storwize:ITSO-V7k:superuser>lsdumps -prefix /dumps/iostats
id filename
0  Nd_stats_7822DFF-2_181101_173808
1  Nn_stats_7822DFF-2_181101_173808
2  Nv_stats_7822DFF-2_181101_173808
3  Nm_stats_7822DFF-2_181101_173808
4  Nm_stats_7822DFF-2_181101_175308
5  Nv_stats_7822DFF-2_181101_175308
6  Nd_stats_7822DFF-2_181101_175308
7  Nn_stats_7822DFF-2_181101_175308
...
60 Nm_stats_7822DFF-2_181101_212314
61 Nn_stats_7822DFF-2_181101_212314
62 Nd_stats_7822DFF-2_181101_212314
63 Nv_stats_7822DFF-2_181101_212314
```

Tip: The performance statistics files can be copied from the Storwize V7000 nodes to a local drive on your workstation by using `pscp.exe` (included with PuTTY) from an MS-DOS command line, as shown in this example:

```
C:\Program Files\PuTTY>pscp -unsafe -load ITSO-V7K
superuser@192.168.100.100:/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 this website.

Real-time performance monitoring

IBM Storwize V7000 supports real-time performance monitoring. Real-time performance statistics provide short-term status information for the Storwize V7000. The statistics are shown as graphs in the management GUI, or can be viewed from the 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 I/Os 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 through 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_Storwize:ITSO-V7k:superuser>lsnodecanisterstats
node_id node_name stat_name          stat_current stat_peak stat_peak_time
1       node1     compression_cpu_pc 14           14        181101214224
1       node1     cpu_pc             2            2         181101214224
1       node1     fc_mb              0            0         181101214224
1       node1     fc_io              5            5         181101214224
1       node1     sas_mb             122          158       181101213804
1       node1     sas_io             485          627       181101213804
1       node1     iscsi_mb           0            0         181101214224
1       node1     iscsi_io           0            0         181101214224
1       node1     write_cache_pc     13           13        181101214224
1       node1     total_cache_pc     79           80        181101213959
1       node1     vdisk_mb           0            0         181101214224
1       node1     vdisk_io           0            0         181101214224
1       node1     vdisk_ms           0            0         181101214224
1       node1     mdisk_mb           52           55        181101213934
1       node1     mdisk_io           205          217       181101213934
1       node1     mdisk_ms           9            10        181101214219
1       node1     drive_mb           123          158       181101213804
1       node1     drive_io           485          627       181101213804
...
2       node2     mdisk_w_ms         8            11        181101214131
2       node2     drive_r_mb         41           50        181101214041
2       node2     drive_r_io         166          242       181101214041
2       node2     drive_r_ms         8            14        181101214041
2       node2     drive_w_mb         254          271       181101214121
2       node2     drive_w_io         1011         1071      181101214121
2       node2     drive_w_ms         14           17        181101213851
2       node2     iplink_mb          0            0         181101214226
2       node2     iplink_io          0            0         181101214226
2       node2     iplink_comp_mb     0            0         181101214226
2       node2     cloud_up_mb        0            0         181101214226
2       node2     cloud_up_ms        0            0         181101214226
```

```
2       node2    cloud_down_mb      0            0            181101214226
2       node2    cloud_down_ms      0            0            181101214226
2       node2    iser_mb            0            0            181101214226
2       node2    iser_io            0            0            181101214226
```

Example A-4 on page 744 shows statistics for the two nodes members of system `ITSO-V7k`. 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 **`lsnodecanisterstats`** command can also be used with a node canister name or ID as an argument. For example, you can enter the command **`lsnodecanisterstats node1`** to display the statistics of node name `node1` only.

The **`lssystemstats`** command lists the same set of statistics that is listed with the **`lsnodecanisterstats`** command, but representing 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_Storwize:ITSO-V7k:superuser>lssystemstats
stat_name          stat_current stat_peak stat_peak_time
compression_cpu_pc 14           14        181101215930
cpu_pc             2            2         181101215930
fc_mb              0            0         181101215930
fc_io              10           10        181101215930
sas_mb             726          778       181101215445
sas_io             2879         3080      181101215845
iscsi_mb           0            0         181101215930
iscsi_io           0            0         181101215930
write_cache_pc     13           14        181101215745
total_cache_pc     79           80        181101215925
vdisk_mb           0            0         181101215930
vdisk_io           0            0         181101215930
vdisk_ms           0            0         181101215930
mdisk_mb           1            2         181101215920
mdisk_io           2            4         181101215900
mdisk_ms           6            7         181101215720
drive_mb           727          778       181101215445
drive_io           2877         3075      181101215445
drive_ms           11           13        181101215855
vdisk_r_mb         0            0         181101215930
vdisk_r_io         0            0         181101215930
vdisk_r_ms         0            0         181101215930
vdisk_w_mb         0            0         181101215930
```

```
vdisk_w_io        0          0        181101215930
vdisk_w_ms        0          0        181101215930
mdisk_r_mb        0          0        181101215930
mdisk_r_io        0          0        181101215930
mdisk_r_ms        0          0        181101215930
mdisk_w_mb        1          2        181101215920
mdisk_w_io        2          4        181101215900
mdisk_w_ms        6          7        181101215720
drive_r_mb        18         34       181101215825
drive_r_io        77         140      181101215525
drive_r_ms        4          13       181101215545
drive_w_mb        708        752      181101215510
drive_w_io        2800       2971     181101215510
drive_w_ms        11         13       181101215855
power_w           374        384      181101215555
temp_c            24         24       181101215930
temp_f            75         75       181101215930
iplink_mb         0          0        181101215930
iplink_io         0          0        181101215930
iplink_comp_mb    0          0        181101215930
cloud_up_mb       0          0        181101215930
cloud_up_ms       0          0        181101215930
cloud_down_mb     0          0        181101215930
cloud_down_ms     0          0        181101215930
iser_mb           0          0        181101215930
iser_io           0          0        181101215930
```

Table A-1 gives the description of the different counters that are presented by the **lssystemstats** and **lsnodecanisterstats** commands.

Table A-1 List of counters in lssystemstats and lsnodecanisterstats

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 traffic on the system. This value includes host I/O and any bandwidth that is used for communication within the system.
fc_io	Displays the total I/O operations that are transferred per second for Fibre Channel 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 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 iSCSI traffic on the system.

Value	Description
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 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.
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.

Value	Description
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 Internet Protocol (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 internet Small Computer System Interface (iSCSI) host I/O operations.
cloud_up_mb	Displays the average number of 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.
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 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 gives you performance at a glance by displaying some information about the system. You can see the entire cluster (the system) performance by selecting the information between 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 and Figure A-2.

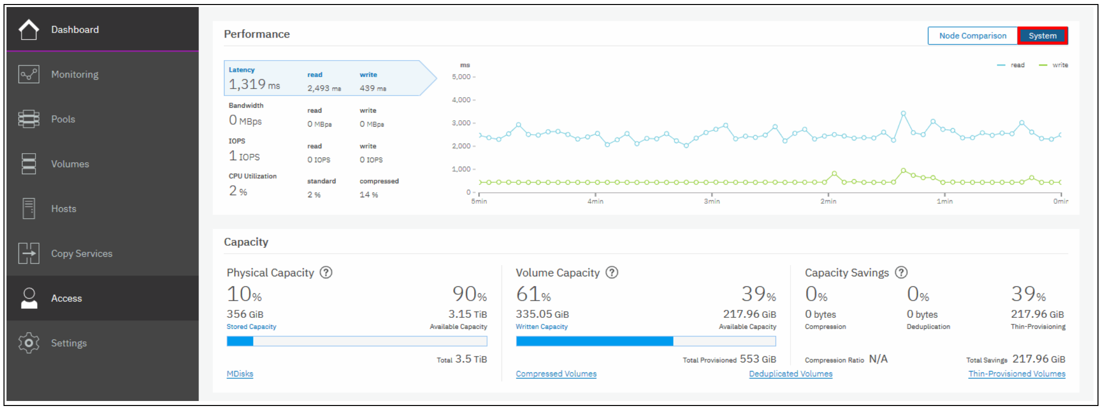


Figure A-1 IBM Spectrum Virtualize Dashboard displaying System performance overview

Figure A-2 shows the display after switching the button.

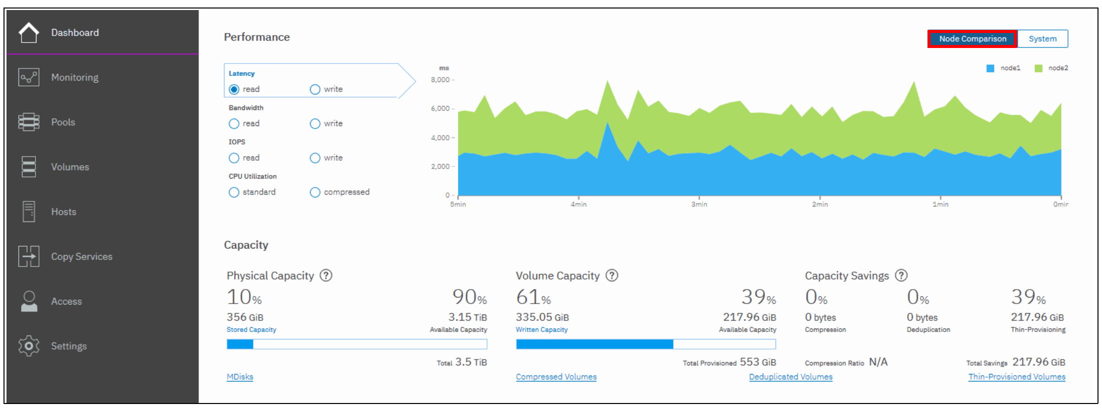


Figure A-2 IBM Spectrum Virtualize Dashboard displaying Nodes performance overview

You can also use real-time statistics to monitor CPU utilization, volume, interface, and 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. Click **Monitoring** → **Performance** (as shown in Figure A-3) to open the Performance Monitoring window.

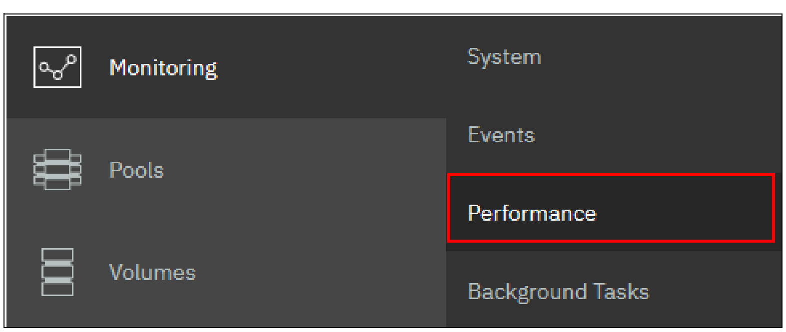


Figure A-3 Selecting performance pane in the monitoring menu

As shown in Figure A-4 on page 751, the Performance monitoring pane is divided into the following sections that provide utilization views for the following resources:

- CPU Utilization: The CPU utilization graph shows the current percentage of CPU usage and peaks in utilization. It can also display compression CPU usage for systems with compressed volumes.
- Volumes: Shows four metrics on the overall volume utilization graphics:
  - Read
  - Write
  - Read latency
  - Write latency
- Interfaces: The Interfaces graph displays data points for FC, iSCSI, serial-attached SCSI (SAS), and IP Remote Copy interfaces. You can use this information to help determine connectivity issues that might affect performance:
  - Fibre Channel
  - 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 also always visible in the bottom of the IBM Spectrum Virtualize window, as shown in Figure A-4 on page 751.

Note: The indicated values in the graphics are averaged on a 1 second based sample.

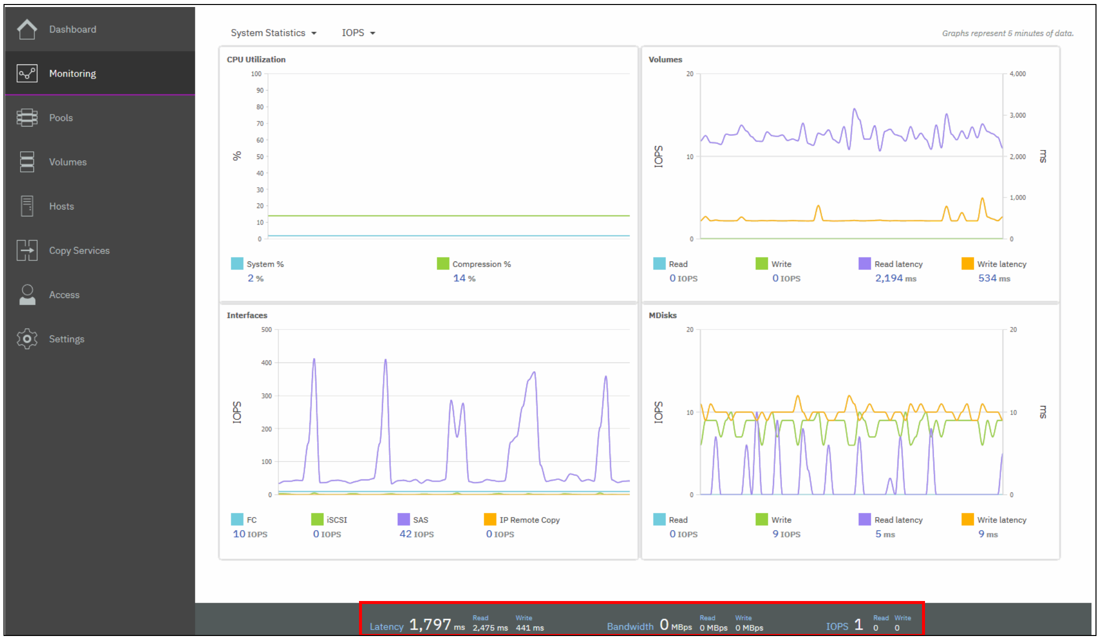


Figure A-4 IBM Spectrum Virtualize performance window

You can also select to view performance statistics for each of the available nodes of the system, as shown in Figure A-5.

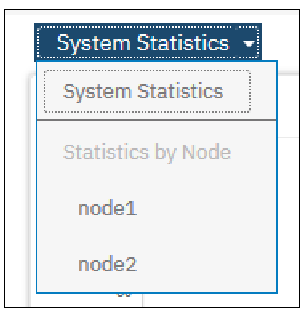


Figure A-5 View statistics per node or for the entire system

You can also change the metric between MBps or IOPS, as shown in Figure A-6.

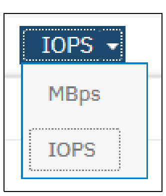


Figure A-6 View performance metrics by MBps or IOps

On any of these views, you can select any point by using your cursor to know 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.

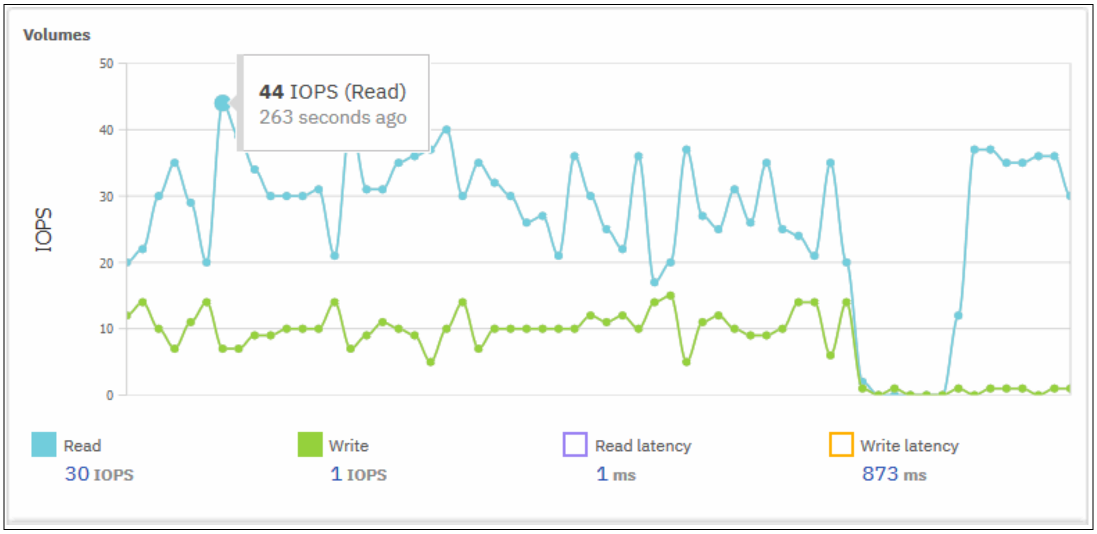


Figure A-7 Viewing performance with details

For each of the resources, various metrics are available and you can select which 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.

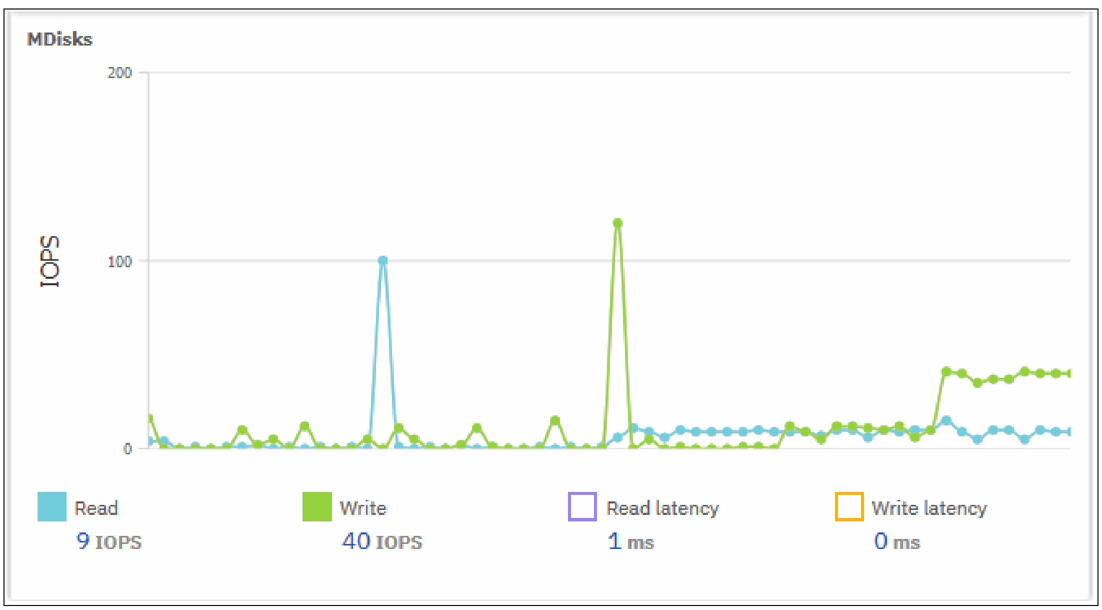


Figure A-8 Displaying performance counters

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 Storwize V7000 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 the use of IBM Spectrum Control to monitor your storage subsystem, see this website.

As an alternative to IBM Spectrum Control, a cloud-based tool 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 this website.

B

CLI setup

This appendix describes access configuration to the command-line interface (CLI) by using the local SSH authentication method.

CLI setup

The IBM Spectrum Virtualize system features a powerful CLI that offers more options and flexibility as compared to the graphical user interface (GUI). This appendix describes how to configure a management system by using Secure Shell (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 Knowledge Center.

> **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 user name and password. CLI uses SSH to connect from a host to the IBM Spectrum Virtualize system. A private and a public key pair or user name and password is necessary.

The use of SSH keys with a passphrase is more secure than a login with user name 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 the use of 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 Storwize V7000 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 Storwize V7000 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 this website.

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 B-1.

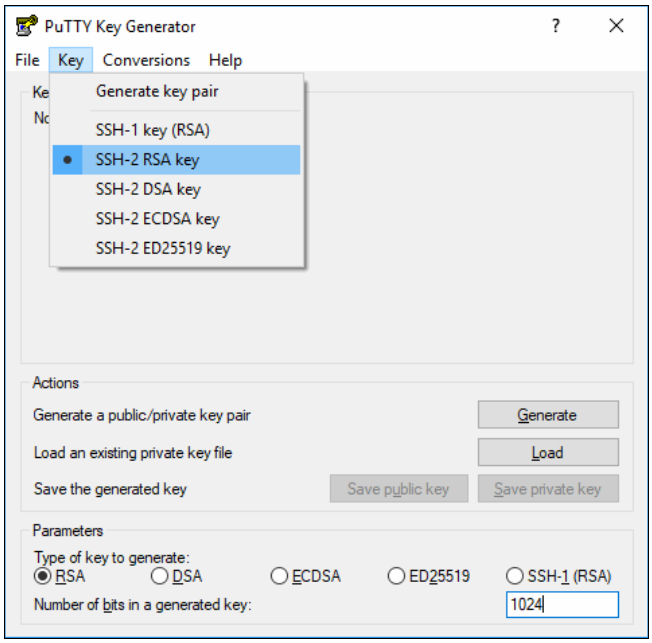


Figure B-1 PuTTY key generator

   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 B-2 on page 758).

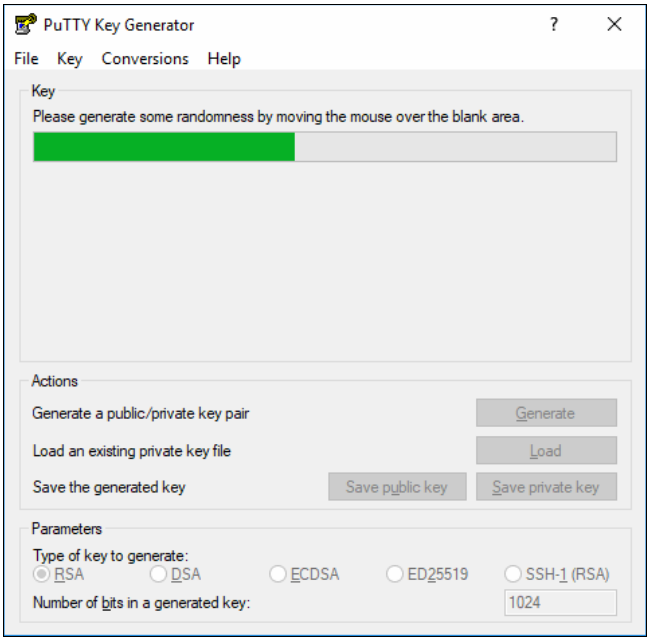


Figure B-2 Generate keys

To generate keys: The blank area that is indicated by the message is the large blank rectangle on the GUI inside the section of the GUI labeled Key. Continue to move the mouse pointer over the blank area until the progress bar reaches the far right. This action generates random characters to create a unique key pair.

3. After the keys are generated, save them for later use. Click **Save public key**, as shown in Figure B-3.

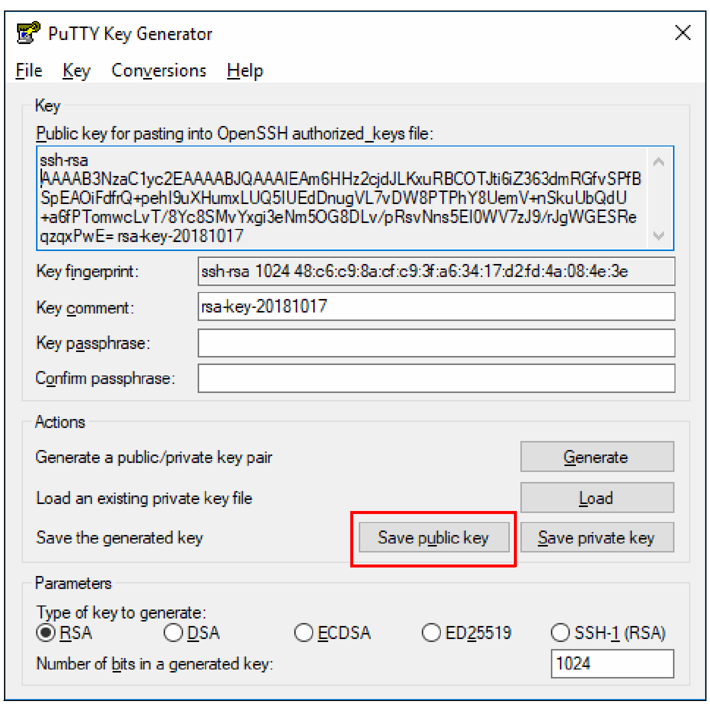


Figure B-3 Save public key

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 B-4). Click **Yes** to save the private key without a passphrase.

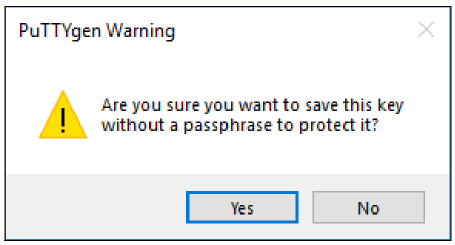


Figure B-4 Confirm the security warning

> **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**.

   > **Key generator:** The PuTTY key generator saves the PuTTY private key (PPK) with the `.ppk` extension.

7. Close the PuTTY key generator.

Uploading the SSH public key to the IBM Storwize V7000

After you create your SSH key pair, upload your SSH public key onto the IBM Storwize V7000. Complete the following steps:

1. Open the user section in the GUI, as shown in Figure B-5.

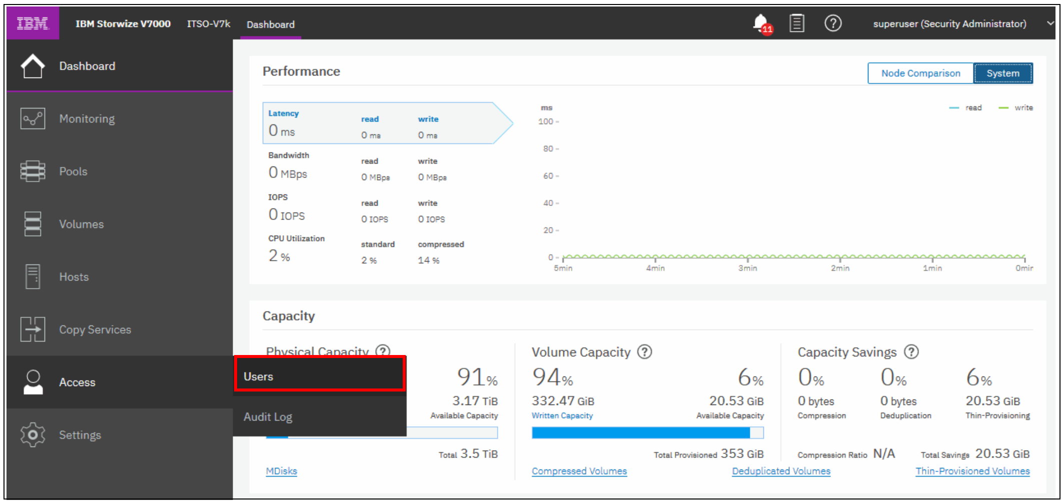


Figure B-5 Open user section

2. Right-click the user name for which you want to upload the key and click **Properties** (see Figure B-6).

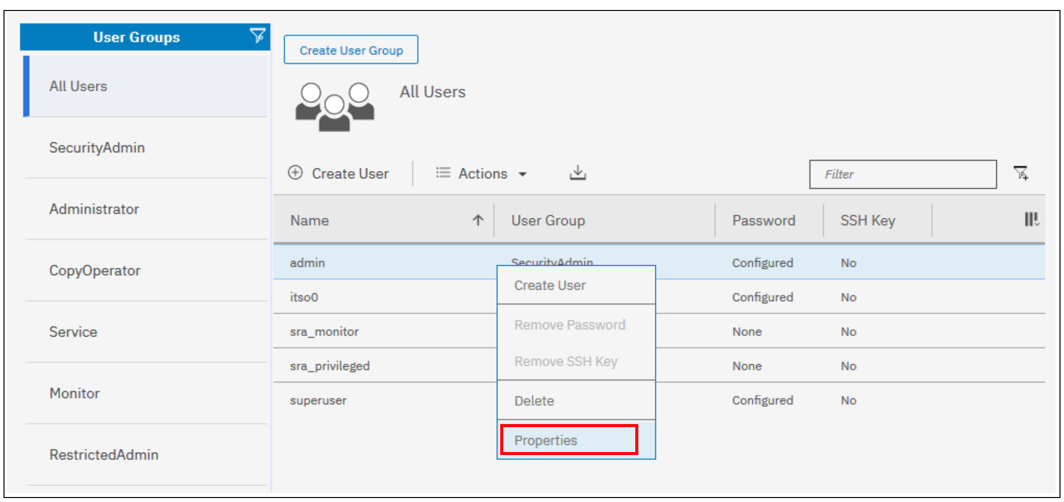


Figure B-6 User properties

3. To upload the public key, click **Browse**, open the folder where you stored the public SSH key, and select the key (see Figure B-7).

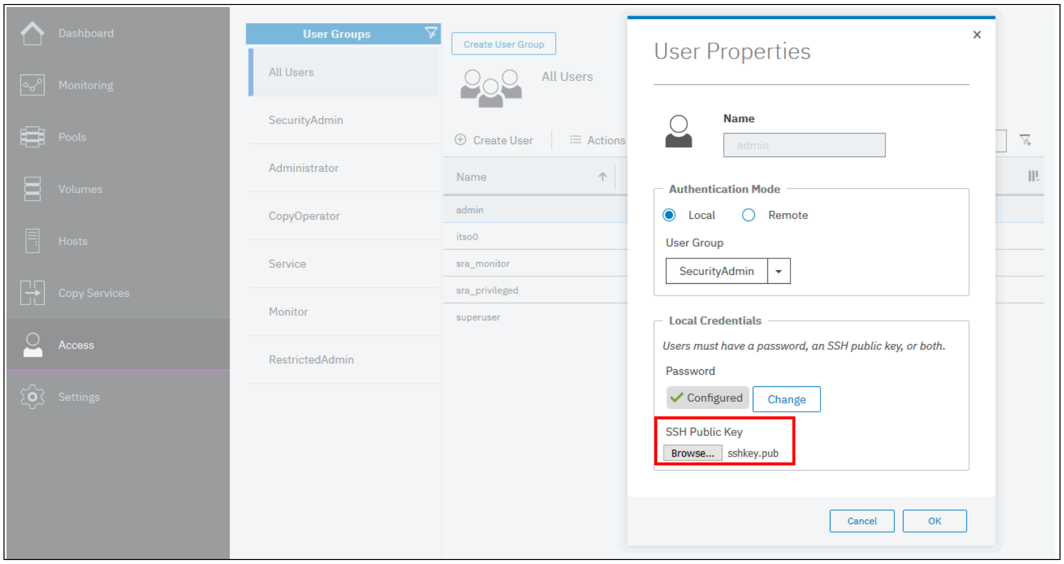


Figure B-7 Select public key

4. Click **OK** and the key is uploaded.
5. Check in the GUI to ensure that the SSH key is imported successfully (see Figure B-8).

Name	User Group	Password	SSH Key
admin	SecurityAdmin	Configured	Yes
itso0	SecurityAdmin	Configured	No
sra_monitor	Monitor	None	No
sra_privileged	RestrictedAdmin	None	No
superuser	SecurityAdmin	Configured	No

Figure B-8 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 B-9).

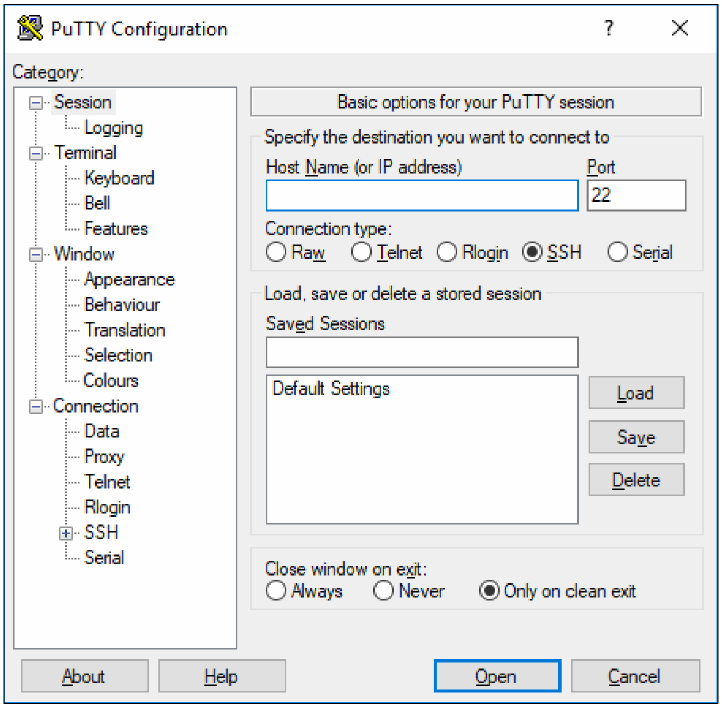


Figure B-9 PuTTY

2. In the right pane, select **SSH** as the connection type. In the Close window on exit section, select **Only on clean exit** (see Figure B-9), which ensures that if any connection errors occur, they are displayed on the user's window.

3. In the Category pane, on the left side of the PuTTY Configuration window, click **Connection** → **Data**, as shown on Figure B-10 on page 763. In the Auto-login username field, enter the 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 B-6 on page 760.

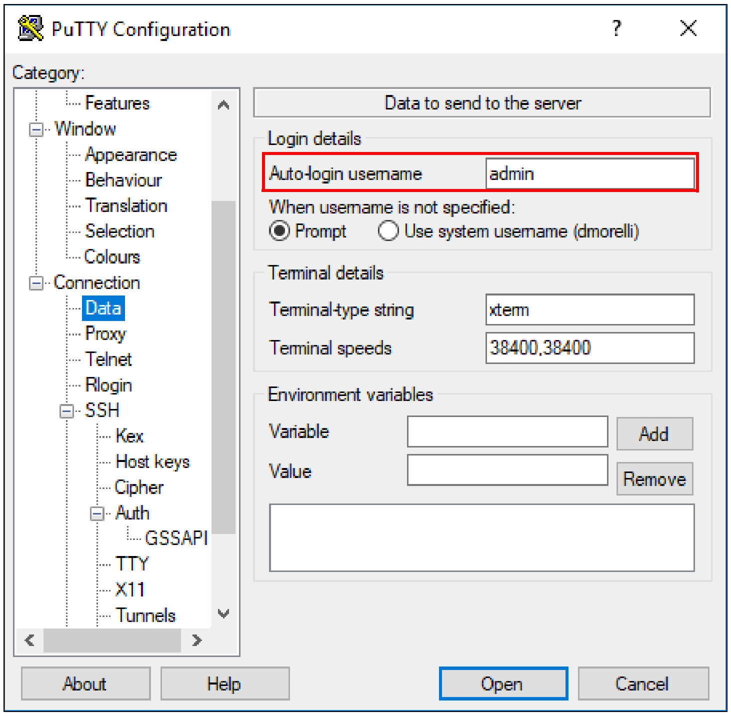


Figure B-10 PuTTY Auto-login user name

4. In the Category pane, on the left side of the PuTTY Configuration window (see Figure B-11), click **Connection** → **SSH** to open the PuTTY SSH Configuration window. In the SSH protocol version section, select **2**.

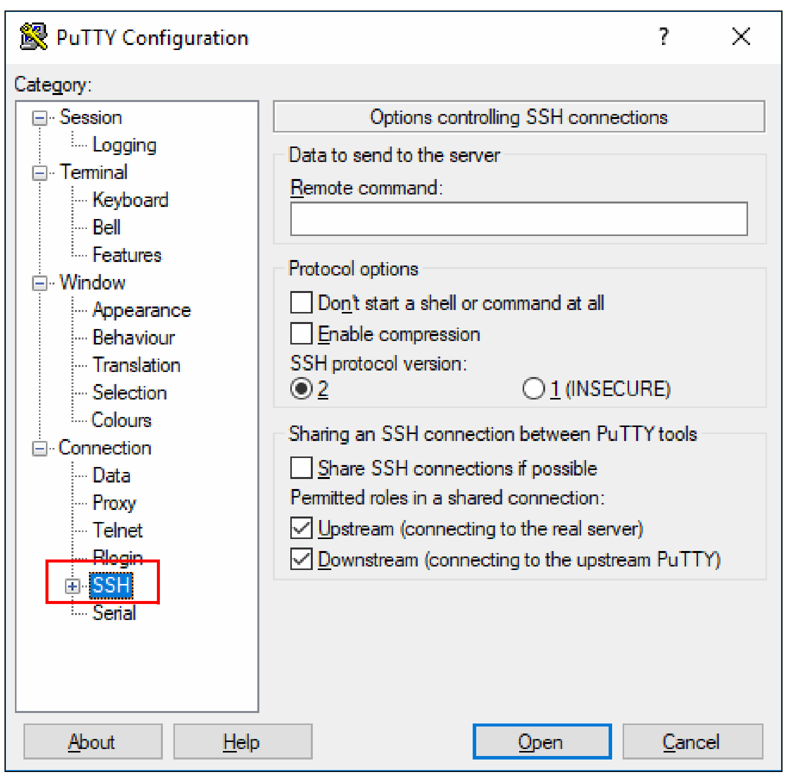


Figure B-11 SSH protocol version 2

5. In the Category pane on the left, click **Connection** → **SSH** → **Auth**. More options are displayed for controlling SSH authentication.
6. In the Private key file for authentication field in Figure B-12, 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:\Keys\sshkey.ppk` is used).

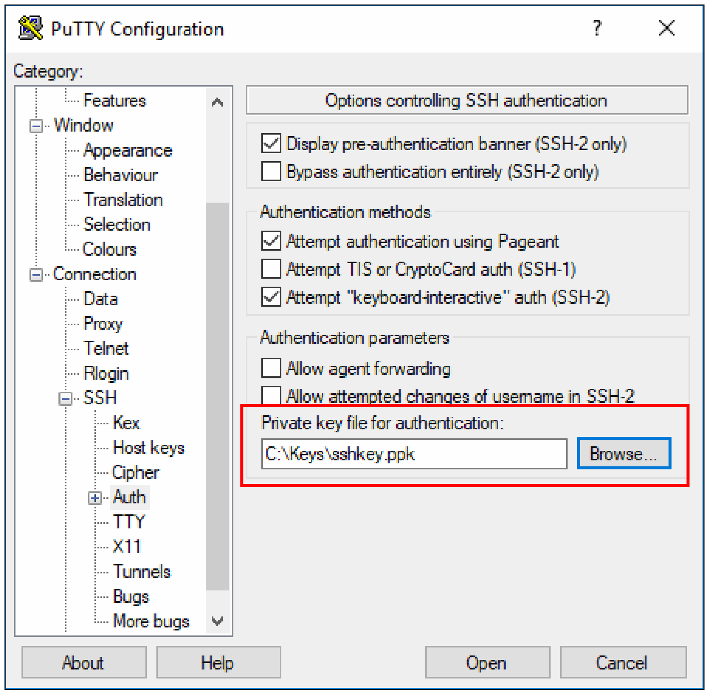


Figure B-12 SSH authentication

7. In the Category pane, click **Session** to return to the Basic options for your PuTTY session view.
8. Enter the following information in the fields in the right pane (see Figure B-13 on page 765):
   – Host Name (or IP address): Specify the host name or system IP address of the IBM Spectrum Virtualize system.
   – Saved Sessions: Enter a session name.
9. Click **Save** to save the new session (Figure B-13 on page 765).

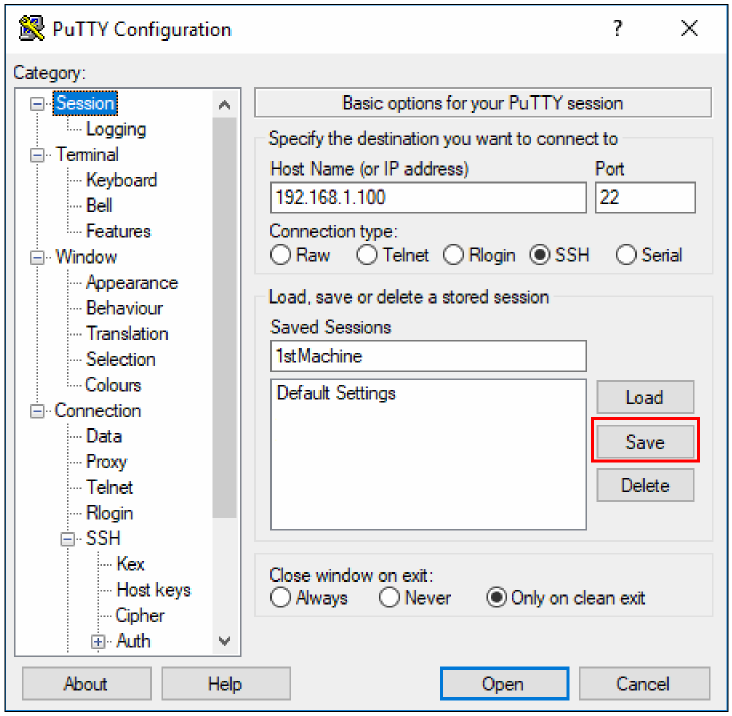


Figure B-13 Session information

10.Select the new session and click **Open** to connect to the IBM Spectrum Virtualize system, as shown in Figure B-14.

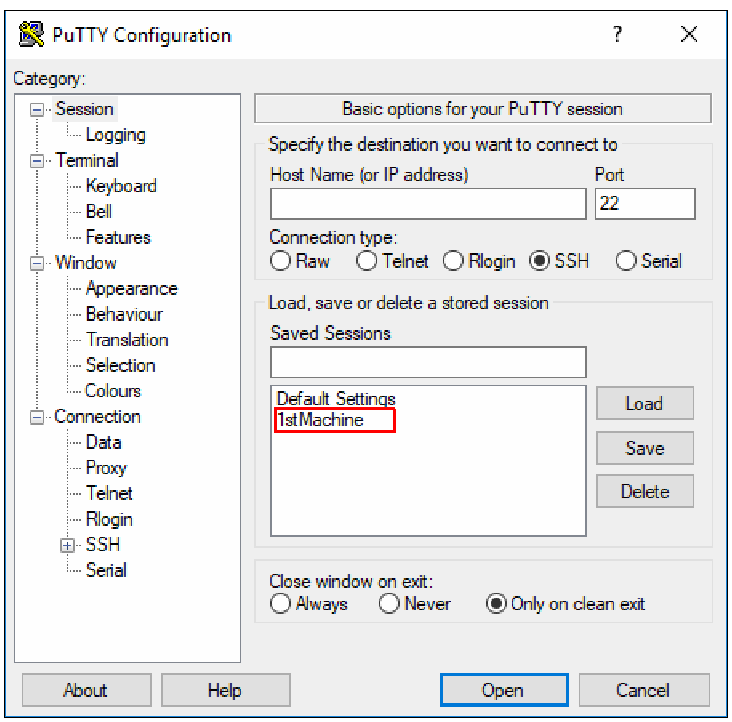


Figure B-14 Connect to system

11. If a PuTTY Security Alert opens as shown in Figure B-15, confirm it by clicking **Yes**.

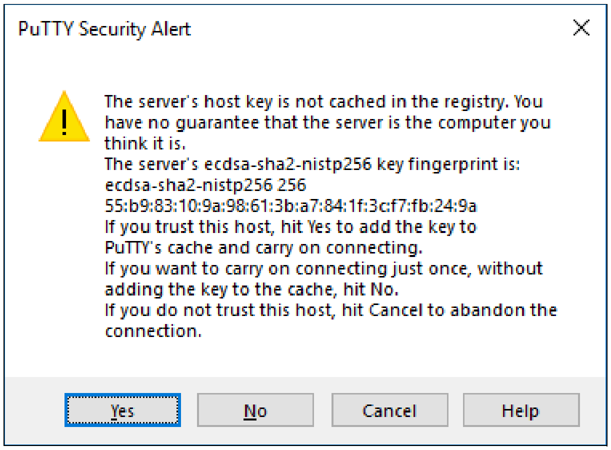


Figure B-15 Confirm Security Alert

12. As shown in Figure B-16, PuTTY now connects to the system automatically by using the user ID that was specified earlier, without prompting for password.

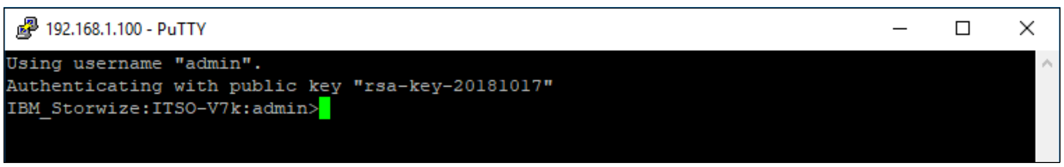


Figure B-16 PuTTY login

The CLI is now configured for IBM Spectrum Virtualize system administration.

Basic setup on an UNIX or Linux host

OpenSSH client is the most common tool that is used on Linux or UNIX operating systems. It is installed by default on most of these types of operating systems. If it is not installed on your system, OpenSSH can be obtained at this website.

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 a private key to connect to IBM Spectrum Virtualize system without entering the user password, run the **ssh-keygen** tool, as shown in Example B-1.

Example B-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]----+
|           .+=BO*|
|           + oB*+|
|          . oo+o |
|         . . .  E|
|        S   .   o|
|               . |
|                 |
|                 |
|                 |
+-----------------+
#
```

In `ssh-keygen`, the parameter **`-t`** refers to the type of SSH key (RSA in Example B-1 on page 766) and **`-b`** is the size of SSH key in bits (Example B-1 on page 766, 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 B-1 on page 766, path is `/.ssh/`, name of private key is `sshkey` and name of public key is `sshkey.pub`.

> **Note:** The use of 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 Storwize V7000

How to upload the new SSH public key to IBM Spectrum Virtualize by using the GUI was described in "Uploading the SSH public key to the IBM Storwize V7000" on page 760.

Complete the following steps to upload the public key by using CLI:

1. On the SSH client (for example, AIX or Linux host), use **`scp`** to copy the public key to IBM Storwize V7000. 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 B-2 shows the command to copy the new generated public key to the IBM Spectrum Virtualize system.

   Example B-2 SSH public key copy to IBM

   ```
   # scp /.ssh/sshkey.pub admin@192.168.1.100:/tmp/
   Password:
   sshkey.pub
   100%  241     0.2KB/s   00:00
   #
   ```

2. Log in to IBM Storwize V7000 by using SSH and run the **chuser** command (as shown in Example B-3) to import the public SSH key to a user.

Example B-3 Importing SSH public key to a user

```
IBM_Storwize:ITSO-V7k:admin>chuser -keyfile /tmp/sshkey.pub admin
IBM_Storwize:ITSO-V7k:admin>lsuser admin
id 4
name admin
password yes
ssh_key yes
remote no
usergrp_id 1
usergrp_name Administrator
IBM_Storwize:ITSO-V7k:admin>
```

When running the **lsuser** command as shown in Example B-3, it is indicated that a user has a configured SSH key in the field `ssh_key`.

Connecting to 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 using the **ssh** command with the following options:

```
ssh -i <SSH_private_key> <user>@<IP_address_or_hostname>
```

Example B-4 shows the SSH command that is running from an AIX server and connecting to IBM Storwize V7000 with an SSH private key and no password prompt.

Example B-4 Connecting to IBM Storwize V7000 with SSH private key

```
# ssh -i /.ssh/sshkey admin@192.168.1.100
IBM_Storwize:ITSO-V7k:admin>
```

C

Terminology

This appendix summarizes the IBM Spectrum Virtualize and IBM Storwize V7000 terms that are commonly used in this book.

For more information about the complete set of terms that relate to the IBM Storwize V7000, see the Glossary section of IBM Knowledge Center.

Commonly encountered terms

This book uses the common IBM Spectrum Virtualize and IBM Storwize V7000 terminology listed in this section.

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 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 786.

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 786.

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 780, and "Relationship" on page 784.

Available (usable) capacity

See "Capacity" on page 771.

Back end

See "Front end and back end" on page 777.

Caching I/O Group

The caching 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:

- Raw capacity

  The reported capacity of the drives in the system before formatting or RAID

- Usable capacity

  The amount of capacity after formatting and RAID available for storing data on a system, pool, array, or MDisk. 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 taken up by data in a system, pool, array, or MDisk after data reduction techniques were applied.

- Available capacity

  The amount of usable capacity that is not yet used in a system, pool, array, or MDisk.

- Effective capacity

  The amount of provisioned capacity that can be created in the system or pool without running out of usable capacity given the current data reduction savings being achieved. This capacity equals the physical capacity divided by the data reduction savings percentage.

- Provisioned capacity

  Total capacity of all volumes in a pool or system.

- Written capacity

  The amount of usable capacity that would be used to store written data in a pool or system before data reduction is applied.

- Overhead capacity

  The amount of usable capacity occupied by metadata in a pool or system and other data used for system operation.

- Total capacity savings

  The total amount of usable capacity saved in a pool, system, or volume through thin-provisioning and data reduction techniques. This capacity saved is the difference between the used usable capacity and the provisioned capacity.

- Data reduction

  The techniques used to reduce the size of data including deduplication and compression.

- Data reduction savings

  The total amount of usable capacity saved in a pool, system, or volume through the application of a compression or deduplication algorithm on the written data. This capacity saved is the difference between the written capacity and the used capacity

- Thin provisioning savings

  The total amount of usable capacity saved in a pool, system, or volume by using usable capacity when needed as a result of write operations. The capacity saved is the difference between the provisioned capacity minus the written capacity.

- Over provisioned

  A storage system or pool where there is more provisioned capacity than there is usable capacity.

- Over provisioned ratio

  The ratio of provisioned capacity to usable capacity in the pool or system.
- Provisioning limit - maximum provisioned capacity - over provisioning limit

  In some storage systems, restrictions in the storage hardware or configured by the user that define a limit the maximum provisioned capacity allowed in a pool or system.

Capacity licensing

Capacity licensing is a licensing model that licenses features with a price-per-terabyte model. Licensed features are FlashCopy, Metro Mirror, Global Mirror, and virtualization. See also “FlashCopy” on page 776, “Metro Mirror” on page 780, and “Virtualization” on page 787.

Chain

A set of enclosures that are 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 user name and password.

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), Internet Protocol (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 managed disks (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. Also, see “Parent pool” on page 781.

Cloud Container

Cloud Container is a virtual object that includes all of the elements, components, or data that are common to a specific application or data.

Cloud Service Provider

Cloud Service Provider (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 Cloud Providers such as IBM SoftLayer, Amazon S3, and deployments of OpenStack Swift.

Cloud Tenant

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 (Storwize V7000)

A clustered system, which was known as a cluster, is a group of up to eight IBM Storwize V7000 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 disk. A cold extent also refers to an extent that needs to be migrated onto an HDD if it is on a Flash disk 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 unused real capacity that is maintained. For thin-provisioned volumes that are not configured to automatically expand, 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 a 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 Storwize V7000 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.

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 771.

Data reduction pool

Data Reduction pools are specific types of pools where more control over volumes capacity is given to specific hosts (for example VMware VAAI/VASA/VVOL, Microsoft ODX). These hosts are able to 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 771.

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 Storwize V7000 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 revolutions per minute (RPM) Fibre Channel (FC) or serial-attached SCSI (SAS) disk, Nearline SAS, or Serial Advanced Technology Attachment (SATA), or even Flash Disks. Therefore, a storage tier attribute is assigned to each MDisk, and the default is *generic_hdd*.

Distributed RAID or DRAID

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 hence improves the overall performance of array. See also “Rebuild area” on page 783.

Easy Tier

Easy Tier is a volume performance function within the IBM Storwize family that provides automatic data placement of a volume’s extents in a multitiered storage pool. The pool normally contains a mix of Flash Disks and HDDs. Easy Tier measures host I/O activity on the volume’s extents and migrates hot extents onto the Flash Disks to ensure the maximum performance.

Effective capacity

See “Capacity” on page 771.

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 server

An internal or external system that receives and then serves existing encryption keys or certificates to a storage system.

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

The 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 Storwize V7000. 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 Storwize V7000 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 logical units 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.

Fibre Channel port logins

FC port logins refer to the number of hosts that can see any one V7000 port. The IBM Storwize V7000 has a maximum limit per node port of FC logins that are allowed.

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.

FlashCopy

FlashCopy refers to a point-in-time copy where a virtual copy of a volume is created. The target volume maintains the contents of the volume at the point in time 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.

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 782.

Flash drive

A data storage device that uses solid-state memory to store persistent data.

Flash module

A modular hardware unit that contains flash memory, one or more flash controllers, and associated electronics.

Front end and back end

The IBM Storwize V7000 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 IBM Storwize V7000.

Global Mirror

Global Mirror (GM) is a method of asynchronous replication that maintains data consistency across multiple volumes within or across multiple systems. Global Mirror is generally 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 relationship. A FlashCopy mapping exists between a primary and its change volume, and a secondary and its change volume.

Grain

A grain is the unit of data that is represented by a single bit in a FlashCopy bitmap (64 KiB or 256 KiB) in the IBM Storwize V7000. 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).

Hop

One segment of a transmission path between adjacent nodes in a routed network.

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, PCI Express. Typically it is referred to the Fibre Channel adapters.

Host ID

A host ID is a numeric identifier that is assigned to a group of host FC ports or internet Small Computer System Interface (iSCSI) host names 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.

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 Disk.

Hot Spare node (SVC only)

Hot Spare Node is an online SVC node defined in a cluster, but not in any IO group. During a failure of any of online nodes in any IO group of cluster, it is automatically swapped with this spare node. After the recovery of an original node has finished, the spare node returns to the standby spare status. This feature is not available for IBM Storwize V7000.

HyperSwap

Pertaining to a function that provides continuous, transparent availability against storage errors and site failures, and is based on synchronous replication.

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 780 and “Unmanaged mode” on page 787.

Image volume

An image volume is a volume in which a direct block-for-block translation exists from the MDisk to the volume.

I/O Group

Each pair of IBM Storwize V7000 canisters 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 V7000 canister is associated with exactly one I/O Group. The canister in an I/O Group provide a failover and failback function for each other.

Internal storage

Internal storage refers to an array of MDisks and drives that are held in IBM Storwize V7000 enclosures.

Internet Small Computer System Interface qualified name

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.<yyyy-mm>.<reversed domain name>`. For example, the default for a Storwize V7000 canister can be in the following format:

```
iqn.1986-03.com.ibm:2076.<clustername>.<nodename>
```

Internet storage name service

The internet storage name service (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 Storwize V7000 environment, the number of ISL hops is counted on the shortest route between the pair of canister that are farthest apart. The V7000 supports a maximum of three ISL hops.

Input/output group

A collection of volumes and canister relationships that present a common interface to host systems. Each pair of canister is known as an I/O group.

I/O throttling rate

The maximum rate at which an I/O transaction is accepted for a volume.

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 hard disk drives and tape changers), an iSCSI initiator sends SCSI commands over an IP network.

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.

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 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 (DAP). 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 (LW) gigabit interface converters (GBICs) or Small Form-factor Pluggables (SFPs), 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 so on) that connect the components (nodes, hosts, and switches) of the local cluster together.

Logical unit and logical unit number

The logical unit (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.

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 Storwize V7000. The MDisk is not visible to host systems on the SAN.

Managed disk group (storage pool)

See “Storage pool (managed disk group)” on page 786.

Managed mode

An access mode that enables virtualization functions to be performed. See also “Image mode” on page 778 and “Unmanaged mode” on page 787.

Maximum replication delay

Maximum replication delay is the number of seconds that Metro Mirror or Global Mirror replication can delay a write operation to a volume.

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 770, and “Relationship” on page 784.

Metro Global Mirror

Metro Mirror Global is a cascaded solution where Metro Mirror synchronously copies data to the target site. This Metro Mirror target is the source volume for Global Mirror that asynchronously copies data to a third site. This solution has the potential to provide disaster recovery with no data loss at Global Mirror distances when the intermediate site does not participate in the disaster that occurs at the production site.

Metro Mirror

Metro Mirror (MM) is a method of synchronous replication that maintains data consistency across multiple volumes within the system. Metro Mirror is generally 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 Storwize V7000 as `copy 0` and the secondary copy is known within the IBM Storwize V7000 as `copy 1`.

Node canister

A node canister is a hardware unit that includes the node hardware, fabric and service interfaces, and serial-attached SCSI (SAS) expansion ports. Node canisters are specifically recognized on IBM Storwize products. In SVC, all these components are spread within the whole system chassis, so we usually do not consider node canisters in SVC, but just the node as a whole.

Node rescue

The process by which a node with no valid software is installed on its hard disk drive can copy software from another node connected to the same Fibre Channel fabric.

NPIV

N_Port ID Virtualization (NPIV) is a Fibre Channel feature whereby multiple Fibre Channel node port (N_Port) IDs can share a single physical N_Port.

Object Storage

Object Storage is a general term that refers to the entity in which Cloud Object Storage organizes, manages, and stores with units of storage or just *objects*.

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.

Over provisioned

See “Capacity” on page 771.

Over provisioned ratio

See “Capacity” on page 771.

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.

Partnership

In Metro Mirror or Global Mirror 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.

Point-in-time copy

A point-in-time copy is an instantaneous copy that the FlashCopy service makes of the source volume. See also “FlashCopy service” on page 777.

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 Metro Mirror or Global Mirror relationship, the target of write operations that are issued by the host application.

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 771.

Public fabric

Configure one SAN per fabric so that it is dedicated for host attachment, storage system attachment, and remote copy operations. This SAN is referred to as a public SAN. You can configure the public SAN to allow Storwize V7000 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.

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.

RACE engine

The RACE engine compresses data on volumes in real time with minimal effect on performance. See “Compression” on page 773 or “Real-time Compression” on page 783.

Raw capacity

See “Capacity” on page 771.

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 771.

Real-time Compression

Real-time Compression is an IBM integrated software function for storage space efficiency. The RACE engine compresses data on volumes in real time with minimal effect on performance.

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 TRAID - Traditional RAID.

RAID 0

RAID 0 is a data striping technique that is used across an array and no data protection is provided.

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

RAID 10 is a combination of a RAID 0 stripe that is mirrored (RAID 1). Therefore, two identical copies of striped data exist, with no parity.

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 of the array's virtual disks in the presence of two concurrent disk failures.

Read intensive drives

The read intensive flash drives that are available on IBM Storwize V7000 Gen2, IBM Storwize V5000 Gen2, and IBM SAN Volume Controller 2145-DH8, SV1, and 24F enclosures are one Drive Write Per Day (DWPD) Read Intensive drives.

Rebuild area

Reserved capacity that is distributed across all drives in a redundant array of drives. 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 RAID or DRAID" on page 775.

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 using the SCSI unmap command, or when a volume is deleted.

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.

Relationship

In Metro Mirror or Global Mirror, 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.

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 fabric

The remote fabric is composed of SAN components (switches, cables, and so on) 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.

Remote Support Server and Client

Remote Support Client is a software toolkit that is in IBM Storwize V7000 and opens a secured tunnel to the Remote Support Server. Remote Support Server is in the IBM network and collects key health check and troubleshooting information that is required by IBM support personnel.

SAN Volume Controller

The 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. The IBM Spectrum Virtualize is a software engine of SVC (and Storwize family) that performs block-level virtualization of disk storage.

Secondary volume

Pertinent to remote copy, the volume in a relationship that contains a copy of data written by the host application to the primary volume.

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 passed between the web server and browsers remain private. To be able to create an SSL connection, a web server requires an SSL Certificate.

Security Key Lifecycle Manager

Security Key Lifecycle Manager (SKLM) centralizes, simplifies, and automates the encryption key management process to help minimize risk and reduce operational costs of encryption key management.

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 American National Standard Institute standard called SAS. In the business enterprise, SAS is useful for access to mass storage devices, particularly external hard disk drives.

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 the request for change (RFC) 2608.

Small Computer System Interface (SCSI)

Small Computer System Interface (SCSI) is an ANSI-standard electronic interface with which personal computers 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 point-in-time view of a volume.

Solid-state disk

A solid-state disk (SSD) or Flash Disk 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 Storwize V7000 as a disk tier `generic_ssd`.

Space efficient

See “Thin provisioning” on page 786.

Spare

An extra storage component, such as a drive or tape, that is predesignated for use as a replacement for a failed component.

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

For more information about a space-efficient volume, see “Thin-provisioned volume” on page 786.

Stand-alone relationship

In FlashCopy, Metro Mirror, and Global Mirror, 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 or SAN

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 pool (managed disk 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 and 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 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 known as extents. These extents are then concatenated, by using various policies, to make volumes. See also "Asymmetric virtualization" on page 770.

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 770.

Tie-break

In a case of a cluster split in two groups of nodes, tie-break is a role of a quorum device used to decide which group continues to operate as the system, handling all I/O requests.

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.

Thin provisioning savings

See "Capacity" on page 771.

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.

Transparent Cloud Tiering

Transparent Cloud Tiering is a separately installable feature of IBM Spectrum Scale™ that provides a native cloud storage tier.

Total capacity savings

See “Capacity” on page 771.

T10 DIF

T10 DIF is a *Data Integrity Field* (DIF) extension to SCSI to enable end-to-end protection of data from host application to physical media.

Unique identifier

A unique identifier (UID) is an identifier that is assigned to storage-system logical units when they are created. It is used to identify the logical unit regardless of the logical unit number (LUN), the status of the logical unit, or whether alternative paths exist to the same device. Typically, a UID is used only once.

Unmanaged 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 778 and “Managed mode” on page 780.

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 772.

Virtualized storage

Virtualized storage is physical storage that has virtualization techniques 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 defined within the storage area network (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 built from physical disk drives attached directly into the ESXi hosts. This solution is not considered for the scope of our publication.

Vital product data

Vital product data (VPD or VDP) is information that uniquely defines system, hardware, software, and microcode elements of a processing system.

Volume

A volume is an IBM Storwize V7000 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 remote-copy 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.

Write-through mode

Write-through mode is a process in which data is written to a storage device at the same time that the data is cached.

Written capacity

See “Capacity” on page 771.

Related publications

The publications that are listed in this section are considered particularly suitable for a more detailed discussion of the topics that are covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide more information about the topic in this document (note that some publications referenced in this list might be available in softcopy only):

- *IBM b-type Gen 5 16 Gbps Switches and Network Advisor*, SG24-8186
- *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521
- *Implementing the IBM Storwize V5000 Gen2 (including the Storwize V5010, V5020, and V5030)*, SG24-8162
- *Implementing the IBM System Storage SAN Volume Controller with IBM Spectrum Virtualize V7.8*, SG24-7933

You can search for, view, download, or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

The following IBM Redbooks publication web pages that are related to this book are also useful resources:

- IBM Storage Networking Redbooks:

  http://www.redbooks.ibm.com/Redbooks.nsf/domains/san
- IBM Flash Storage Redbooks:

  http://www.redbooks.ibm.com/Redbooks.nsf/domains/flash
- IBM Software Defined Storage Redbooks:

  http://www.redbooks.ibm.com/Redbooks.nsf/domains/sds
- IBM Disk Storage Redbooks:

  http://www.redbooks.ibm.com/Redbooks.nsf/domains/disk
- IBM Storage Solutions Redbooks:

  http://www.redbooks.ibm.com/Redbooks.nsf/domains/storagesolutions
- IBM Tape storage Redbooks:

  http://www.redbooks.ibm.com/Redbooks.nsf/domains/tape

Other resources

The following publications are also relevant as further information sources:

- *IBM System Storage Master Console: Installation and User's Guide*, GC30-4090
- *IBM System Storage Open Software Family SAN Volume Controller: CIM Agent Developers Reference*, SC26-7545
- *IBM System Storage Open Software Family SAN Volume Controller: Command-Line Interface User's Guide*, SC26-7544
- *IBM System Storage Open Software Family SAN Volume Controller: Configuration Guide*, SC26-7543
- *IBM System Storage Open Software Family SAN Volume Controller: Host Attachment Guide*, SC26-7563
- *IBM System Storage Open Software Family SAN Volume Controller: Installation Guide*, SC26-7541
- *IBM System Storage Open Software Family SAN Volume Controller: Planning Guide*, GA22-1052
- *IBM System Storage Open Software Family SAN Volume Controller: Service Guide*, SC26-7542
- *IBM System Storage SAN Volume Controller - Software Installation and Configuration Guide,* SC23-6628
- *IBM System Storage SAN Volume Controller V6.2.0 - Software Installation and Configuration Guide,* GC27-2286
- *IBM System Storage SAN Volume Controller 6.2.0 Configuration Limits and Restrictions*, S1003799
- *IBM TotalStorage Multipath Subsystem Device Driver User's Guide*, SC30-4096
- *IBM XIV and SVC Best Practices Implementation Guide*

  http://ibm.co/1bk64gW
- *Considerations and Comparisons between IBM SDD for Linux and DM-MPIO*

  http://ibm.co/1CD1gxG

Referenced websites

These websites are also relevant as further information sources:

- IBM Storage home page

  http://www.ibm.com/systems/storage
- SAN Volume Controller supported platform

  http://ibm.co/1FNjddm
- SAN Volume Controller IBM Knowledge Center

  http://www.ibm.com/support/knowledgecenter/STPVGU/welcome
- Cygwin Linux-like environment for Windows

  http://www.cygwin.com

- Open source site for SSH for Windows and Mac

  http://www.openssh.com/
- Windows Sysinternals home page

  http://www.sysinternals.com
- Download site for Windows PuTTY SSH and Telnet client

  http://www.chiark.greenend.org.uk/~sgtatham/putty

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

F

G

H

I

K

L

T

Redbooks

Implementing the IBM Storwize V7000 with IBM Spectrum Virtualize V8.2.1

SG24-7938-07

ISBN 0738457779

IBM

(1.5” spine)
1.5”<-> 1.998”
789 <->1051 pages

SG24-7938-07

ISBN 0738457779

Printed in U.S.A.