

Ensuring Business Continuity: Policy-Based Replication and Policy-Based High Availability for IBM Storage Virtualize Systems

Bernd Albrecht
Byron Grossnickle
Carsten Larsen
Erwan Auffret
Thomas Vogel
Vasfi Gucer



Storage

Infrastructure Solutions



IBM Redbooks

**Ensuring Business Continuity with Policy-Based
Replication and Policy-Based HA**

July 2024

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

First Edition (July 2024)

This edition applies to IBM Storage Virtualize Version 8.7.

This document was created or updated on July 12, 2024.

© Copyright International Business Machines Corporation 2024. All rights reserved.

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

Contents

Figures	vii
Tables	xi
Examples	xiii
Notices	xv
Trademarks	xvi
Preface	xvii
Authors	xvii
Now you can become a published author, too!	xix
Comments welcome	xix
Stay connected to IBM Redbooks	xix
Chapter 1. Introduction	1
1.1 Recovery Time and Point Objectives (RTO & RPO)	2
1.2 Synchronous, asynchronous and policy-based replication	3
1.2.1 Synchronous replication	4
1.2.2 Asynchronous replication	5
1.2.3 Asynchronous replication with snapshots	5
1.2.4 IBM policy-based replication	6
1.3 Data consistency	8
1.4 Policy-based HA	9
1.5 Summary of storage business continuity strategies	10
1.6 IBM Flash Grid	10
Chapter 2. Policy-based replication	31
2.1 Introduction	32
2.2 Policy-based replication - Asynchronous	32
2.2.1 Asynchronous replication - Journaling	33
2.2.2 Asynchronous replication - Cycling	34
2.2.3 Quality of Service (QOS)	35
2.3 Recovery and testing	36
2.3.1 Enable access	36
2.3.2 Snapshot and thin clone	38
2.3.3 Recovery test	38
2.4 More information	40
Chapter 3. Policy-based high availability	31
3.1 Introduction	32
3.1.1 Driving forces behind the development of a new policy-based HA solution	32
3.1.2 Simplifying storage with IBM Flash Grid and policy-based HA	32
3.2 Policy-based HA concepts	34
3.2.1 Simplified management with storage partitions	34
3.2.2 Draft partition	35
3.2.3 Making partitions highly available	36
3.3 Behavior examples of policy-based HA	37
3.3.1 Comparison of policy-based HA with SVC stretched cluster and HyperSwap	44

Chapter 4. Implementing policy-based replication	47
4.1 Implementing policy-based replication	48
4.1.1 Configuring policy-based replication using GUI	48
4.2 Converting Global Mirror to policy-based replication.	62
Chapter 5. Managing policy-based replication.	73
5.1 Managing partnerships using the GUI.	74
5.1.1 Stopping a partnership	75
5.2 Managing pool links	76
5.3 Managing volume groups using the GUI.	78
5.3.1 Adding volumes to a volume group.	80
5.3.2 Removing volumes from a volume group	81
5.3.3 Taking snapshots of volume groups	82
5.3.4 Restoring a volume group from snapshots	85
5.4 Managing replication policies using the GUI	87
5.5 Checking the RPO and the status of policy-based replication	89
5.5.1 Checking the RPO and status using the management GUI	89
5.5.2 Checking the RPO and status using REST APIs	91
Chapter 6. Implementing policy-based HA	97
6.1 Implementing policy-based high availability	98
6.1.1 Storage partitions	98
6.2 Policy-based HA versus HyperSwap	98
6.2.1 Migrating from HyperSwap to policy-based HA	99
6.3 Configuring policy-based HA.	99
6.4 Migrating storage partitions between systems	113
Chapter 7. Managing policy-based high availability	117
7.1 Evaluate the current status of policy-based HA	118
7.1.1 Evaluate the current environment using GUI	118
7.1.2 Evaluate the current environment using CLI.	120
7.2 Volume management	123
7.2.1 Create a new volume in a partition	123
7.2.2 Delete a volume in a partition	124
7.2.3 Add data volumes to a partition and merge partitions.	125
7.3 Host management.	134
7.3.1 Add hosts to partition	135
7.3.2 Optimize policy-based HA internal data flow: Assign host location.	135
7.3.3 Remove host from partition.	136
7.4 Partition management.	136
7.4.1 Change replication policy	136
7.4.2 Delete partition	136
7.5 Migration options for policy-based HA partitions.	139
7.5.1 Migration of policy-based HA data by temporary HA protection removal	139
7.5.2 Migration of policy-based HA data to a third site by keeping the policy-based HA protection	139
7.5.3 Migration of policy-based HA or policy-based replication data to a third site ...	139
7.6 Snapshots and policy-based HA	140
7.6.1 Key differences from the previous snapshot solutions	140
Chapter 8. Configuring FlashSystem and SVC partnerships over high-speed Ethernet	
141	
8.1 Introduction to replication over high-speed Ethernet.	142
8.2 Short-distance partnership using RDMA.	142

8.3 Setup considerations	142
8.3.1 Initial setup considerations	142
8.3.2 Network requirements	143
8.3.3 Deployment of short-distance partnership using RDMA	144
8.3.4 Configuring a short-distance partnership using RDMA	146
8.3.5 Bandwidth utilization	157
8.3.6 Policy-based replication configuration checklist	157
8.3.7 General guidelines	158
8.3.8 Troubleshooting	158
Related publications	161
IBM Redbooks	161
Online resources	161
Help from IBM	161

Figures

1-1 Non-zero RPO and non-zero RTO	2
1-2 Zero RPO and non-zero RTO	3
1-3 Zero RPO and zero RTO	3
1-4 Synchronous replication and RTT impact	4
1-5 Cycle-based asynchronous replication	6
1-6 Policy-based replication with journaling	8
1-7 Storage partitions	9
1-8 IBM Flash Grid concept	11
2-1 Journaling methodology	33
2-2 Cycling methodology	35
2-3 Enable access 1	37
2-4 Enable access 2	37
2-5 Enable access 3	38
2-6 Recovery test In progress	39
2-7 Recovery volumes online	39
3-1 Storage partition	35
3-2 FlashSystem example - single IO group with two partitions and other local volumes	35
3-3 Highly available storage partition	37
3-4 HA without host locations	38
3-5 HA with host locations	38
3-6 HA with host locations - Storage failure	39
3-7 HA storage partitions - split brain scenario 1	39
3-8 HA storage partitions - split brain scenario 2	40
3-9 HA storage partitions - split brain scenario 3	40
3-10 HA storage partitions - normal running, asymmetric preferences	41
3-11 HA storage partitions - quorum decision per partition	42
3-12 HA disconnected - System 1 cannot get to the quorum	42
3-13 Operation while in fault state	43
3-14 Problem fixed - Resynchronization	43
3-15 Paths return to normal	44
3-16 Management returns to normal	44
4-1 Topology for our policy-based replication setup	48
4-2 Verify software version	49
4-3 Create partnership	50
4-4 Create Partnership window	51
4-5 Partnership ready for policy-based replication	52
4-6 Setup policy-based replication	52
4-7 Set up policy-based replication - Link pools	53
4-8 Link pools on the recovery system	54
4-9 Create replication policy	55
4-10 Create Volume Group	56
4-11 Policy-based replication create summary	56
4-12 Volume Group created including no volumes	57
4-13 Create new volumes within Volume Group	57
4-14 Create Volumes wizard begins	57
4-15 Create four volumes each 50 GB	58
4-16 Four volumes created and added to the Volume group	58
4-17 Initial copy ongoing	59

4-18	Volume group policy status	59
4-19	Enable volume access from recovery system	60
4-20	Replication stopped and access enabled to recovery volumes	60
4-21	Restarting replication in reverse direction	61
4-22	Copy direction reversed - initial copy ongoing	62
4-23	Update the existing partnership	65
4-24	Partnership properties	65
4-25	Verify that the partnership is ready for policy-based replication enabled	66
4-26	Setup policy-based replication wizard	67
4-27	Volumes exist for Remote Copy as well as for policy-based replication	67
4-28	Verify the current state of the consistency group	68
4-29	Stop Remote-Copy consistency group	68
4-30	State of the consistency group to Idling	69
4-31	Delete relationship	69
4-32	Confirm relationship deletion	70
4-33	Delete group	70
4-34	GUI preferences - remove Remote Copy features	71
5-1	Partnerships menu	74
5-2	Fully configures partnership	74
5-3	Supported multiple partnerships	75
5-4	Stopping a partnership	75
5-5	Last recovery point on stopped partnership	76
5-6	Pools management menu	77
5-7	Adding a pool link	78
5-8	Modifying a pool link	78
5-9	Volume Groups menu	79
5-10	Supported multiple replication policies	80
5-11	Unsupported multiple replication policies	80
5-12	Adding volumes to a Volume group	81
5-13	Removing volumes from a volume group	82
5-14	Taking volume groups snapshot	83
5-15	Assigning an internal snapshot policy to a volume group	83
5-16	Selecting a snapshot policy for a volume group	84
5-17	Displaying external Safeguarded backup policy in volume groups	85
5-18	A Volume Group page with external Safeguarded backup policies available	85
5-19	Restoring a volume group snapshot	86
5-20	Replication policies menu	87
5-21	Creating a replication policy	88
5-22	Assigning a replication policy to a volume group	88
5-23	Assigning a replication policy to a volume group	89
5-24	Volume group RPO and status	90
5-25	Volume groups and their RPO	90
5-26	An outside policy RPO alert	91
6-1	The topology of the systems we are configuring	100
6-2	The Setup policy-based replication wizard begins	101
6-3	Download IP quorum	101
6-4	Link pools between systems	102
6-5	Create partition	102
6-6	Create replication policy	103
6-7	Select volume groups	103
6-8	Volume group selected	104
6-9	Summary review	105
6-10	Storage Partition view	105

6-11	Create volumes from within the storage partition	106
6-12	Define volume properties	107
6-13	Create volumes wizard	107
6-14	Volumes in the storage partition of the FS9100 and FS7300	108
6-15	Volumes on the FS7300	108
6-16	Hosts menu in policy-based HA	109
6-17	Select host location	109
6-18	Select host WWNs	110
6-19	Review hosts settings	110
6-20	Host created	111
6-21	Create volume mappings to host	111
6-22	Create mapping wizard	112
6-23	Select volumes to map	112
6-24	Storage Partition overview page	113
7-1	Replication partnership	118
7-2	Replication policy for 2-site high availability (policy-based HA)	119
7-3	Configuration entry point for partitions	119
7-4	Partition configuration: Overview	120
7-5	Delete a policy-based HA volume	124
7-6	Delete a policy-based HA volume and confirm host mapping removal	125
7-7	Original storage partition, replicated to the remote site	126
7-8	Check current policy-based HA volume status	127
7-9	Create new storage partition	128
7-10	Create new storage partition and select volume group	128
7-11	Create new storage partition and select volume group	129
7-12	Create new storage partition and select volume group finish	129
7-13	Check all storage partitions	130
7-14	Partition: Check for details for the new partition	130
7-15	Partition: Select replication policy	131
7-16	Partition: Review and finalize partition changes	131
7-17	Start partition merge process	132
7-18	Set options for partition merge	132
7-19	Review the partition merge settings and start the merge process	133
7-20	Verify the partition after merge	133
7-21	Check for policy-based HA volumes in detail after merge	134
7-22	Remove replication policy	137
7-23	Delete the storage partition	137
8-1	Configuration topology for short-distance partnership using RDMA	144
8-2	Configuration of a short-distance partnership using RDMA over ISL	145
8-3	Configuration of a short-distance partnership using RDMA using direct-attach connections	146
8-4	Creating a port-set	147
8-5	Specify port-set name and type	147
8-6	Listing of portsets	148
8-7	Right click on RDMA port	148
8-8	Add IP address	149
8-9	Enter IP address, subnet mask, VLAN, and gateway	149
8-10	Select a portset	150
8-11	IP address of RDMA port	150
8-12	Creating a partnership	151
8-13	Select a short-distance partnership using RDMA	151
8-14	Test a connection for partnership	152
8-15	Select Portset Link1 and Portset Link2	152

8-16	Partially configured partnership	153
8-17	Configured partnership	153
8-18	Select partnership properties	154
8-19	Partnership properties	154
8-20	mkportset and lspportset command output	155
8-21	lspportethernet command output	155
8-22	mkip and lspip command output	155
8-23	mkippartnership and lspartnership command output	156
8-24	lspartnership command output	156
8-25	sainfo lsnodeipconnectivity command output	156
8-26	Throughput versus time	157
8-27	View the configuration status in GUI	158
8-28	View the configuration status in CLI	159
8-29	Ethernet Connectivity page	159
8-30	Displaying connectivity between nodes attached through Ethernet network	159
8-31	IP addresses configured on the ports	160

Tables

1-1 Business continuity options	10
3-1 Comparing policy-based HA with SVC stretched cluster and HyperSwap	45
5-1 RPO status	90
5-2 Replication status	91

Examples

5-1	Authenticating and getting a token	91
5-2	Requesting volume groups' replication status	92
5-3	Viewing the volume groups and their replication status	92
5-4	Details of the replication status for a given volume group	92
5-5	Requesting RPO alerts list from the event log	93
5-6	Eventlog with exceeded RPO events only	93
5-7	Example of specific event details	94
5-8	Example of details for a specific exceeded RPO alert	94
7-1	Verify remote copy partnerships	120
7-2	List available replication policies.	121
7-3	Replication policy in detail.	121
7-4	List of partition and one partition in detail	121
7-5	Partition: Identification of active management system	122
7-6	Partition: Change active management system	122
7-7	Identify storage location and verify the host location	123
7-8	Create new policy-based HA volume	124
7-9	Assign policy-based HA volume to both hosts	124
7-10	Policy-based HA volume deletion	125
7-11	Identify storage locations - optional needed for host definition	134
7-12	Create new hosts with location settings	135
7-13	Assign volume to the host.	135
7-14	Change host location setting.	135
7-15	Host removal.	136
7-16	Assign a different replication policy to a partition	136
7-17	Remove replication policy	138
7-18	Delete the partition	138

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 <https://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.

Redbooks (logo) ®

FlashCopy®

HyperSwap®

IBM®

IBM FlashSystem®

IBM Spectrum®

Redbooks®

The following terms are trademarks of other companies:

Evolution, are trademarks or registered trademarks of Kenexa, an IBM Company.

ITIL is a Registered Trade Mark of AXELOS Limited.

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

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

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

Preface

In today's digital age, downtime is not an option. Businesses rely on constant access to critical data to maintain productivity and ensure customer satisfaction. IBM® Storage Virtualize offers functionalities to safeguard your data against various threats. Policy-based replication and policy-based HA (policy-based-HA) protect against site failures by automatically failing over to a secondary site, ensuring business continuity.

This IBM Redbooks® delves into the powerful tools of IBM policy-based replication and IBM policy-based high availability, empowering you to create a robust disaster recovery plan that minimizes downtime and maximizes data protection.

Whether you are a seasoned IT professional or just starting to explore business continuity solutions, this book provides a comprehensive guide to navigating these essential technologies and building a resilient IT infrastructure.

Authors

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



Bernd Albrecht is a Storage Advisory Partner Technical Specialist for DACH with over 32 years of technical sales experience at IBM. He brings deep expertise to his role. Previously, he spent 3 years as an OEM technical business development manager for Lenovo storage within IBM Germany. Based in Berlin, Bernd focuses on the IBM FlashSystem Family and IBM SAN Volume Controller products, with experience dating back to their launch in 2003. He's also a published IBM Redbooks author.



Byron Grossnickle is a member of the IBM Advanced Technology Group and serves North America as a Subject Matter Expert on Storage Virtualize and IBM FlashSystem. He helps technical sales teams engineer IBM storage and communicate its benefits. He also trains technical sales teams worldwide. Byron has worked in IBM Storage for 18 years. Prior to that, he engineered storage for the data centers of a large national telecommunications provider. Byron has also worked a number of years in IT in the healthcare industry.



Carsten Larsen is an IBM Certified Senior IT Specialist working for the Technical Services Support organization at IBM Denmark, where he delivers consultancy services to IBM clients within the storage arena. Carsten joined IBM in 2007 when he left HP, where he worked with storage arrays and UNIX for 10 years. While working for IBM, Carsten obtained several Brocade and NetApp certifications. Carsten is the author of several IBM Redbooks publications.



Erwan Auffret has been a Storage Consultant for IBM Technology Expert Labs, based in Montpellier, France since 2016. He was previously an IBM FlashSystem Client Technical Specialist and also worked as an IBM System x servers IT specialist for six years. He was co-author of the *Implementing the IBM System Storage SAN Volume Controller with IBM Spectrum Virtualize V8.1* IBM Redbooks. Erwan has also worked for Amadeus, a global IT solutions provider for airlines as a Product Manager for two years and for ATOS, as a global storage Product Manager. He holds a Master's degree in computer sciences from the EFREI computers and electronics engineering school in Paris, and has been working in the IT industry for 19 years.



Thomas Vogel is a Consulting IT Specialist working for the EMEA Advanced Technical Support organization at IBM Systems Germany. His areas of expertise include solution design, storage virtualization, storage hardware and software, educating the technical sales teams and IBM Business Partners, and designing DR and distributed high availability (HA) solutions. For the last 16 years, he has been designing and selling solutions for IBM Storage Virtualize and FlashSystem, and assisting with customer performance and problem analysis. He holds a degree in electrical engineering and has achieved VMware VCP Certification.



Vasfi Gucer leads projects for the IBM Redbooks team, leveraging his 20+ years of experience in systems management, networking, and software. A prolific writer and global IBM instructor, his focus has shifted to storage and cloud computing in the past eight years. Vasfi holds multiple certifications, including IBM Certified Senior IT Specialist, PMP, ITIL V2 Manager, and ITIL V3 Expert.

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

Elias Luna, Andrew Greenfield

IBM USA

Lucy Harris, Evelyn Perez, Chris Bulmer, Chris Canto, Daniel Dent, Bill Passingham, Nolan Rogers, David Seager, Russell Kinmond, Joanne E Borrett

IBM UK

Abhishek Jaiswal, Aakanksha Mathur, Akshada Thorat, Akash Shah and Santosh Yadav

IBM India

Diana Laura Silva Gallardo

IBM Mexico

Now you can become a published author, too!

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

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

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

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

ibm.com/redbooks

- ▶ Send your comments in an email to:

redbooks@us.ibm.com

- ▶ Mail your comments to:

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

Stay connected to IBM Redbooks

- ▶ Find us on LinkedIn:

<https://www.linkedin.com/groups/2130806>

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

<https://www.redbooks.ibm.com/subscribe>

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

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



Introduction

Business continuity ensures an organization can deliver services even during disruptions. While some applications might tolerate temporary outages, major disasters can cause significant downtime and data loss, leading to immense costs for recovery. Organizations should minimize data loss and downtime to lessen business impact and financial strain.

From a storage perspective, business continuity involves maintaining data consistency and availability for uninterrupted application access. Two key concepts contribute to this: Disaster recovery (DR) and high availability (HA). DR focuses on replicating data to remote locations for recovery, while HA prioritizes continuous data accessibility.

Disasters can range from entire site outages to data corruption or theft. Data protection typically involves local or remote data backups. IBM Storage Virtualize offers functionalities to safeguard your data against various threats. Policy-based replication and policy-based high availability protect against site failures by automatically failing over to a secondary site, ensuring business continuity. While not covered here, Storage Virtualize offers additional features like Snapshots and Safeguarded Snapshots to protect against data corruption or cyberattacks.

This chapter has the following sections:

- ▶ “Recovery Time and Point Objectives (RTO & RPO)” on page 2
- ▶ “Synchronous, asynchronous and policy-based replication” on page 3
- ▶ “Data consistency” on page 8
- ▶ “Policy-based HA” on page 9
- ▶ “Summary of storage business continuity strategies” on page 10
- ▶ “IBM Flash Grid” on page 10

1.1 Recovery Time and Point Objectives (RTO & RPO)

After a disastrous event, the priority is to recover the business-critical applications as quickly as possible and to use the most recent data available.

In a disaster recovery environment, where a production site runs the applications and replicates on a recovery site, depending on the replication mode, the data on the recovery site can be older than the one on production site. The time gap between these two versions represents the amount of data potentially lost in case there is a disaster. It is referred to as the *Recovery Point Objective (RPO)*.

The time needed to recover from the latest available data is the *Recovery Time Objective (RTO)*. It is typically the time needed to reload latest available data and to mount volumes to servers on the recovery site; it corresponds to the application downtime.

When cycle-based asynchronous replication is used, the cycle period will define the recovery point. See Figure 1-1 on page 2.

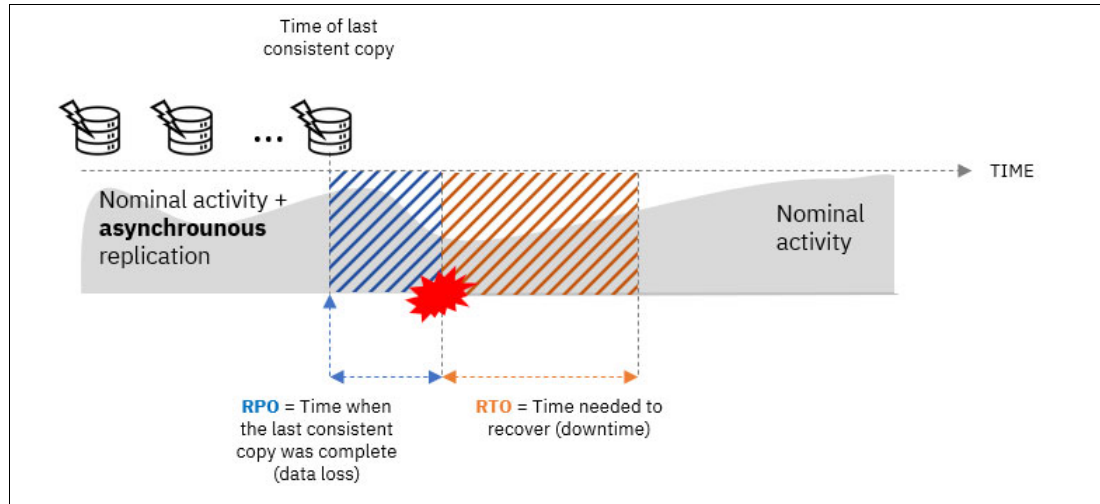


Figure 1-1 Non-zero RPO and non-zero RTO

When synchronous replication is used, the recovery point is reduced to zero because the available version of data on the recovery site is equivalent to the latest on production site. There is no data loss in the event of a disaster. See Figure 1-2.

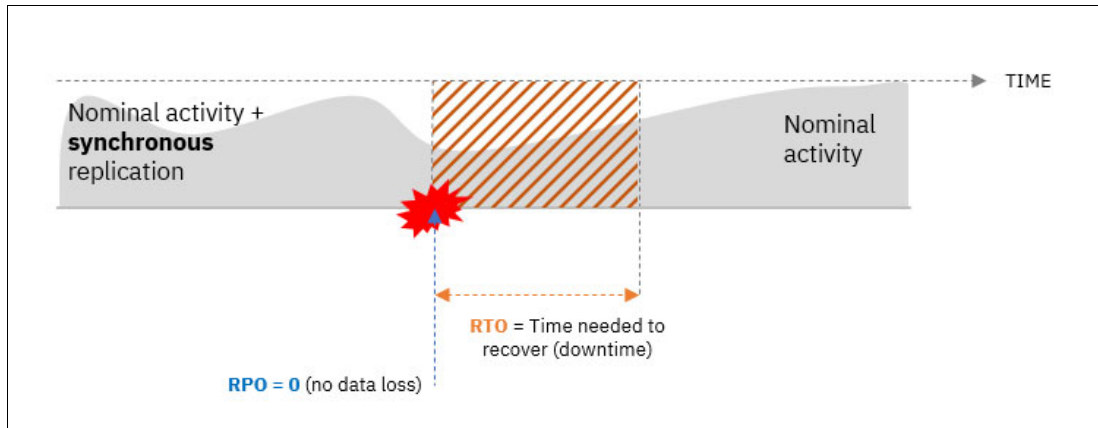


Figure 1-2 Zero RPO and non-zero RTO

Policy-based replication is an adaptive replication solution. When conditions are optimal, policy-based replication is a near-synchronous replication, the recovery point is very small (several milliseconds to a few seconds).

The RTO is not related to the type of replication between production and recovery sites. Whenever a disaster occurs, even with a synchronous replication, recovered data still needs to be presented to servers on the recovery site. The recovery time can be reduced when servers are pre-attached to the recovery system, but they still need to mount the recovered data. This is generally made manually (or scripted) and there is still a downtime for business-critical applications.

With policy-based HA, servers are pre-mapped to volumes that are instantly accessible. In case of a disaster, they automatically failover the surviving site to access the data. The recovery time in that case is reduced to zero as there is no downtime. See Figure 1-3.

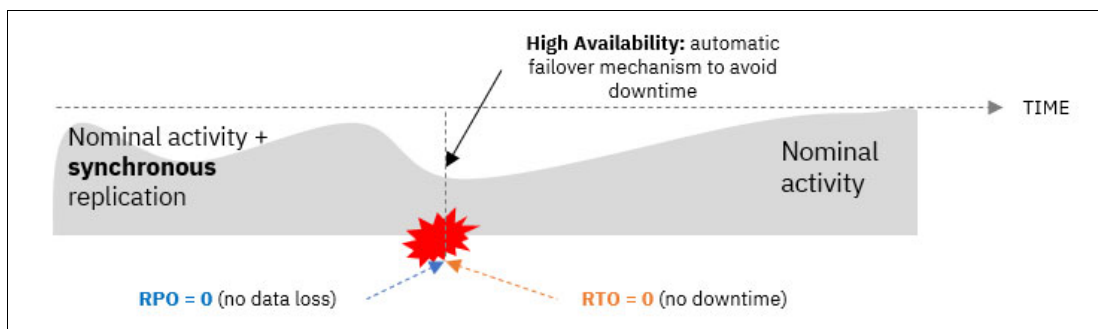


Figure 1-3 Zero RPO and zero RTO

1.2 Synchronous, asynchronous and policy-based replication

In a storage infrastructure, disaster recovery (DR) is the ability to return some data from a system after a disaster occurred. DR is implemented by replicating data, locally or remotely, depending on the nature of the risks, the localization of systems and the amount of data that clients are ready to lose.

A typical DR implementation involves a production site where applications run and access local data. Additionally, a secondary or recovery site stores copies of this production data.

This ensures that even if the primary site becomes unavailable, you can access and restore critical data from the secondary location.

1.2.1 Synchronous replication

Synchronous replication prioritizes data equivalence between the production and recovery sites. In this approach, every write operation to the production system is first copied to the recovery system. The write operation is acknowledged only after successful confirmation of the copy operation.

This method guarantees that both production and recovery systems maintain identical data copies. However, there are trade-offs:

- ▶ **Increased write response times:** Since writes involve sending data to the recovery site and waiting for confirmation, application performance can be impacted.
- ▶ **Impact of Round-Trip Time (RTT):** The longer the distance between the production and recovery sites (measured in milliseconds or ms), the higher the write response times due to the additional data travel and confirmation cycle.

Therefore, synchronous replication is best suited for scenarios with very low RTT (ideally less than 1 millisecond) to minimize performance drawbacks.

See Figure 1-4.

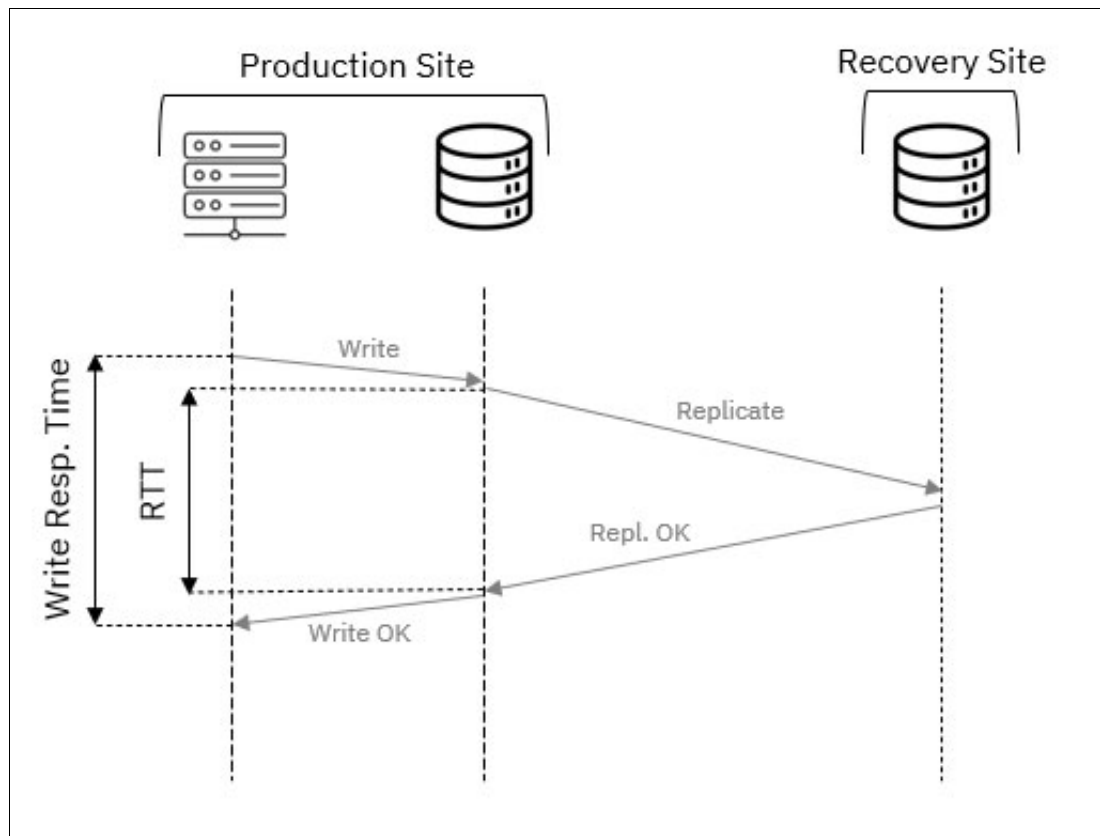


Figure 1-4 Synchronous replication and RTT impact

In earlier versions of IBM Storage Virtualize, synchronous replication was facilitated by the "Metro Mirror" remote copy service. However, for geographically dispersed sites where

distance creates high Round-Trip Time (RTT), synchronous replication becomes impractical. This is where asynchronous replication comes into play, which is covered next.

1.2.2 Asynchronous replication

Asynchronous replication consists in dissociating replication processing from hosts writes. Unlike synchronous replication, hosts do not have to wait for write operation completion on the recovery storage system.

This approach demands sufficient bandwidth between the production and recovery sites. The bandwidth needs to be able to handle the write throughput of the production system to minimize the amount of data waiting to be replicated.

Additionally, since the hosts do not wait for the replica to be completed on recovery site, there might be a gap between data in production and in recovery sites if a disaster occurs and the replica are not completed.

Asynchronous replication was managed by “Global Mirror” on previous versions of Storage Virtualize.

1.2.3 Asynchronous replication with snapshots

To optimize the bandwidth utilization, asynchronous replication with cycling mode (snapshots) can be used. This type of asynchronous replication captures periodic snapshots of the production data. Only the changes that occur between these snapshots are copied to the recovery site, reducing the amount of data transferred.

By dissociating application server activity and data replication, this method optimizes overall system efficiency. Applications and replication operate independently, minimizing performance bottlenecks. See Figure 1-5.

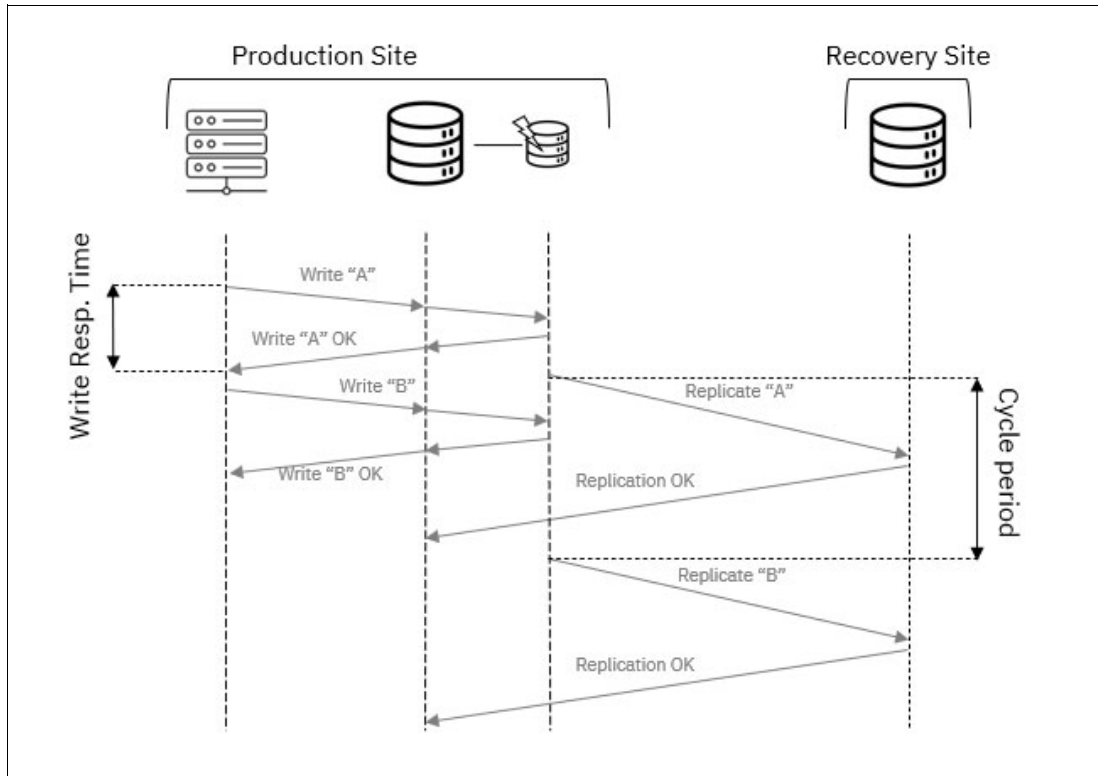


Figure 1-5 Cycle-based asynchronous replication

In this mode, replicated data on the recovery site can be older than the production one because data is very likely to change since the last completed cycle. In the example above, Recovery Site started receiving "A" while "B" was being written on Production. Recovery site will receive "B" on next cycle.

It is the write change rate on data that will determine the size of the snapshots and therefore the amount of data to be replicated. Some areas of data volumes can change several times between cycles, but only the latest are replicated, reducing the amount of data to be replicated.

The frequency of the cycles will dictate the age of the latest available copy on recovery site. The frequency of the cycles should be high to minimize the time gap between a disaster event and the latest completed cycle.

In earlier versions of IBM Storage Virtualize, asynchronous replication was managed by two primary remote copy services:

- ▶ **Global Mirror:** This service facilitated basic asynchronous replication.
- ▶ **Global Mirror with Change Volumes (cycling-mode):** This advanced version offered asynchronous replication with snapshots, similar to the functionality described in this section.

1.2.4 IBM policy-based replication

IBM policy-based replication employs a single algorithm that incorporates both asynchronous snapshot-based and full data replication methods. This intelligent approach automatically switches between these modes (cycling mode and journaling mode) based on the available replication bandwidth, ensuring efficient data transfer.

Consequently, the system always strives to provide the best possible recovery point based on the current workload and available bandwidth. With journaling mode, this is achieved by using a journal to record every write operation on the production volumes. The system monitors this journal and triggers replication operations dynamically, eliminating the need for predefined replication cycles.

Journals are used in journaling mode and snapshots are used in cycling mode, maintaining consistency at all times. To ensure consistent data on the recovery site, the system automatically creates a snapshot before initiating the resynchronization process. This snapshot guarantees that the order of writes on the recovery site mirrors the production site, maintaining data integrity.

To achieve a high frequency replication and maintain the most recent data on the recovery site, the bandwidth between the two sites needs to be sufficient to handle the write throughput of the production site.

Policy-based replication has three operation modes:

- ▶ A "Change Recording" mode that simply tracks the changes on production site, without replicating to the recovery site.
- ▶ A "Journaling" mode which tracks and replicates, in order, the changes made on production environment.
- ▶ A "Cycling" mode where new host writes are tracked and periodically replicated from a snapshot of the production volume.

Journaling mode is the preferred replication method because it offers a lower RPO. However, the system might switch to cycling mode if it cannot sustain the write volume required for journaling due to bandwidth limitations. In cycling mode, the system captures periodic snapshots of the production volume and replicates only the changes since the last snapshot. This reduces the amount of data transferred but increases the potential recovery point.

The frequency of these cycles in cycling mode is determined by the acceptable recovery point objective. More frequent cycles minimize data loss but require more bandwidth. The system automatically balances these factors to meet the RPO for the volumes. The administrator can use different replication policies with different RPOs to prioritize replication between different applications.

See Figure 1-6 on page 8.

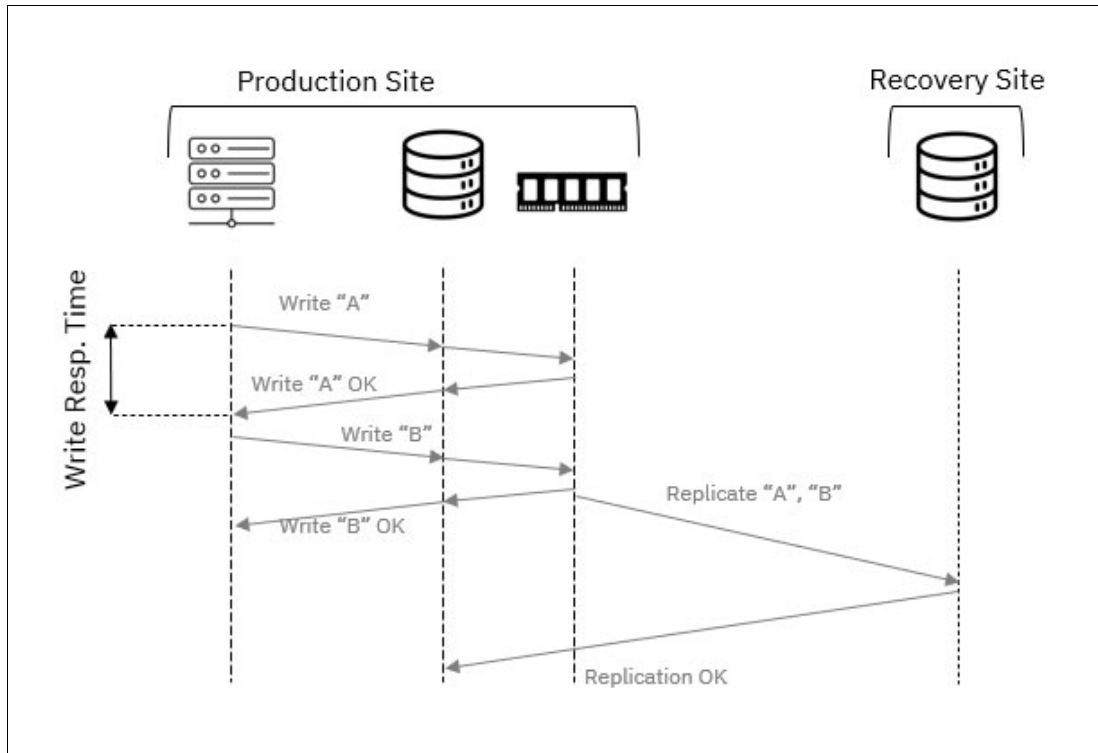


Figure 1-6 Policy-based replication with journaling

1.3 Data consistency

Regardless of the replication method that is used between two systems, it is essential for business continuity to ensure the integrity of the data replicated on the recovery site. Consistent data on the recovery site means it can be read by an application or operating system in case of a failover.

Data consistency applies within each volume and across volumes. Some applications need blocks from different volumes to assemble exploitable data, so consistency needs to be maintained between volumes associated to the same application.

The order in which data changes are applied on the recovery site (within a volume or across volumes of a volume group) is crucial to maintain data consistency. IBM Storage Virtualize policy-based replication uses an in-memory journal while in journaling mode. The journal tracks the changes that are made on volumes within the volume groups, in sequenced order. The journal in journaling mode acts as a buffer for write I/Os on the production site. This allows data to be written locally without waiting for the entire replication process to finish. Hosts can continue operations without delays caused by replication.

In cycling mode, to ensure data consistency during resynchronization, the system automatically creates a snapshot of the volumes before initiating the process. This snapshot provides a known, consistent state of the data. If the resynchronization fails, the system can revert to this snapshot, guaranteeing data integrity on the recovery site.

1.4 Policy-based HA

In storage infrastructure, high availability (HA) ensures that applications on hosts can access their data continuously, even if there's a failure in the primary storage system. This solution is achieved by maintaining a full copy of the data and synchronization on a peer system that allows for application access through either system, so that data access is maintained even during a disaster.

Policy-based HA (PBHA) is an active/active high availability solution. Both copies of the volume are accessible while HA is established where the hosts can submit I/O to either copy and synchronization is maintained between the copies.

Policy-based HA uses a synchronous replication so data equivalence between production volumes copies are guaranteed. Volume groups are used to manage the consistency across volumes which are application interdependent. In a highly available solution, the hosts from all sites must have access to the same set of production data. To facilitate this behavior in an easy-to-use way, IBM Storage Virtualize introduces a concept of Storage Partitions. Storage partitions are a collection of related volume groups, hosts and mappings. See Figure 1-7.

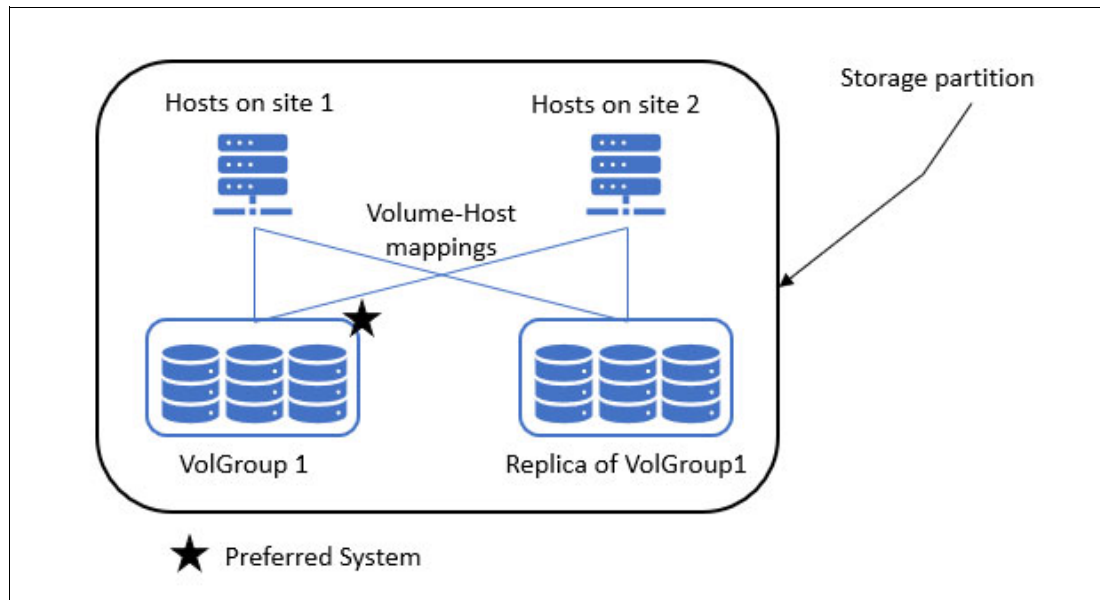


Figure 1-7 Storage partitions

With storage partitions, users do not have to worry about mapping manually the hosts to the volume groups replica, as it is already prepared on both systems. They do not have to worry either about volumes recognition by the hosts, as the UID is the same.

To ensure complete consistency, HA replicates not only data, but also the configuration of the storage partition (including host definitions and mappings) to the remote site whenever changes occur. This configuration information is typically managed on a 'preferred system' for the selected partition and stored on both systems.

In HA configurations, a *preferred management system* is designated per storage partition. The preferred management system is typically the system where the partition is managed. It is the system that will be preferred for the storage partition to continue to be accessible and managed through in the event of a disconnect between systems. If host locations are not set, hosts will access the volumes through this system while HA is established.

When the location is explicitly set for hosts, read and write operations become *localized*. This means hosts at a specific site will access the copy of the data available on the storage system at the same site, assuming the host location is configured correctly.

Additionally, management of the volume groups, storage partitions and the policy are centralized on an *active management system*. The active management system will usually be the preferred management location.

If an outage or other failure happens on the current active management system, the active management system will automatically fail over to the other system.

In case of a system failure on a local site, hosts from the local site will automatically switch to the system on the remote site to access the data, using their ALUA-compliant multipath policy.

1.5 Summary of storage business continuity strategies

A business continuity strategy should consider many factors. First of all, not all applications (hosts and volumes) need the same level of protection or continuity. The standard approach is to determine what RPO and RTO must be achieved for a given set of data. Costs, ease of management or administrators' knowledge are also factors that should be considered.

From a pure storage infrastructure perspective, IBM Storage Virtualize policy-based replication protects your business from data loss on production site by replicating them on a recovery site. The RPO is optimized to be minimum and tracked so you can control or avoid data loss. Policy-based HA adds a different layer of protection for your business by avoiding downtime with automatic fail-over from a production site to a recovery site. See Table 1-1.

Table 1-1 Business continuity options

Replication and HA technique	RPO	RTO	Constraints
Legacy Metro Mirror	0	> 0 (not HA)	Short distance / low RTT
Legacy GM/GMVCV	Seconds to hours	> 0 (not HA)	Dependent on link quality and bandwidth
Legacy HyperSwap®	0	0	No storage partitioning (system-centric solution)
Policy-based Replication	Adaptive: from near-zero to hours/days	> 0 (not HA)	Adaptive RPO, must be monitored
Policy-based HA	0	0	Short distance / low RTT

1.6 IBM Flash Grid

IBM Storage Virtualize, with the adoption of storage partitions and volume groups, dissociates the business continuity requirements (HA, replication) from hardware systems and moves

further in the creation of multiple software-defined virtual storage systems within a single FlashSystem deployment.

The Flash Grid approach allows users to create federated, scalable clusters of independent storage devices and failure domains. From an application angle, through the use of storage partitions, it allows users to add HA and/or DR resilience to applications through manual or automated non-disruptive data movement. It also enables easier device migration and consolidation and rebalancing of storage capacity and performance over several systems.

Clients can aggregate IBM FlashSystem or SVC systems and manage them as a single scalable storage grid, engineered for high availability, replication, and non-disruptive application data migration. Systems that are involved in replication and HA can participate to the same Flash Grid.

The historical approach of clustering nodes with IBM Storage Virtualize was a “per I/O group” one. Pairs of nodes were the bricks of a cluster solution design that was more “scale-up” oriented. With the introduction of IBM Flash Grid, the clustering granularity is slightly different. It is now the systems themselves that can scale-out, and form a single solution, with a single point of management. There are fewer requirements around hardware compatibility and the performance and capacity can now scale linearly.

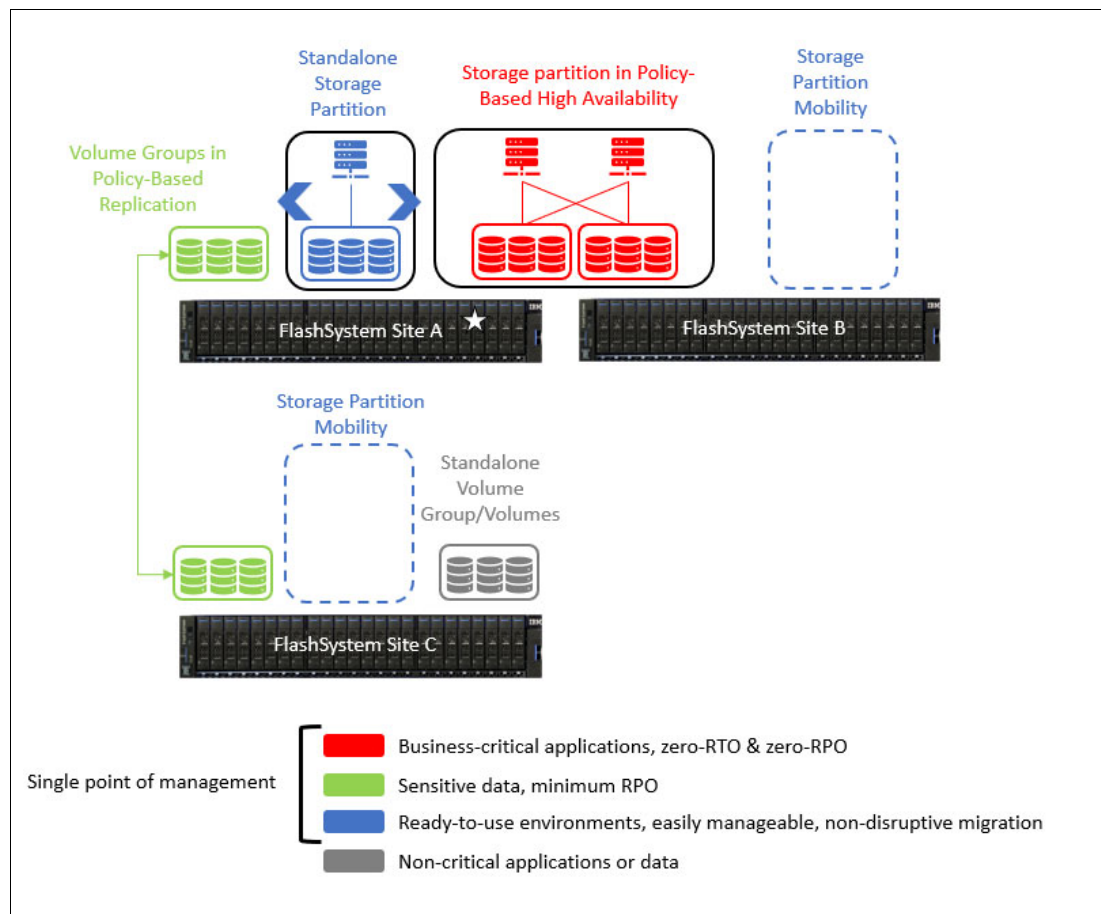


Figure 1-8 IBM Flash Grid concept

Once storage partitions are configured, they can easily be moved from a system to another, manually balanced by users over several systems and sites. They can also be stretched over two sites for high availability. See Figure 1-8 on page 11.

At the time of this writing, Flash Grid features (partitions mobility) are manageable with IBM Storage Insights Pro only.

It is possible to use the CLI to create a Flash Grid and add or remove systems in a Flash Grid.



Policy-based replication

IBM policy-based replication is a significant advancement in managing data replication for FlashSystem and SVC storage, making it easier and faster to achieve data protection and improve storage efficiency. This chapter discusses policy-based replication and has the following sections:

- ▶ “Introduction” on page 32
- ▶ “Policy-based replication - Asynchronous” on page 32
- ▶ “Recovery and testing” on page 36
- ▶ “More information” on page 40

2.1 Introduction

Policy-based replication is the method by which IBM Storage Virtualize provides replication services. Policy-based replication was introduced into Storage Virtualize 8.5.2 and is a replacement for Remote Copy in previous versions of code. It is a *ground up implementation of code* using the most recent coding advancements and techniques. Policy-based replication offers many performance and automation advantages over Remote Copy.

True to its name, this solution uses policies (such as provisioning and replication policies) to define the overall replication behavior. Volume groups serve as the smallest unit, and the assigned replication policy dictates how data is replicated. The replication policy states what systems to replicate between and the desired recovery point objective (RPO). Different volume groups can have differing policies to allow an organization to replicate to one or multiple (3) DR systems as well as prioritize the value of sets of data. This prioritization only occurs if resources are somehow restricted. If no restriction exists, all volume groups are treated with equal value. By nature of the volumes being in a volume group, the system will maintain consistency between them.

Policy-based replication was designed to free a system administrator from having to manually configure and maintain replication. With older Remote Copy, an administrator had to manually configure and maintain remote replication targets, relationships, consistency groups, change volumes etc. By using policy-based replication, all these manual steps have been automated, freeing system administrators to focus on higher value tasks.

Note: Currently, policy-based replication is asynchronous only. It is possible that in the future a synchronous option will be added.

2.2 Policy-based replication - Asynchronous

Asynchronous policy-based replication is a direct replacement for the Remote Copy Services: Global Mirror, Global Mirror with Consistency Protection and Global Mirror with Change Volumes. In the past, each of these modes had to be chosen manually and if the mode needed to change this was also a manual intervention. Policy-based replication combines all three of these methodologies and will automatically switch between methodologies on a volume group by volume group basis if necessary. It uses the RPO designated on the policy to use QOS to try to keep all volume groups within their stated RPO.

The bandwidth limit on the partnership used for asynchronous policy-based replication will dictate how many data can be sent between systems. This is helpful so a particular system does not overload the shared inter-site WAN link. The bandwidth limit value does not take into consideration data compression, so if the data is being compressed by native IP based replication or FCIP based replication, set the limit accordingly. The bandwidth limit is per I/O group on a multi-I/O group system.

There are generally two types of methodologies for asynchronous replication that will be explored in more depth in the following sections. These methods are journaling, which is used by Global Mirror/Global Mirror with Consistency Protection, and cycling, which is used by Global Mirror with Change Volumes.

2.2.1 Asynchronous replication - Journaling

Journaling is a method that captures frames in a journal/buffer, sequences them and sends them over in order. When a write from a host comes into the system, an acknowledgement is sent to the host and then the write is sent to a journal/buffer in either memory or on disk where it is stored until it can be sent across the link to the remote site. Assuming no constraints, this method can produce a very low RPO. With policy-based replication, this can be as low as one half the round-trip time to the secondary site. The disadvantage to this method is that when the amount of data overruns the link or if the link has problems, the journal can run out of space and potentially slow down or stop replication. Policy-based replication monitors journal resources proactively and will take action to prevent journaling from slowing down or stopping replication.

In IBM Storage Virtualize the journaling is implemented, as shown in Figure 2-1.

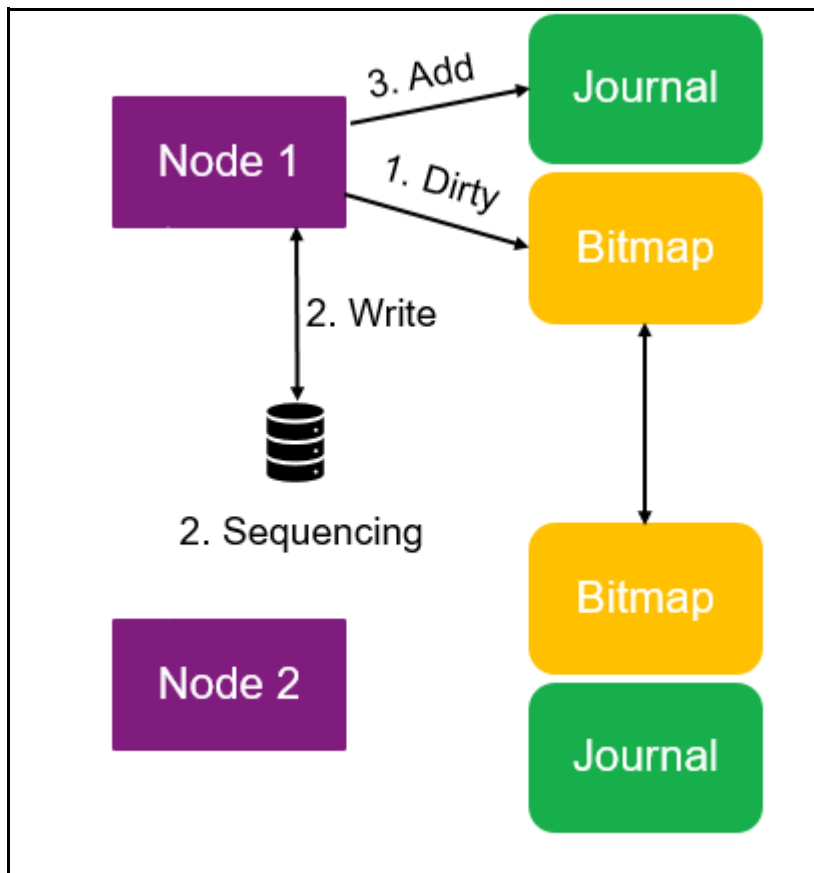


Figure 2-1 Journaling methodology

Note: IBM FlashSystem storage systems implement a dual controller architecture where there is a control enclosure, and each controller is known as a node. For SAN Volume Controller, each appliance is a node, and nodes are deployed in pairs.

In IBM's DR systems, each controller or node employs a non-volatile bitmap and a volatile journal. To understand this process, a volume is divided into fixed-size regions called *grains*, where each grain is a contiguous 128 KiB segment. Each bit in the bitmap represents the status of a corresponding grain.

When a write request arrives at a controller/node, it is mirrored to the other controller/node. The bitmap for the affected grain(s) is marked as *dirty*, indicating a pending write operation. This updated bitmap is then synchronized with the other controller or node. Subsequently, the write is sequenced, assigned a unique identifier, and acknowledged back to the host system.

Asynchronously, the write data is written to the volatile journal. This journal acts as a temporary buffer, holding the write data until it is transmitted across the network connection to the remote system for redundancy.

Since DR systems rely on sequence numbers to ensure the data gets written to the remote storage in the correct order, there's always a consistent point-in-time reflection of the data at the recovery site, even though it might not be the most recent version. The journal can be volatile because the bitmap can be used to figure out what has/has not been sent if something were to happen to the journal. The bitmap takes less resources than the journal and keeps the traffic between nodes to a minimum.

In IBM's DR systems, maintaining a consistent point-in-time copy at the recovery site relies on sending data sequentially. However, if the link fails or replication stops, restarting requires the bitmap to identify which data needs transmission.

Since the journal might be unavailable during this interruption, the order of transactions and potentially some data might be lost. In such scenarios, change volumes are used to create a (potentially outdated) recovery point on the target site while resynchronization occurs. This ensures a recoverable state until data synchronization is complete.

Change volumes are always used during resynchronization, so factor them in when designing policy-based replication.

Tip: As a general rule, change volumes can consume up to 10% of the storage capacity on both the production and recovery systems.

While not visible through the GUI, change volumes can be accessed using the `lsvdisk -showhidden` command or the `lscmap -showhidden` command.

2.2.2 Asynchronous replication - Cycling

The other method of asynchronous replication deploys a recurring cycle. That cycle involves taking a point-in-time snapshot at the primary site, sending the changes to the secondary site, when the changes are at the secondary site, take a point-in-time snapshot at the secondary site to have a consistent point to recover from if necessary and then repeat the cycle. The change volumes that are associated with the relationship will be used for these point-in-time snapshots.

The advantage to this method is that it tolerates low bandwidth links and problems with site connectivity well. The disadvantage to this method is that it is hard to maintain a very low RPO. See Figure 2-2.

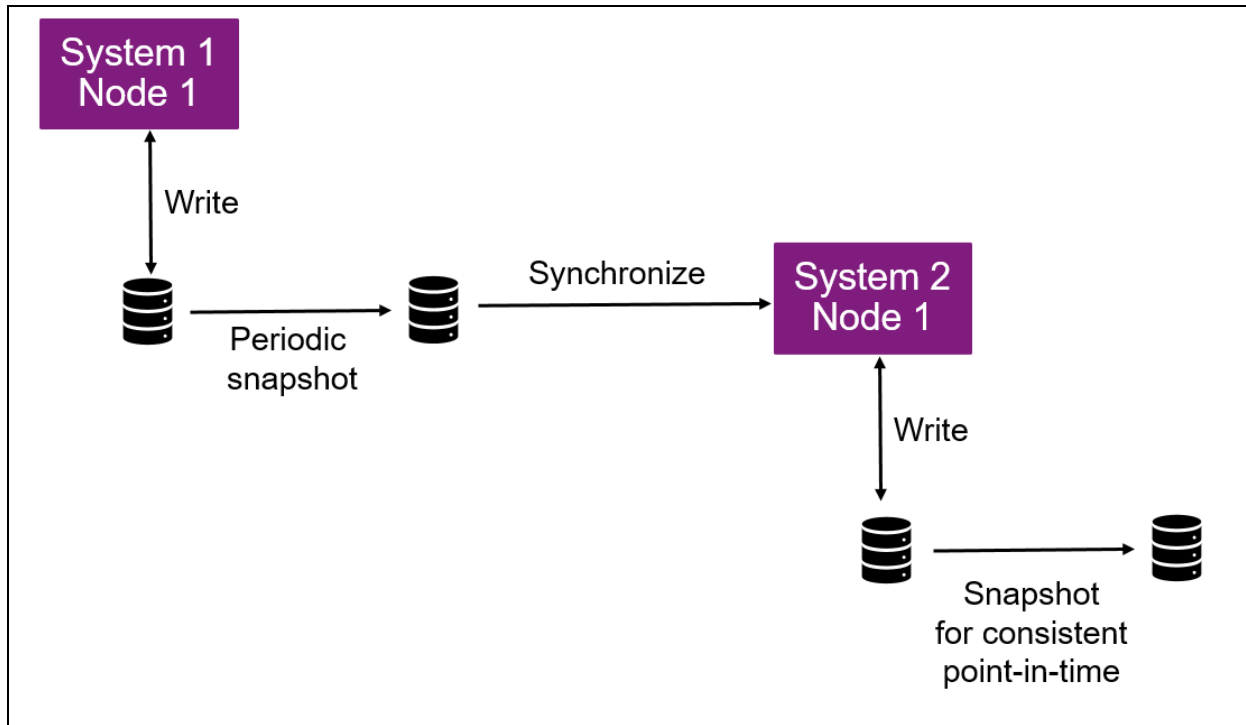


Figure 2-2 Cycling methodology

2.2.3 Quality of Service (QOS)

Unlike remote copy, policy-based replication offers quality of service based on the RPO alert on the replication policy attached to specified volume groups. This allows the system to prioritize some traffic over other traffic. The RPO alert on a replication policy can be set from 1 minute up to 1440 minutes (24 hours). The system will always attempt to keep all volume groups within the RPO alert stated on their replication policy.

Barring constraints, policy-based replication will *always* prefer journaling mode regardless of the stated RPO on the policy. This will keep all volume groups to the lowest possible RPO.

If constraints begin to appear on the system or connections, policy-based replication may choose to convert some or all the volume groups into cycling mode and base that decision on the stated RPO alert on the policy. If or when the constraint no longer exists, the system will choose to convert some or all of the volume groups back to journaling mode. The end user does not control whether policy-based replication is using journaling or cycling, nor should they care as long as the volume groups are within the stated RPO.

Take a system, for example, with two replication policies to the same target system. One policy has an RPO alert of 5 minutes and the other has an RPO alert of 60 minutes. In this example the client has a peak workload in the evening that overloads their connection bandwidth between the two sites. When constraints appear on the system, it may convert some or all of the volume groups with the 60-minute RPO policy to cycling mode, keeping them within their 60-minute RPO so it can dedicate more bandwidth to the volume groups with the stated 5-minute RPO keeping them in journaling mode. When the constraint no longer exists, the system will convert the volume groups that it put into cycling mode, back into journaling mode.

2.3 Recovery and testing

When performing replication between sites it is sometimes necessary to recover to the recovery site as well as perform intermittent testing to prove recovery mechanisms work and provide the level of recovery desired. Enabling independent access at the recovery site provides the ability to recover or test. The downside to enabling access at the recovery site is that replication is suspended and the recovery point increases while testing occurs. To mitigate this, some choose to take a snapshot of the volume group on the recovery site and then create a thin clone to test with. While this provides a testing mechanism, it does not prove that the actual target volumes are recoverable. A third method is to start a recovery test on the volume group in question. This method allows the target volumes to come online and be tested while replication continues. The recovery test function was released in Storage Virtualize 8.6.2. These methods will be explored in more depth.

2.3.1 Enable access

Target volumes in a policy-based replication volume group are offline and therefore inaccessible to the servers. To bring them online and mount them to a server, access must be enabled. Enabling access at the target site *must* be done from that site. The command cannot be issued from the production storage device or site. Enabling access creates two independent copies of the data. Both copies are tracked for changes so that replication can be started in either direction. Only the deltas are copied across the link. Changes can be forced from site 1 to site 2 if that is to be the primary copy or changes can be forced from site 2 to site 1 if necessary. *Replication must be restarted from the storage device that is to be the primary copy.* For example, if a real disaster was declared and production is now running on site 2 (it was on site 1 before), when replication is restarted, it must be started from site 2 for changes to be replicated from site 2 to site 1. Figure 2-3 on page 37.

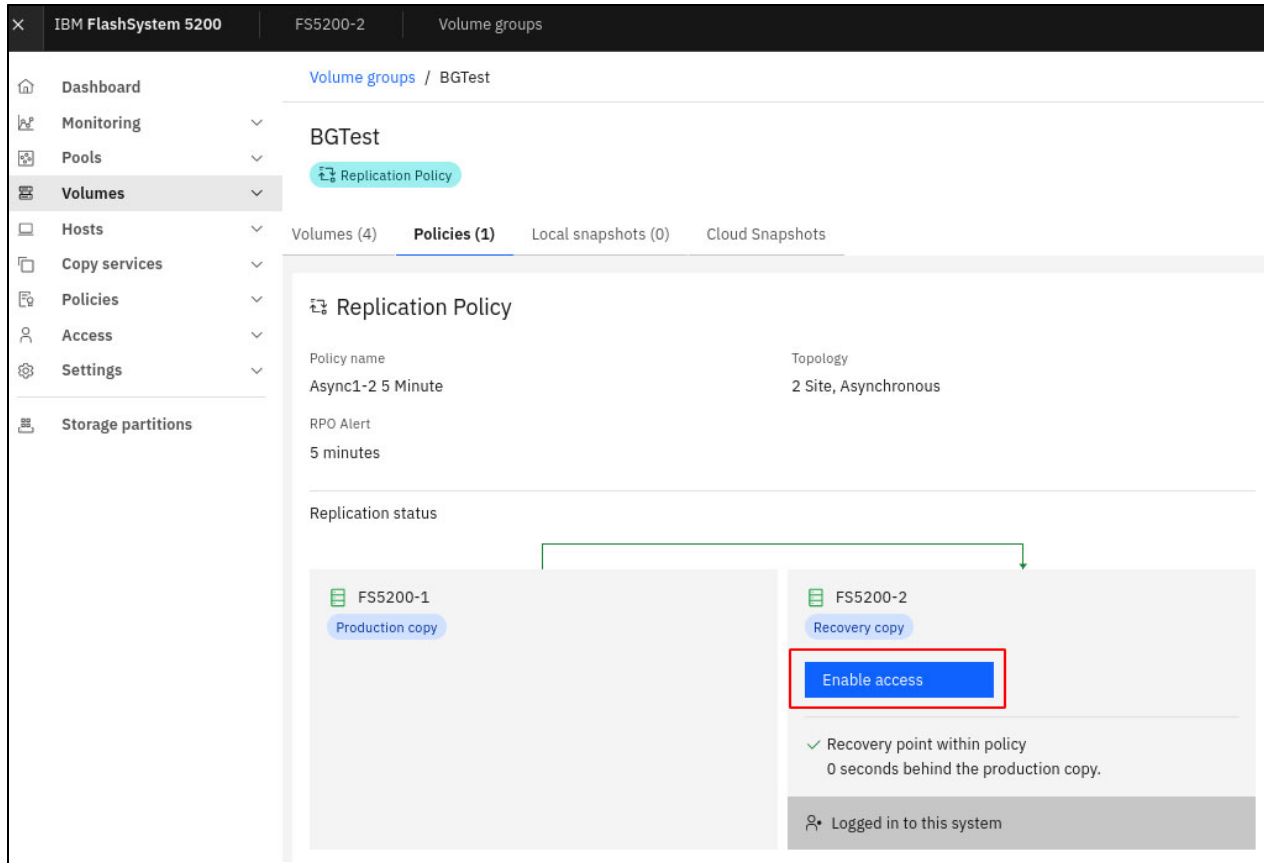


Figure 2-3 Enable access 1

Enabling access is the means for disaster recovery and a means of testing also. However, if testing is the primary function, one of the other two methods are better suited. See Figure 2-4 and “Enable access 3” on page 38.

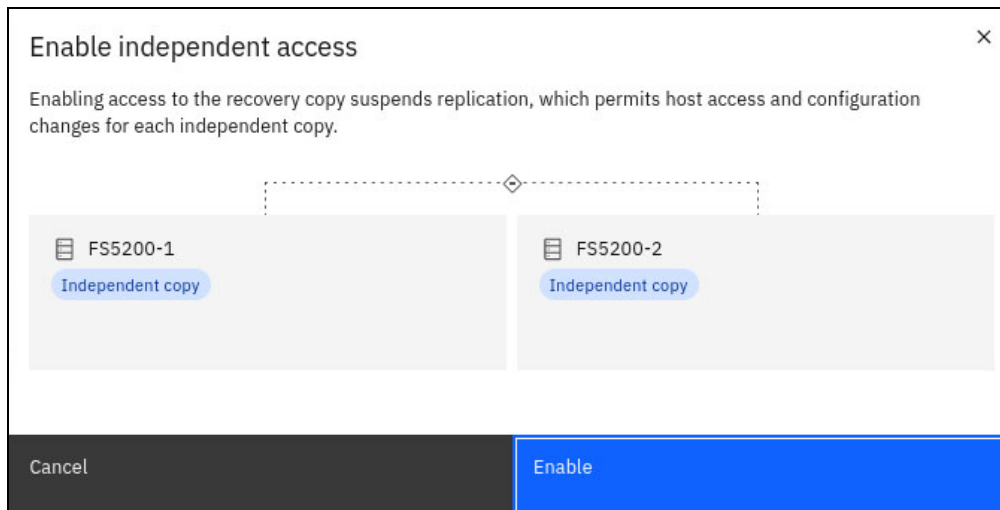


Figure 2-4 Enable access 2

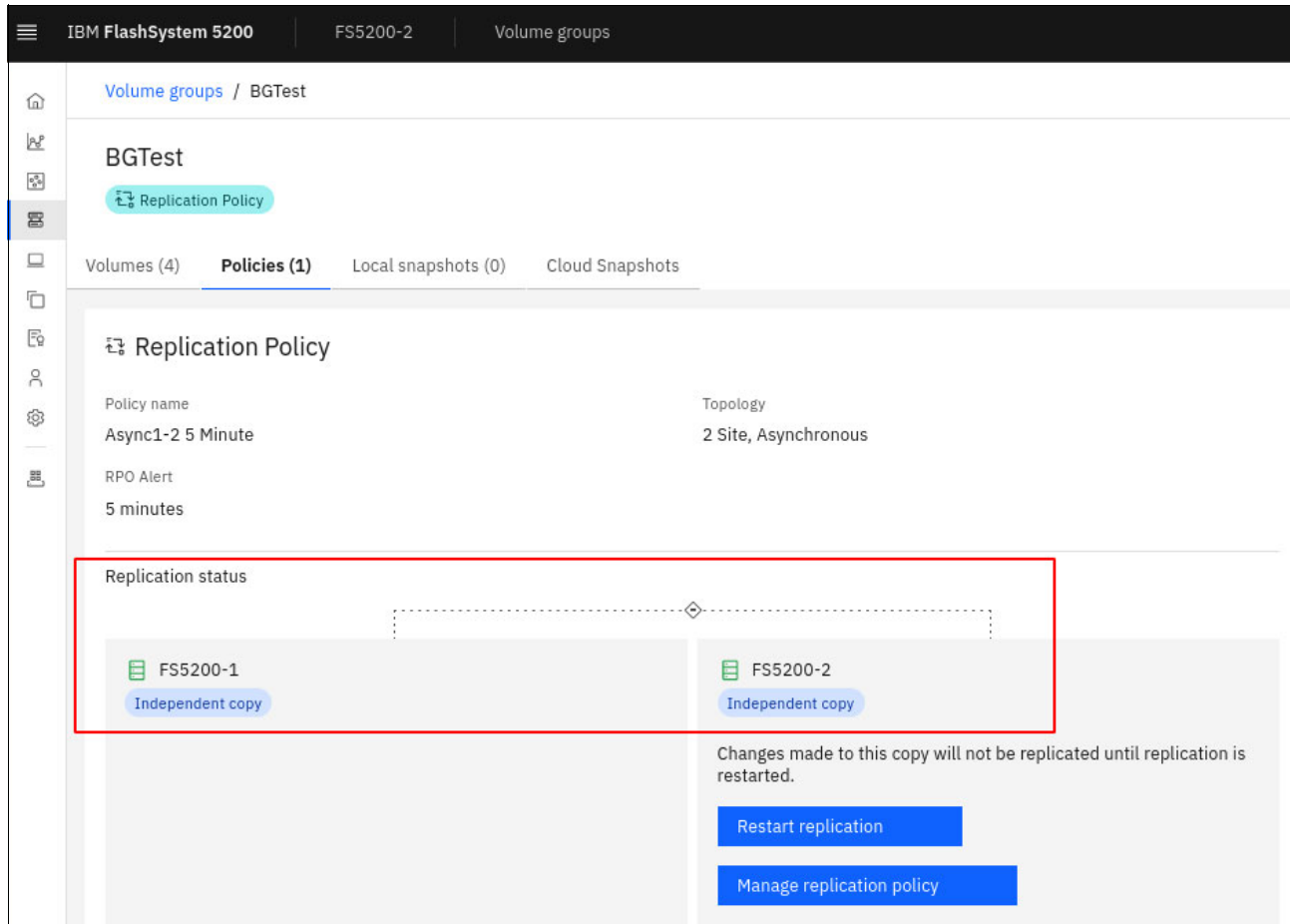


Figure 2-5 Enable access 3

2.3.2 Snapshot and thin clone

Enabling access allows for testing, but also stops replication to do it. If testing is the primary objective a snapshot and thin clone can be used to allow for testing at the secondary site. This would require taking a snapshot of the secondary volume group and creating a thin clone to mount to servers for testing. Once testing is completed the thin clone volumes/volume group can be deleted as well as the snapshot. While this method works, some might not like it since the actual secondary volumes are not tested.

2.3.3 Recovery test

Starting with Storage Virtualize code 8.6.2 there is an option for a recovery test allowing the target volumes to come online and be mounted to a server while replication continues in the background. This allows for the actual target volumes to be tested, proving that they can be recovered from in the case of a disaster. This option is only available from the command line and must be issued on the target storage array.

To initiate a recovery test on a volume group named BGTest, enter the following command on the command line interface (CLI) or REST API:

```
chvolumegroupreplication -startrecoverytest BGTest
```

Figure 2-6 shows that the recovery test is in progress.

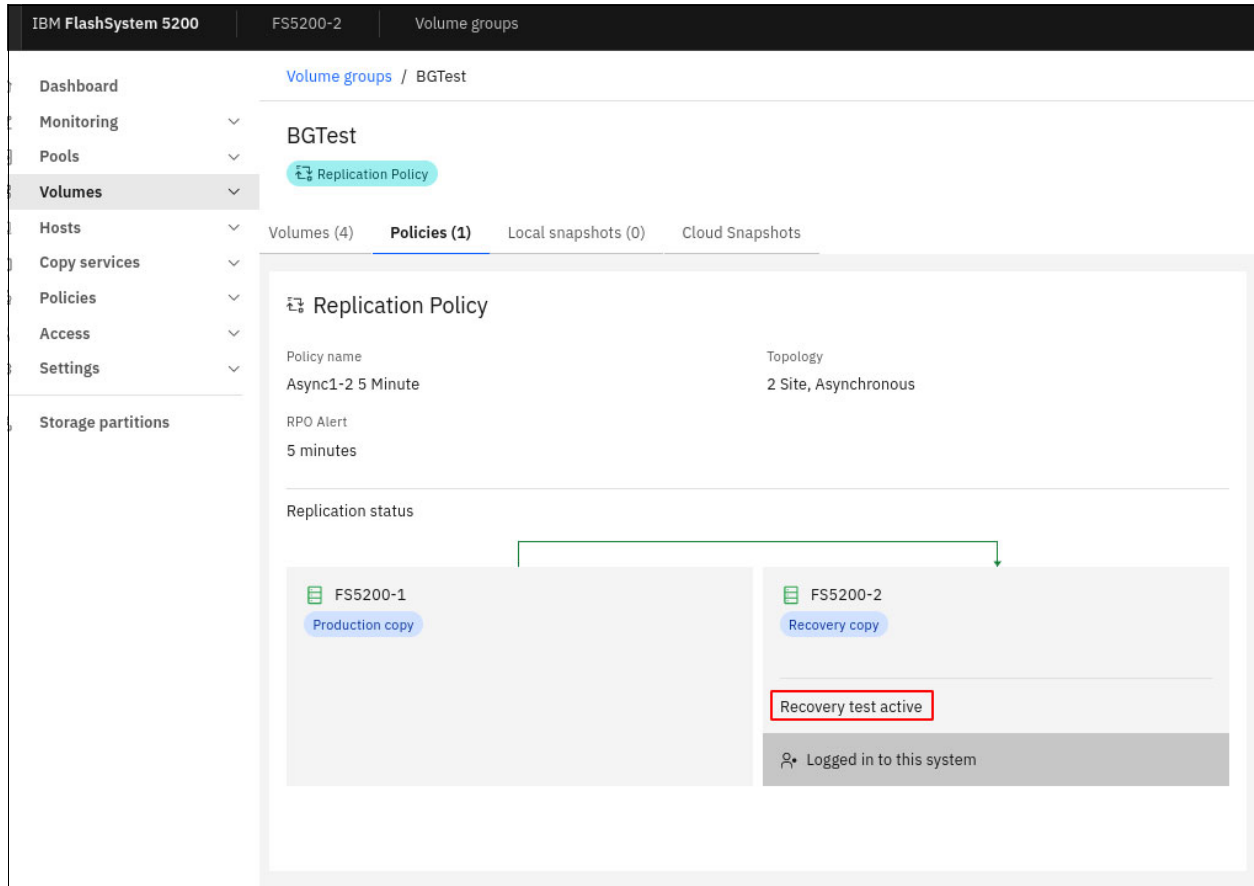


Figure 2-6 Recovery test In progress

Recovery volumes are offline, but when a recovery test is initiated these volumes come online and are able to be mounted to a server with read/write access for testing. See Figure 2-7 on page 39.

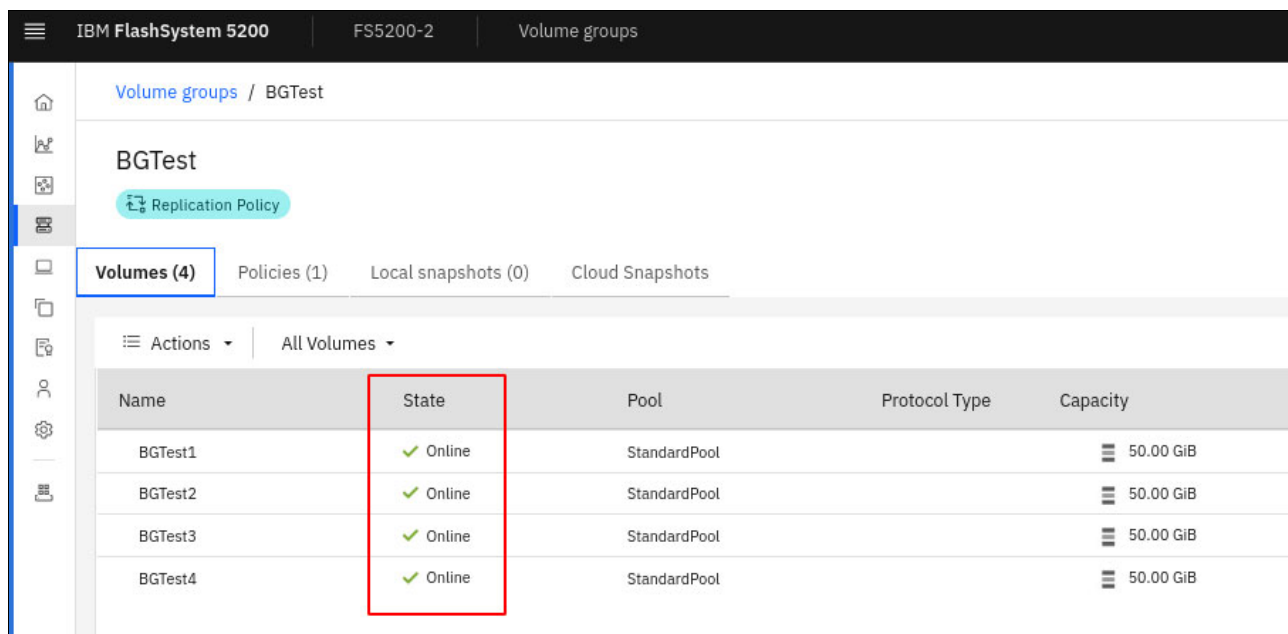


Figure 2-7 Recovery volumes online

When the recovery test is terminated, all changes to the volumes in the recovery volume group will be overwritten by what has changed on the source and the recovery volumes will once again be offline. To stop the recovery test on a volume group named BGTest enter the following command:

```
chvolume group replication -stoprecoverytest BGTest
```

Using the recovery test method allows testing on the target volumes and less data needs to be updated when the test is done (compared to the enable access method).

Recommendation: It is recommended that a snapshot of the target volume group be taken before the recovery test is initiated.

2.4 More information

A more thorough and in-depth discussion of policy-based replication can be found in the following IBM Redpaper: *Policy-Based Replication with IBM Storage FlashSystem, IBM SAN Volume Controller and IBM Storage Virtualize*, REDP-5704.



3

Policy-based high availability

In this chapter, we discuss policy-based high availability (policy-based HA). This chapter has the following sections:

- ▶ “Introduction” on page 32
- ▶ “Policy-based HA concepts” on page 34
- ▶ “Behavior examples of policy-based HA” on page 37

3.1 Introduction

Within the last couple of years, IBM embarked on a journey to modernize FlashSystem and SVC management, achieving a more efficient and scalable storage infrastructure. The first step involved implementing policy-based management, introduced in 2022. This approach streamlines storage provisioning and simplifies overall management tasks.

With Storage Virtualize 8.6.1, IBM introduced storage partitions. This critical step laid the groundwork for a future Flash Grid architecture, offering greater flexibility and scalability for our storage needs. (Refer to Chapter 1, “Introduction” on page 1 for a detailed explanation).

3.1.1 Driving forces behind the development of a new policy-based HA solution

The emergence of policy-based HA can be attributed to several key factors:

- ▶ **Evolution of hardware:** Traditional solutions like HyperSwap were designed for older hardware environments with limited processing power (for example, up to 8 core systems). Policy-based HA aims to leverage the capabilities of modern, multi-core processors for improved scalability and performance.
- ▶ **Integration with policy-based management:** A core tenet of modern storage management is automation and ease of use. Policy-based HA seeks to seamlessly integrate with policy-based management tools, allowing for streamlined configuration and ongoing management of high availability solutions.
- ▶ **Addressing HyperSwap and stretched cluster limitations:** While HyperSwap and stretched cluster have served as a reliable high availability solutions for many years, they have some limitations that policy-based HA aims to address:
 - **Limited configuration flexibility:** In HyperSwap, all HA volumes must have their targets in the same cluster. Policy-based HA, on the other hand, offers more granular control. It allows users to selectively replicate specific volumes based on their needs.
 - **Split brain requires half the cluster to stop:** In split-brain scenarios (where communication between systems is lost), HyperSwap and stretched cluster can require shutting down half the system to maintain data integrity. Only one site stay online. Also local volumes may go offline by a decision. Policy-based HA could potentially introduce improved split-brain handling mechanisms to minimize downtime.
 - **Performance constraints:** They might have performance limitations in certain configurations. Policy-based HA could be designed to leverage modern hardware for improved replication performance.
 - **Hardware and software restrictions:** HyperSwap and stretched cluster may require identical hardware at both sites, because of clustering support. Also the software level needs to be equal in a cluster. So software upgrades needs across the cluster at the same time. Policy-based HA might offer more flexibility in hardware configurations and potentially introduce rolling software updates to minimize disruption.

Note: Storage Virtualize 8.7.0.x is the final release to support HyperSwap.

3.1.2 Simplifying storage with IBM Flash Grid and policy-based HA

IBM is revolutionizing storage management with two key advancements designed to simplify your operations, optimize performance, and minimize disruptions:

► IBM Flash Grid technology:

IBM Flash Grid will allow clients to manage storage systems as a highly available and independently scalable environment, from a single control pane, with the ability to move workloads between FlashSystem devices.

Policy-based high availability fits perfectly into the Flash Grid architecture and is ready for the future.

► Policy-based high availability:

Flash Grid builds upon the foundation of policy-based HA to deliver a future-proof HA solution with unmatched flexibility and ease of use. The major advantages are:

- **Simplified management:** Simplified management using policies to configure entire storage partitions to be highly available and the system manages the HA for the user.
- **Modular design:** Policy-based HA introduces storage partitions for granular and simplified management. These partitions act as building blocks, allowing you to group hosts, host-to-volume mappings, volume groups, and volumes. Grouping the related storage resources allows for simpler management where HA is scoped to only the resources within the partition. This modular design allows you to manage these resources as a single entity.
- **Increased fault tolerance:** An issue on one FlashSystem or SVC, whether hardware or software related, remains isolated and does not impact the other systems. This eliminates downtime by providing a highly-available solution where the volumes are always available.
- **Flexibility in hardware and software versions:** Policy-based HA allows for mixed hardware and software versions across FlashSystem units or SVC in an HA relationship. This provides greater flexibility during upgrades or maintenance cycles, as you can upgrade systems independently without affecting high availability.
- **Enhanced performance:** Policy-based HA delivers significant improvements in performance, with up to a 4x increase in throughput and a reduction in latency compared to HyperSwap.
- **Zero impact on non-HA volumes:** HA operations do not disrupt non-critical workloads running on the same system.
- **Increased scalability:** Policy-based HA supports a higher maximum of HA volumes compared to both HyperSwap and stretched cluster.
- **Automated HA configuration:** Simply assign an HA policy to a partition, and policy-based HA automatically configures everything within it for high availability. This includes remote provisioning, eliminating the need for complex manual setup.
- **Isolated recovery:** Storage partitions can be configured to prefer to continue running on different systems in the event of a split brain. This isolated recovery approach minimizes downtime for critical applications while ensuring non-HA volumes on the same system remain unaffected.

See 3.3, “Behavior examples of policy-based HA” on page 37 for some examples.

Limits and restrictions: Refer to [V8.7.0.x Configuration Limits for IBM FlashSystem and SAN Volume Controller](#) for configuration limits and restrictions apply to policy-based high availability.

Statement of general direction: In the second half of 2024, IBM intends to further enhance these features to support highly available storage with replication to a third system.

Note that all statements regarding future enhancements are subject to change.

3.2 Policy-based HA concepts

Policy-based HA is an active/active high availability solution.

- ▶ Volumes in the storage partition will be highly-available in an active/active manner while HA is established. When HA is not established, access is only through the active system.
- ▶ Writes to the preferred system for a storage partition may have lower response times as there is only one round-trip over the ISL for writes. Writes to the non-preferred system will involve an additional round trip, but data traverses the inter-site link one once to reduce network traffic. The preferred system can be changed at any time. Reads are processed by the local system.
- ▶ Policy-based HA leverages host location awareness. When a host location is defined, volumes automatically report preferred access to the storage system in the same location as the host. This optimizes performance by minimizing ISL traffic and latency for geographically distributed deployments.
- ▶ All volumes and volume groups within a storage partition do have the same copy direction. The active partition management system is always acting as copy source system. Changing the active management system to the other site will change the copy direction for all volumes within this partition.
- ▶ As policy-based HA is based on the new grid architecture it can be implemented only between two single I/O group systems.

3.2.1 Simplified management with storage partitions

Policy-based HA introduces storage partitions to manage your high availability storage. Partitions offer a modular and user-friendly approach:

- ▶ **Logical grouping:** Create custom partitions that logically group your hosts, host-to-volume mappings, volume groups, and volumes. This allows you to manage these resources as a single entity.
- ▶ **Efficient management:** Keep resources self-contained so that HA can be configured and it can be guaranteed that the storage will configure HA on all objects that the user intends to be HA.
- ▶ **Volume group requirement:** All volumes within a partition must belong to a volume group. This ensures proper organization and management of your storage resources.

Figure 3-1 shows a storage partition.

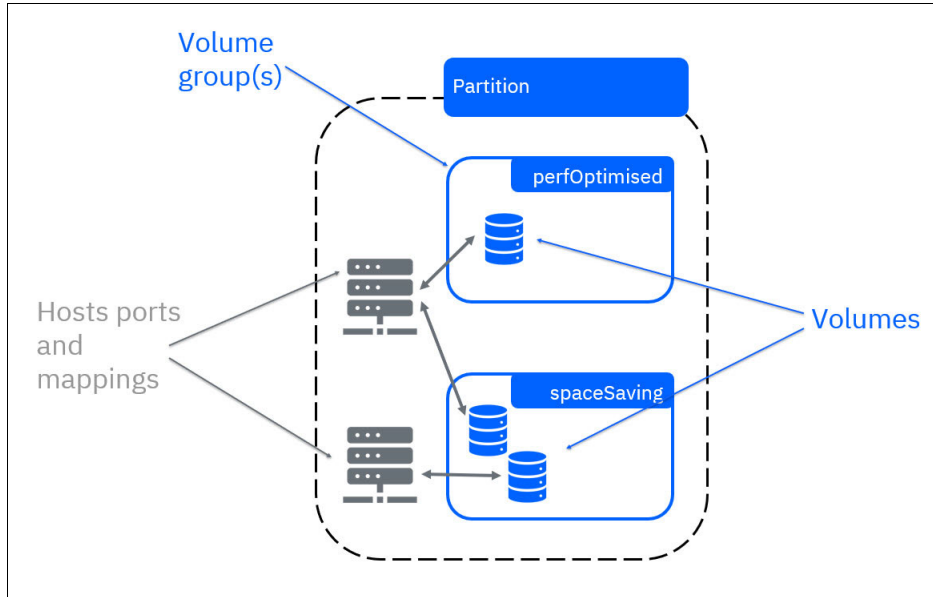


Figure 3-1 Storage partition

Currently, a maximum of four storage partitions are supported per FlashSystem. However, there is no limit on the number of volumes, volume groups, hosts, and host-to-volume mappings you can configure within a partition. You can add more resources as needed, either to existing partitions or by creating new ones. It is possible to merge partitions, if they have the same replication policy. A partition cannot be split into separate partitions.

Figure 3-2 on page 35 shows a FlashSystem example with a single IO group with 2 partitions and other local volumes.

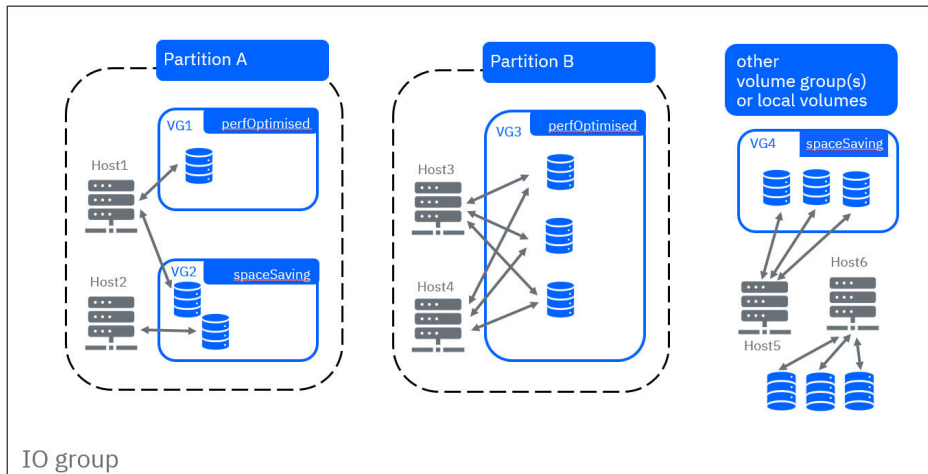


Figure 3-2 FlashSystem example - single IO group with two partitions and other local volumes

3.2.2 Draft partition

Policy-based HA lets you create a new storage partition and specify existing volume groups. The GUI will guide you through this process. A draft partition has no effect until it is published. See [IBM Documentation](#) for more details.

3.2.3 Making partitions highly available

Policy-based HA leverages familiar connection infrastructure, replicating data between sites using either Fibre Channel SCSI or High-Speed Ethernet (iWARP with RDMA).

► **Configuration steps for high availability:**

- **Partnership, pools, and policies:** Set up the foundation for HA by configuring partnerships between your FlashSystem units, storage pools, pool links, and provisioning policies.
- **2-Site-HA topology:** Apply a "2-site-HA" topology to a specific partition to enable high availability between your two independent FlashSystem units or SVC.

► **Partition management:**

- **Active versus preferred management system:** Each partition that is associated with an HA replication policy has two properties - the *preferred management system* and the *active management system*. All configuration actions on a storage partition must be performed on the active management system. The storage partition can be monitored on either system.

The preferred management system is the system that you would like to be the active management system under ideal conditions. In the event of a situation where the active management system and the preferred management system are not the same system, the system will automatically fail over the active management system back to the preferred management system when it is able. The preferred management system can be changed by the user.

- **Dynamic partition configuration:** You can create or delete volumes, volume groups, hosts, and host-to-volume mappings within a partition at any time. A volume removal from a partition requires a replication policy unassignment. A partition must include all volumes mapped to any hosts within it. The assigned HA policy automatically configures all hosts and volumes within the partition for high availability. policy-based HA utilizes an IP quorum application to determine the active management system and prevent "split brain" scenarios (both systems managing the same partition).

Best practice: Configure a second IP quorum as a backup for situations where the primary quorum fails or requires maintenance.

SAN zoning: SAN zoning to isolate traffic needs to be configured manually – it is not automated by policy-based HA.

Figure 3-3 on page 37 shows a highly available storage partition.

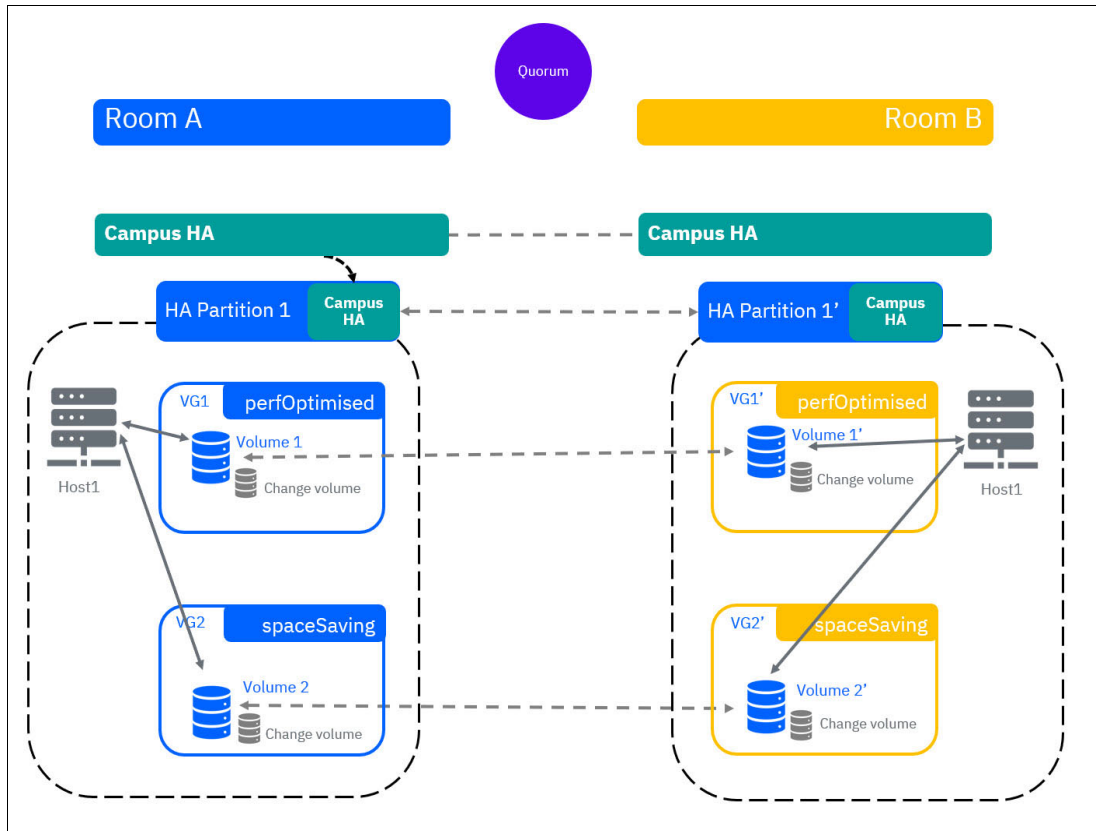


Figure 3-3 Highly available storage partition

To summarize: Policy-based HA offers a simplified and automated approach to setting up and managing high availability. With familiar connection infrastructure, pre-defined configuration steps, and automatic failover mechanisms, policy-based HA helps ensure continuous data access and simplifies storage management.

3.3 Behavior examples of policy-based HA

Policy-based HA delivers robust high availability with active/active access and site affinity for your storage systems. This means applications can access data simultaneously from either FlashSystem, preventing disruptions during failures.

Policy-based HA leverages site awareness. Host site awareness is the (optional) ability to set a location for each host such that when HA is established the I/Os are directed to the storage system in the same location as the host.

What happens if you do not configure site affinity: Without site affinity configured, all hosts will prioritize the preferred partition, potentially causing brief disruptions during a failover as the multipathing driver reconfigures paths. Additionally, all traffic from hosts at the non-preferred site to the active management system will traverse the public SAN. Therefore, it is crucial to consider the required bandwidth when sizing the public SAN to avoid bottlenecks.

Figure 3-4 on page 38 shows an HA configuration without host locations.

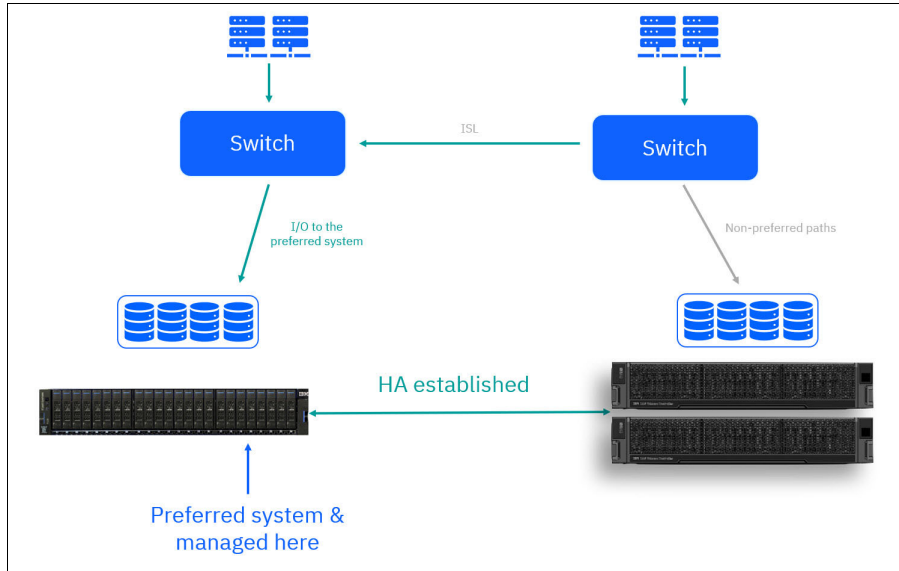


Figure 3-4 HA without host locations

Recommendation: For optimal performance and efficient high availability, we strongly recommend assigning site attributes to all your hosts. This simple configuration step unlocks significant benefits for your applications and simplifies storage management.

The site attribute is the name of the IBM FlashSystem or SVC. Figure 3-5 shows the data flow, if site attributes are used.

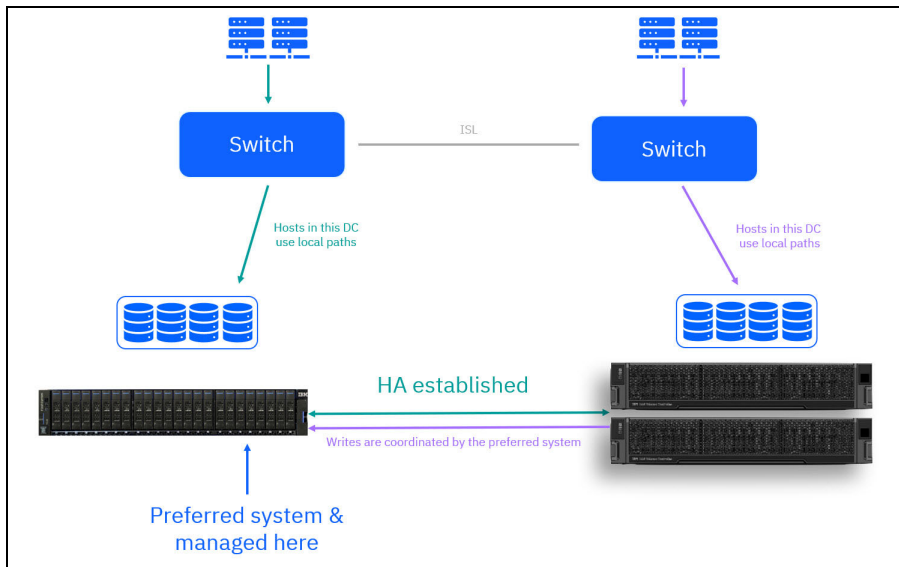


Figure 3-5 HA with host locations

Figure 3-6 on page 39 shows that the host multipath driver is directing I/O to the one surviving system.

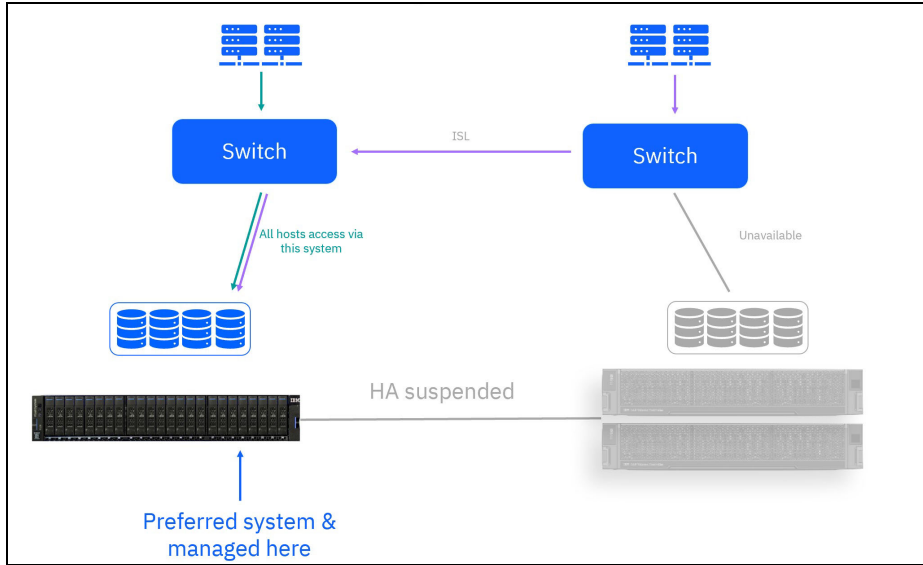


Figure 3-6 HA with host locations - Storage failure

Split brain scenario

In a split-brain scenario (lost communication between sites), each policy-based HA partition leverages its own IP quorum to determine the active management system. If the preferred site is available and has the majority of quorum votes within a partition, that partition will likely remain managed by the preferred site. However, quorum votes ultimately dictate the active management system, not just the preferred designation.

The preferred management system attribute is defined on a per-partition basis. The system-wide "preferred quorum" parameter has minimal influence on policy-based HA behavior, so you can typically leave it at its default setting. See Figure 3-7.

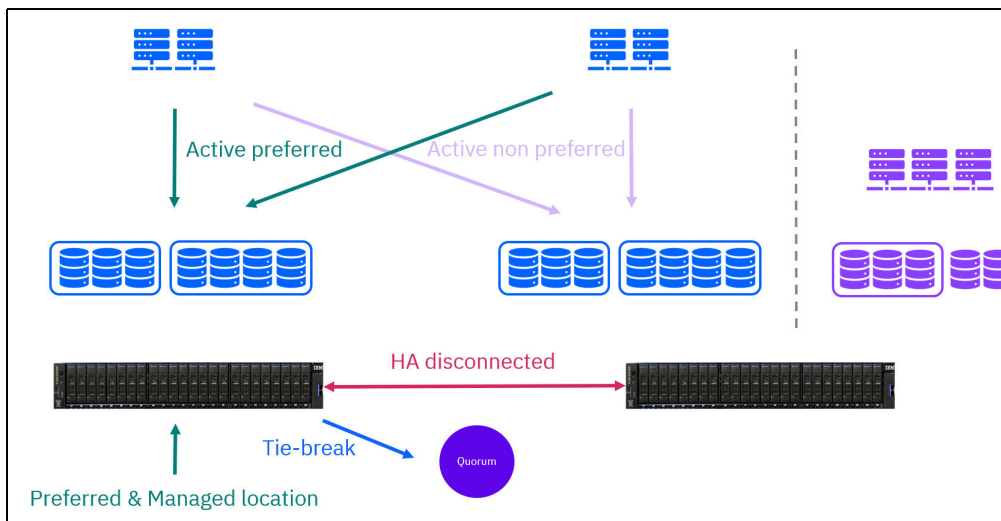


Figure 3-7 HA storage partitions - split brain scenario 1

In a split-brain scenario, only the affected partition on the secondary site and its access paths become unavailable with policy-based HA. Local non-HA volumes on the secondary site critically remain accessible. This is a significant advantage over HyperSwap, where an entire secondary site goes offline during a split-brain, potentially impacting all data stored there. See Figure 3-8 on page 40.

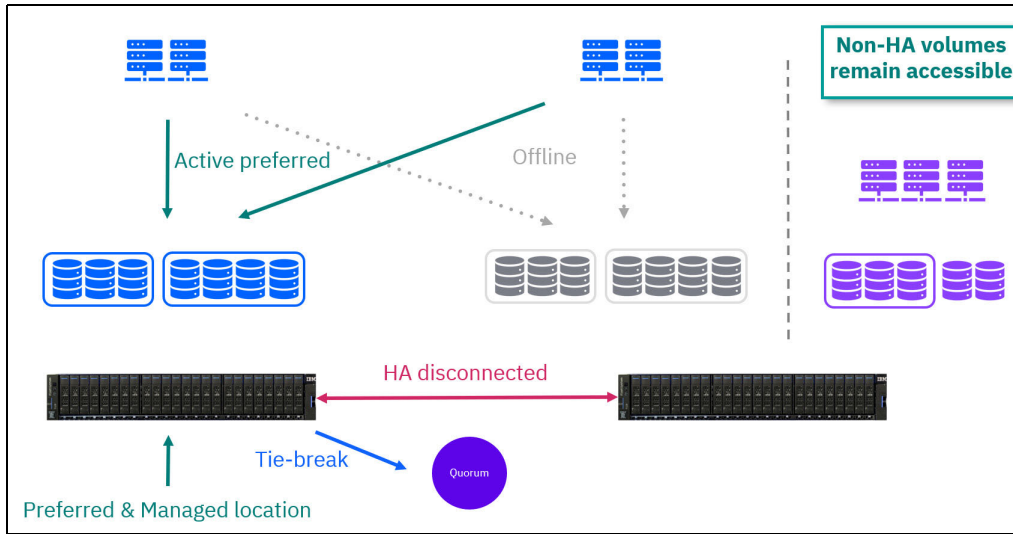


Figure 3-8 HA storage partitions - split brain scenario 2

During a failover event in policy-based HA, the multipathing driver on the host will automatically switch the paths to the active partition from the secondary site to the preferred site. To ensure optimal performance during a failover, the public SAN must have sufficient bandwidth to handle the additional workload from the non-preferred site. See Figure 3-9.

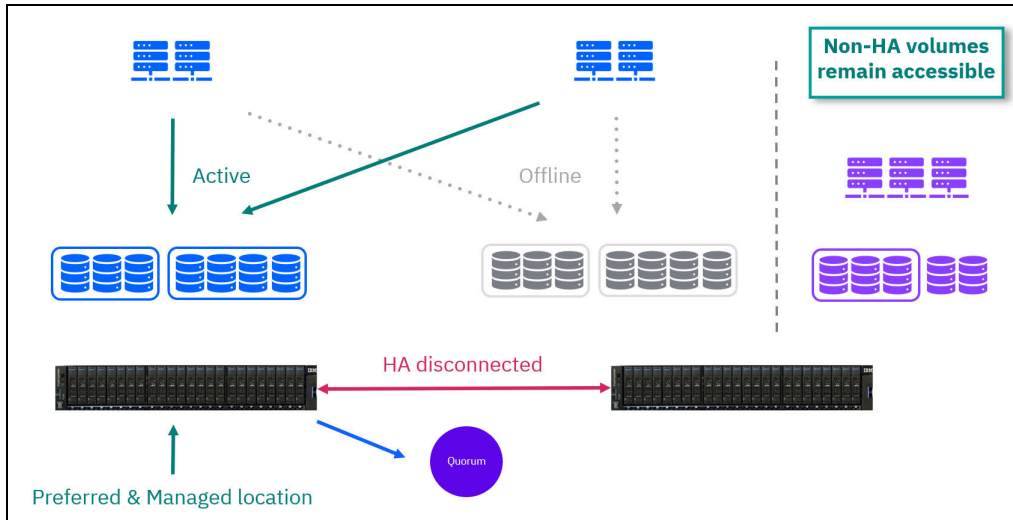


Figure 3-9 HA storage partitions - split brain scenario 3

If the connection between sites is re-established, resynchronization will automatically begin. Change volumes are used automatically during synchronization to maintain a consistent data copy at the secondary site.

In case of cascading failures where HA was not fully re-established, disaster recovery like access can be enabled to the most recently synchronized copy of volumes within the partition. HA is only established once all volumes in the partition are synchronized. HA becomes available after synchronization completes for all volume groups, and the partition reaches the established state. This typically occurs immediately after synchronization finishes.

Management of the partition follows the active management system, which may or may not be the same as the preferred system depending on the failure scenario. Following a failure, management and data access will be routed through only one of the FlashSystem units until HA is re-established.

Note: If policy-based HA detects issues that could compromise high availability, it will automatically suspend the affected partnership for 15 minutes. This 15-minute window allows the systems and network links to stabilize before attempting to re-establish HA. Following the 15-minute suspension, a resynchronization process will occur to ensure data consistency before HA resume.

While automatic restart occurs after 15 minutes, you can manually initiate the partnership restart on both FlashSystem units to potentially expedite HA re-establishment.

HA storage partitions - with preferred workload in different locations

In scenarios with two data centers and active workloads on each site, defining different partitions for each site is considered best practice for policy-based HA. See Figure 3-10.

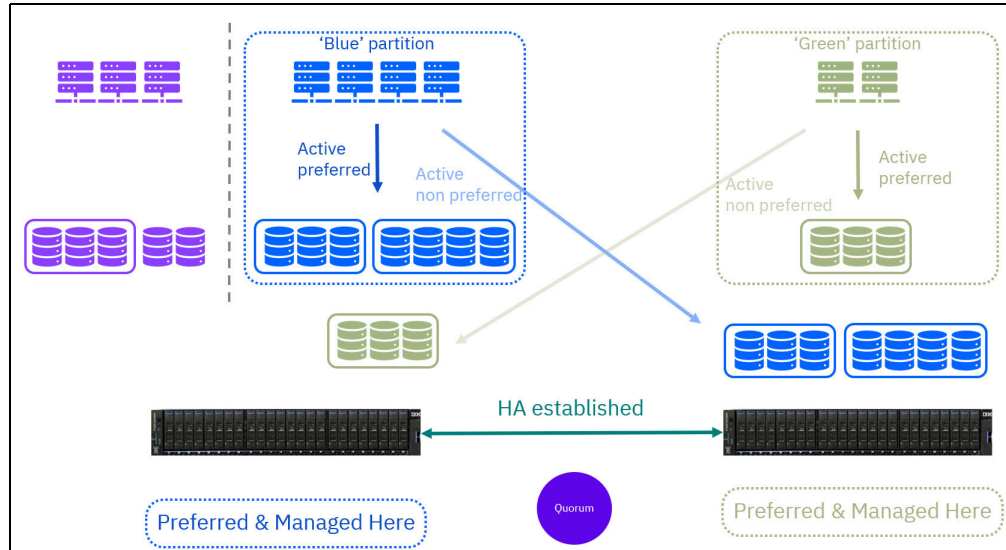


Figure 3-10 HA storage partitions - normal running, asymmetric preferences

Split brain scenario with preferred workload in different locations

Policy-based HA offers a significant advantage over HyperSwap when the connection between HA systems is lost, especially with preferred partitions defined:

- ▶ **Per-partition quorum decisions:** Unlike HyperSwap, policy-based HA relies on per-partition quorums. Each partition independently determines its active management system based on quorum votes within itself.
- ▶ **Preferred site advantage:** If the connection is lost and the preferred site for a partition is available with a majority of quorum votes, that partition remains managed by the preferred site. This avoids unnecessary failovers and keeps data accessible from the preferred location.
- ▶ **Remote partition handling:** Partitions associated with the unavailable site will likely become offline due to the lack of communication and quorum.
- ▶ **Non-HA disk accessibility:** Even during a connection loss, non-HA volumes on the available FlashSystem typically remain accessible, ensuring continued access to critical data not part of an HA policy.

See Figure 3-11 on page 42.

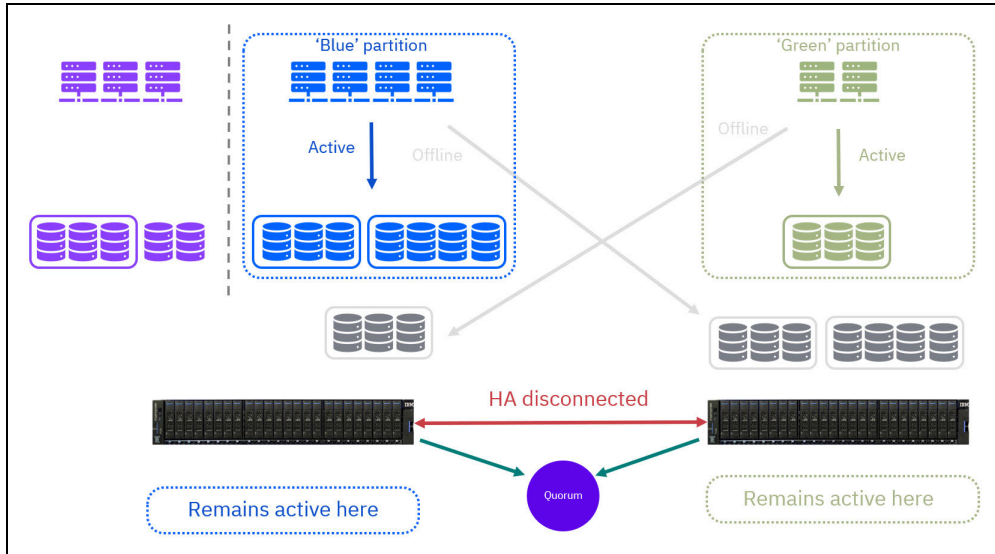


Figure 3-11 HA storage partitions - quorum decision per partition

HA disconnected - System 1 cannot get to the quorum

Let us explore what happens in policy-based HA if the preferred site loses connection to the quorum, a critical decision-making component for high availability. See Figure 3-12.

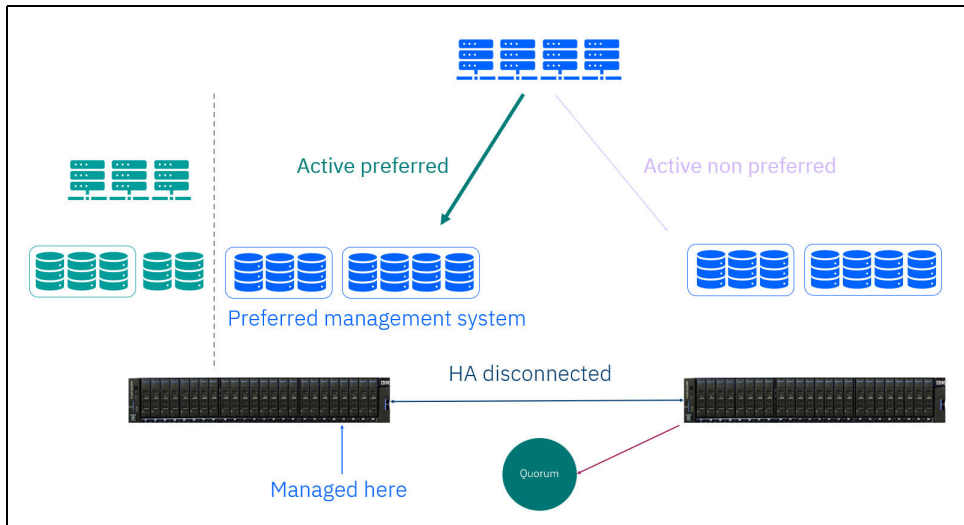


Figure 3-12 HA disconnected - System 1 cannot get to the quorum

1. In the event of a failure at the primary (preferred) site, policy-based HA automatically initiates a failover. Here is what happens:
 - **Secondary site takes over:** The designated secondary site assumes control, ensuring minimal disruption to your applications and data access.
 - **Preferred site goes offline:** The primary site is brought offline to prevent potential data inconsistencies during the failover process. See Figure 3-13 on page 43.

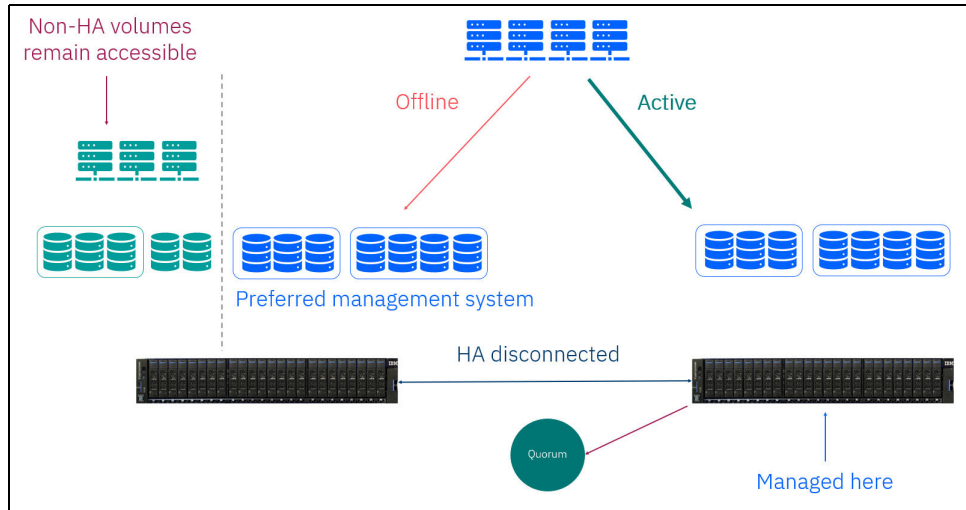


Figure 3-13 Operation while in fault state

Upon reconnection, policy-based HA captures a snapshot of the "change volume" - a temporary storage area that holds modifications during the outage. Leveraging the change volume snapshot, policy-based HA performs an efficient resynchronization to ensure both sites are back in sync. Once resynchronization is complete, policy-based HA seamlessly re-establishes high availability, ensuring your data remains protected.

See Figure 3-14.

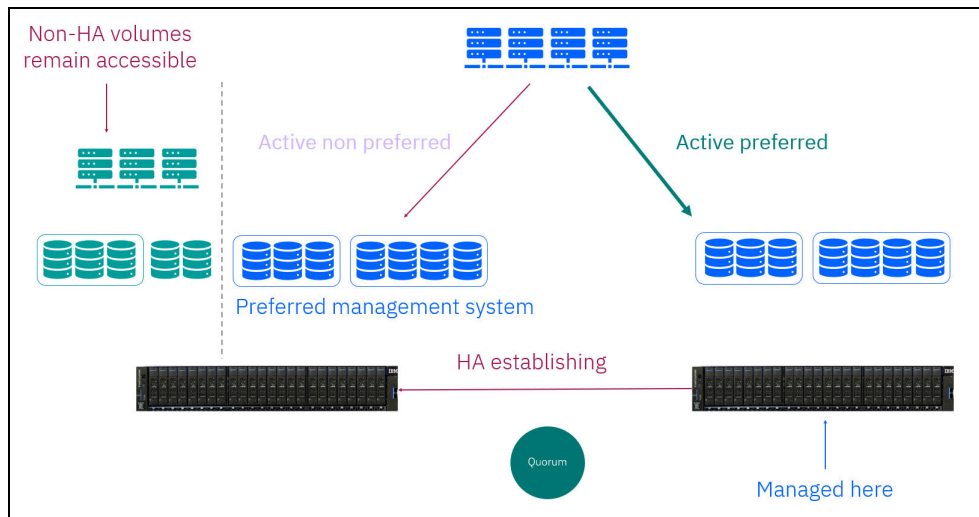


Figure 3-14 Problem fixed - Resynchronization

2. After HA is re-established, the paths will fail back to the preferred site. See Figure 3-15 on page 44.

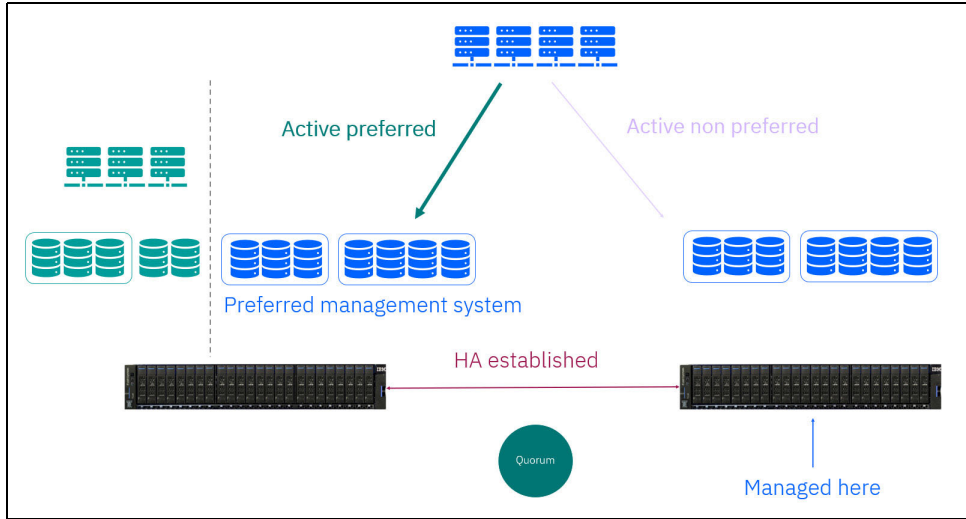


Figure 3-15 Paths return to normal

3. Also, the management will fallback to the preferred site. See Figure 3-16.

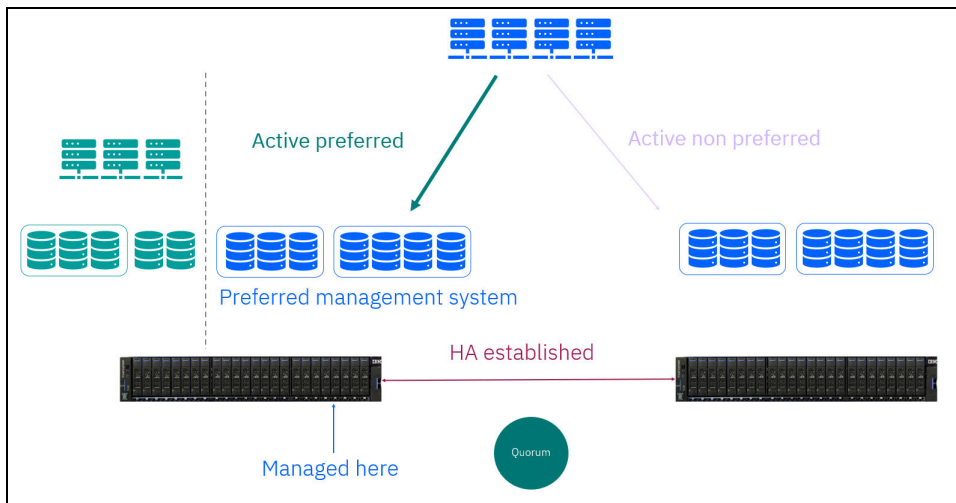


Figure 3-16 Management returns to normal

4. Now, the system runs again in the desired preferred HA configuration.

3.3.1 Comparison of policy-based HA with SVC stretched cluster and HyperSwap

This document focuses on the current capabilities of policy-based HA. While some limitations exist in this initial version, continuous development is underway to address them and introduce even more advanced features. Table 3-1 on page 45 compares policy-based HA with SVC stretched cluster and HyperSwap.

Table 3-1 Comparing policy-based HA with SVC stretched cluster and HyperSwap

	Policy-based HA	HyperSwap	Enhanced Stretched Cluster
Supported on	SVC and NVMe products. 1 I/O group systems	2+ I/O group systems	2+ I/O group SVC only
Maximum HA capacity	Up to 4PiB	Up to 2 PiB per pair of I/O groups, 4PiB per system	40 PiB per I/O group (SV3), 1PiB (SV2, SA2)
Maximum HA volume count	32,500	2,000	7,932
Non-HA volumes	Unaffected by HA problems	Link problems cause offline volumes	Link problems cause offline volumes
Protects against outage of entire system	Yes	No	No
Protects (mutual) consistency during re-synchronization	Yes	Yes (mutual consistency with consistency groups)	No
Host interoperability support	Limited (Refer to IBM Documentation)	Full	Full
HA snapshots (including Safeguarded)	No*	No*	Yes
3-site support	Statement of direction	Yes - externally orchestrated asynchronous	Synchronous with Metro Mirror, asynchronous with policy-based replication
Volume resize support	No	Expand thin volumes	Expand/shrink
Quorum	App only, preferred-site behavior per storage partition	App or controller, default/preferred/winner per system	App or controller, default/preferred/winner per system
Maximum nodes per site	2 (1 I/O group)	4 (2 I/O groups)	4 (half of each I/O group)
Mixed hardware models	Yes, unrestricted	Limited (must cluster)	Limited (must cluster)
Mixed software levels	Yes (future, within range)	No	No
Licensing	Remote Copy, included in LMC	Remote copy, included in LMC	Base

* Volume Group Snapshots would be taken on both sides rather than making the snapshots HA.



4

Implementing policy-based replication

In this section we demonstrate how to implement policy-based replication for the Storage Virtualize 8.7 solutions. We also show how to stop and reverse the mirror direction in case of failure of the production system.

This chapter has the following sections:

- ▶ “Implementing policy-based replication” on page 48
- ▶ “Converting Global Mirror to policy-based replication” on page 62

4.1 Implementing policy-based replication

Policy-based replication can help the setup, administration, and oversight of replication by employing volume groups and replication policies. This approach offers a simplified means of configuring, managing, and monitoring replication between two systems.

Policy-based replication brings several key advantages to asynchronous replication:

- ▶ PBR uses volume groups to ensure that all volumes are replicated based on the assigned policy.
- ▶ By eliminating the need to manage relationships and change volumes manually, policy-based replication can help simplify the overall administration process.
- ▶ The feature automatically manages provisioning on the remote system, reducing the administrative burden.
- ▶ During a site failover, policy-based replication can enable easier visualization of the replication process, which can facilitate effective management and decision-making.
- ▶ IBM Storage Virtualize Version 8.5.2 and later provides asynchronous replication over Fibre Channel and IP partnerships.
- ▶ Automatic notifications are generated when the recovery point objective exceeds the specified threshold, ensuring timely awareness of any deviations.
- ▶ Policy-based replication provides easy-to-understand status updates and alerts regarding the overall health of the replication process, enhancing monitoring and troubleshooting capabilities.

4.1.1 Configuring policy-based replication using GUI

To demonstrate the full functionality and capabilities of policy-based replication, this section shows the roles of the storage architect and the storage administrator.

Figure 4-1 on page 48 shows the topology for the systems we are configuring.

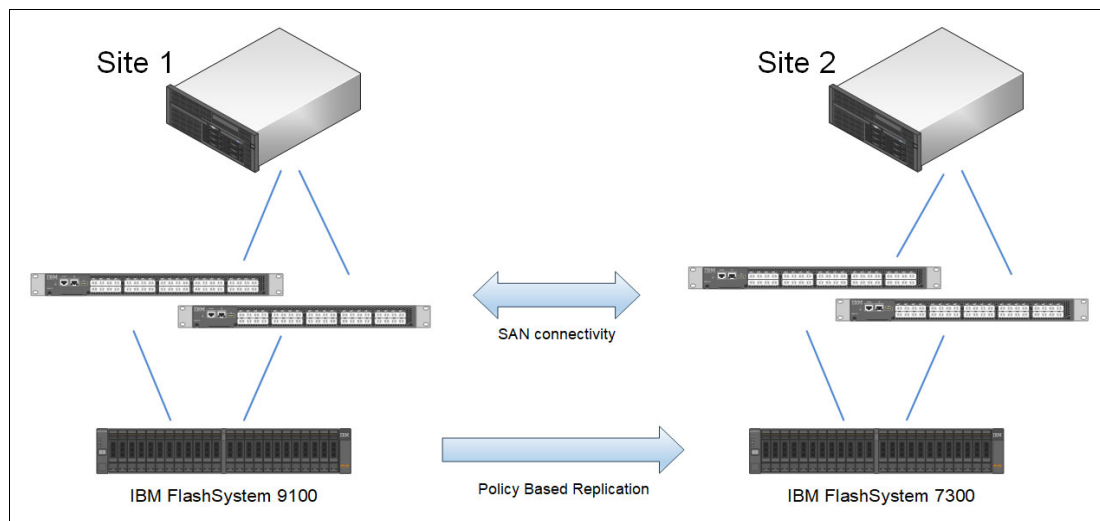


Figure 4-1 Topology for our policy-based replication setup

All aspects of replication can be managed using the GUI.

Upon creating the first partnership for policy-based replication, the management GUI guides you through the steps. The main steps include the following:

1. Completing the partnership setup by creating the partnership from the remote system.
2. Linking pools between systems, optionally using provisioning policies on each pool.
3. Creating a replication policy.
4. Creating a volume group and assigning a replication policy to the group.
5. Creating new volumes, or adding existing volumes, to the group.

The monitoring and management of replication is performed from the Volume Groups page

Note: In the example below we are demonstrating the setup of policy-based replication on two connected FlashSystem systems. Policy-based replication can also be configured on SAN Volume Controller (SVC).

The systems below are SAN-zoned together with dedicated ports for node-node communication. Other connectivity options include high speed ethernet networking.

Check code levels on both connecting systems

To ensure compatibility for policy-based replication, use the GUI to verify that both systems you want to replicate between have the preferred Storage Virtualize code version. In our example, we make sure both systems have version 8.7.0.0. Go to **Settings**, and select **Update System**, as shown in Figure 4-2.

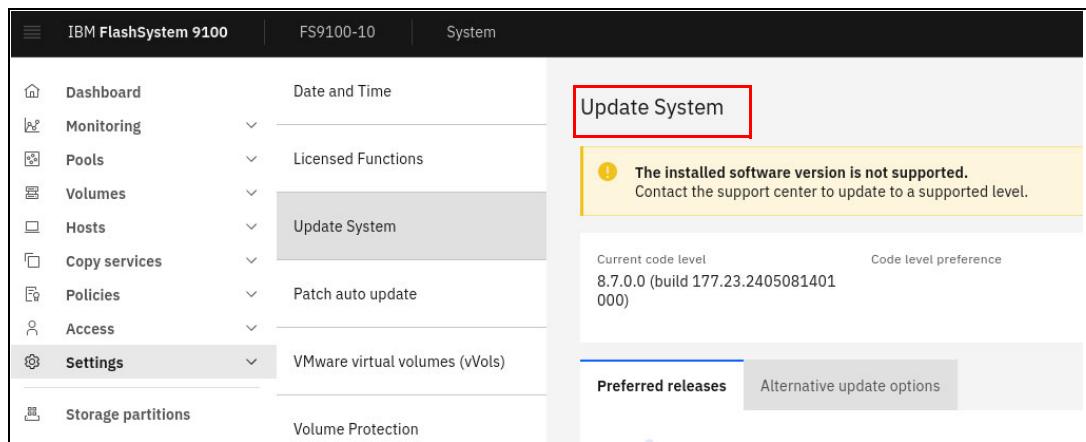


Figure 4-2 Verify software version

Define partnership for replication

Before starting the configuration process, it is important to ensure that a partnership has been established between the two FlashSystem boxes. This partnership is used for data replication and synchronization. To initiate the partnership, perform the initial system setup on both systems to be partners. This setup involves establishing a partnership between the systems, which can be done using either FC or IP connectivity.

Additionally, as part of the partnership setup, a certificate exchange must be performed. This exchange allows each system to have the necessary configuration access to the other system using the REST API.

Note: A prerequisite for creating a partnership via SAN-zoning, is that the two systems are correctly zoned together with dedicated ISL-links for node to node traffic.

Creating a partnership

To create the partnership between systems, complete these steps:

1. On the first system, select **Copy Services** and then **Partnerships**.
2. Click on **Create Partnership**.
3. Enter the type of partnership (FCP, IP TCP or IP RDMA).
4. Select the remote system.
5. Ensure that the Use **Policy-Based Replication** checkbox is selected.
6. Enter the value, in megabits per second (Mbps), for the total bandwidth available between the two systems that can be used for replication. If all the bandwidth is available for policy-based replication, ensure the background copy rate (%) is set to 100.
7. Click **Create**.

The Create Partnership wizard is shown in Figure 4-3 on page 50.

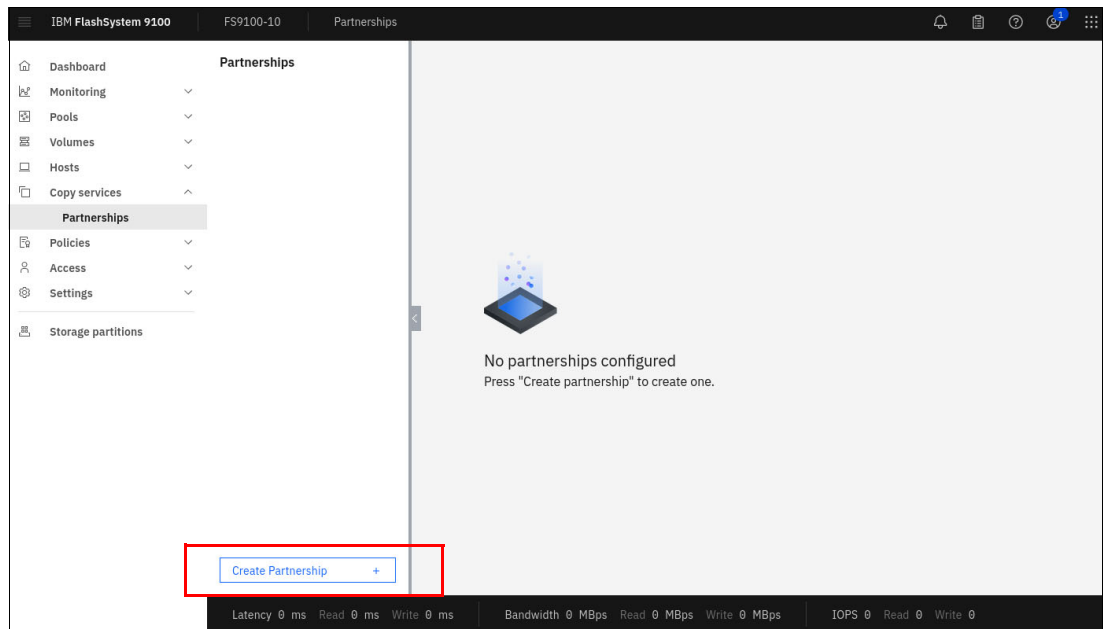


Figure 4-3 Create partnership

8. The Create Partnership wizard options are shown in Figure 4-4 on page 51.

Create Partnership ×

Create a partnership to establish a connection to a remote system for replication.

Type

Fibre Channel

IP (long distances using TCP)

IP (short distances using RDMA)

Partner system name ⓘ

FS7300-2 ▾

Use policy-based replication ⓘ

View certificate

⚠ The remote system is using a CA-signed certificate. Review the certificate to ensure that it matches what you expect

Certificate from 192.168.61.184

[See details](#) ▾

Link specification

Link bandwidth is available between systems, in megabits per second (Mbps)

Link Bandwidth (Mbps) ⓘ Background Copy Rate (%) ⓘ

32000 100

Cancel Create

Figure 4-4 Create Partnership window

The partnership has to be created from both sides production and recovery systems.

- After the partnership is created, the **CopyServices** → **Partnership** menu looks like shown in Figure 4-5.

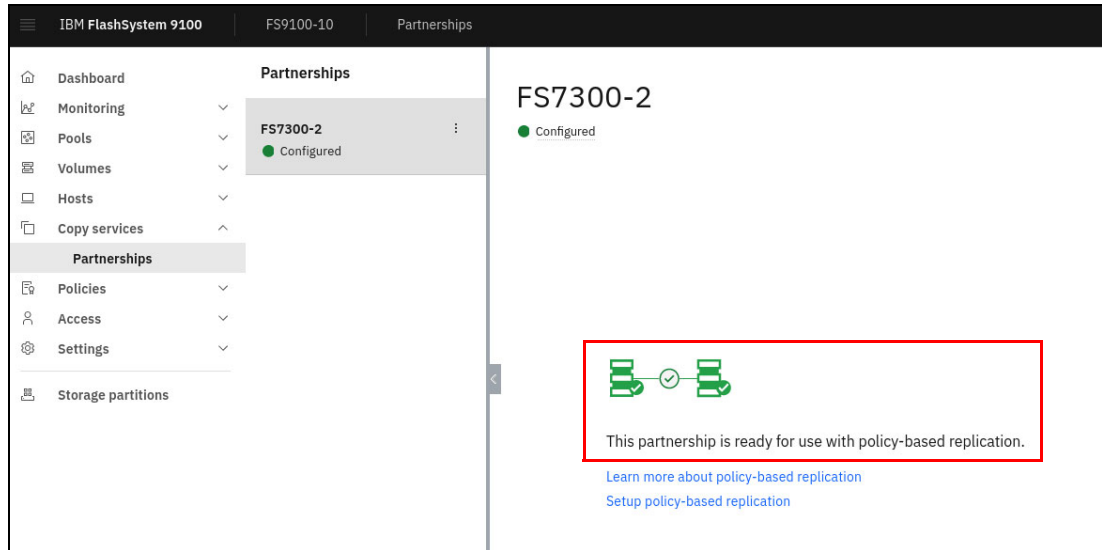


Figure 4-5 Partnership ready for policy-based replication

Since the two systems are now partners the above view will look the same from both sides.

Setup policy-based replication wizard

On the window **Copy Services** → **Partnerships** click **Setup Policy-based replication**. The resulting window, after creating the partnership, looks like shown in Figure 4-6 on page 52.

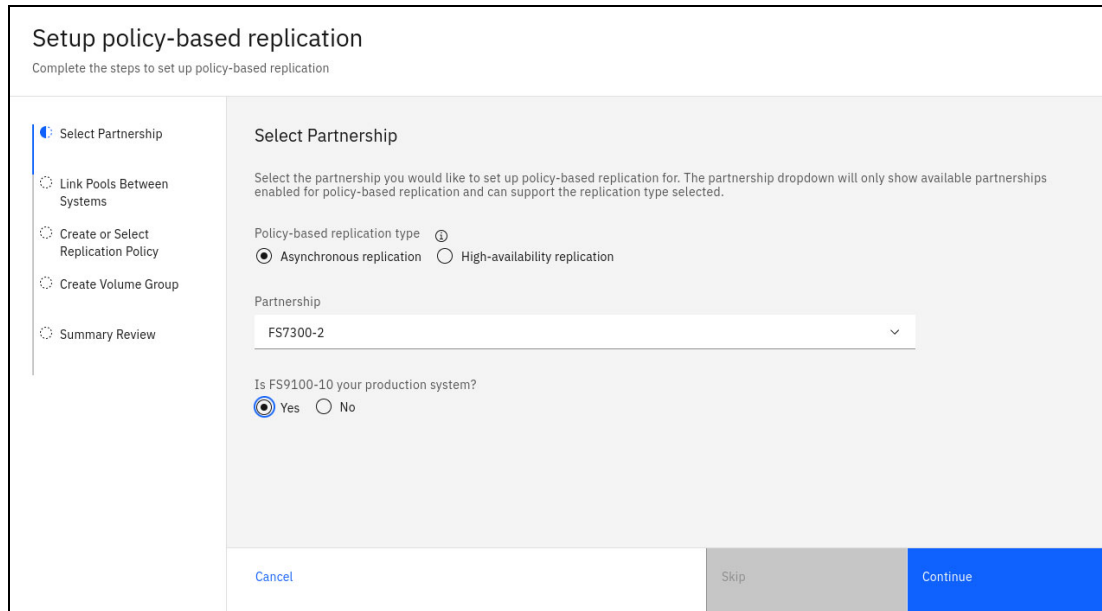


Figure 4-6 Setup policy-based replication

- a. Select the replication type **Asynchronous** or **High-Availability**. We select **Asynchronous** replication.
- b. Select the system to replicate with; we have only one option, the **FS7300-2**.
- c. Confirm that this system is your production system, that is we are mirroring from this system to the recovery system. If you select **No**, you will be prompted to make the changes from the partner system instead.

d. Click **Continue**.

Link storage pools together

To establish a linked pool in policy-based replication, the storage pool links determine where the recovery copy of a volume is stored on the recovery system based on the production volume pool. It is necessary to have links between the pools on the production and recovery systems when using policy-based replication.

At least one linked pool is required on each system for policy-based replication to function properly. There are two approaches to creating linked pools. The first option involves creating pools on each storage system and then linking them together manually.

Alternatively, the Setup Policy-Based Replication wizard shown below guides you through the processes involved.

Select the storage pools to link together. We only have single pools on each system and we select these as shown in Figure 4-7. Click **Link Pools** to proceed.

Setup policy-based replication
Complete the steps to set up policy-based replication

Link Pools Between Systems
Select the local pool that you want to link for replicating between volumes. You must link at least one pool.

Systems to link

Remote system name	Local system name
FS7300-2	FS9100-10

Pools to link

Select remote pool to link	Local pool
StandardPool ▲	StandardPool

Provisioning policy

Remote provisioning policy	Select local pool provisioning policy
capacity_optimized <small>Thin-provisioned, not deduplicated</small>	capacity_optimized

Buttons: Cancel, Skip, Link Pools

Figure 4-7 Set up policy-based replication - Link pools

The wizard will prompt the user for:

- a. Systems to link:
 - We select our local system **FS9100-10** and the remote system **FS7300-2**.
- b. Pools to link:
 - We select the single available pools on each system.
 - We get a warning because one system has its storage pool encrypted which is not recommended. Encryption should be enabled on both or none.
- c. Provisioning policy:
 - We apply a provisioning policy to the storage pools. We select **capacity_optimized** to get thin provisioning.
 - The wizard will direct the user to the remote system to enable a provisioning policy to that systems storage pool.

The Link Pools Between systems wizard shown above will initially provide a link to the recovery system on which the user will be directed to the Pools menu. The user can then right-click the pool and add the link. In Figure 4-8 on page 54 we show how to add or remove pool links directly on the target system.

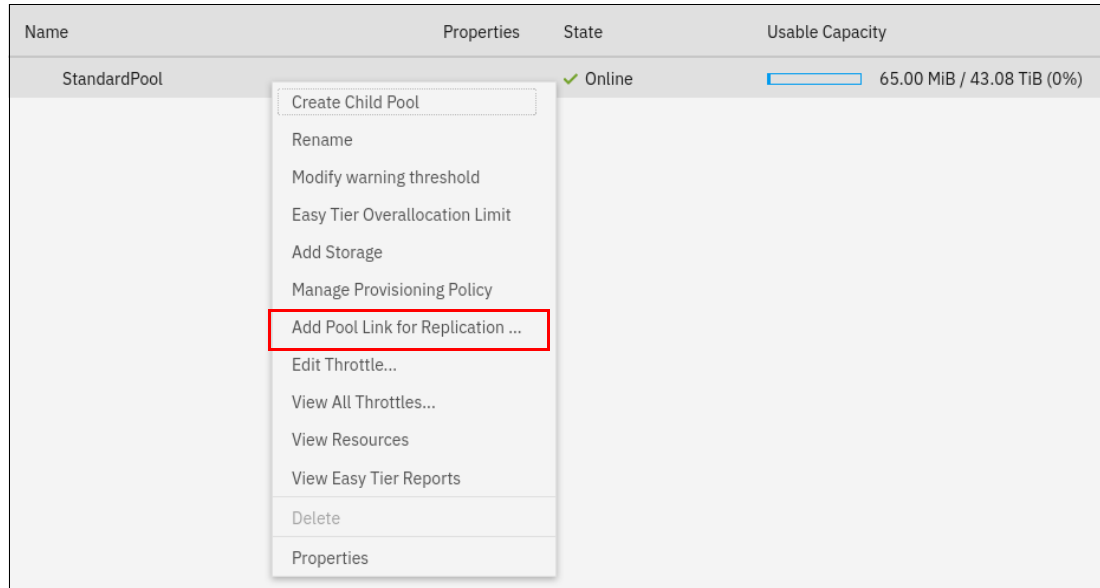


Figure 4-8 Link pools on the recovery system

Once the replication is active the pool links cannot be modified.

Create replication policy

Replication policies play a crucial role in policy-based replication because they define the configuration of replication between I/O groups in partnered systems and define how it should be applied to volume groups and their associated volumes. These policies are established between two fully configured and partnered systems capable of policy-based replication.

A replication policy can be linked to multiple volume groups. However, each volume group can have a maximum of one replication policy associated with it.

In Figure 4-9 on page 55 we show how to create the replication policy. Click **Create replication policy** to proceed.

Setup policy-based replication

Complete the steps to set up policy-based replication

- Select Partnership
- Link Pools Between Systems
- Create or Select Replication Policy
- Create Volume Group
- Summary Review

Create replication policy

You can create a replication policy or select an existing one to define how volume groups are replicated between systems. When you create a replication policy on this system, the policy will automatically be created on the other system.

Create new policy Use existing policy

Replication Policy
A replication policy cannot be changed after it is created. If you want to use different settings in a policy, you must create a new replication policy and assign the new policy to your volume groups.

Name

Topology

Location 1	Location 2
System <input type="text" value="FS9100-10"/>	System <input type="text" value="FS7300-2"/>

Recovery point objective (RPO)
Specify the desired recovery point objective for the policy. An alert will be sent if the recovery point exceeds this value.

Send an alert if data on the recovery copy is older than: min

Figure 4-9 Create replication policy

Creating a volume group

To deploy and manage replications in policy-based replication, volume groups and replication policies are used. After a replication policy is created, the wizard guides us to create a volume group. This process ensures consistent replication by replicating the source volumes as a group to the recovery system.

The recovery copies of volume groups are immutable, meaning they cannot be modified or altered. Policy-based replication greatly simplifies the configuration, management, and monitoring of replication between two systems.

To create a volume group, the Create Policy-Based Replication wizard now prompts us to name a new volume group as shown in Figure 4-10 on page 56. Click **Create Volume Group** to proceed.

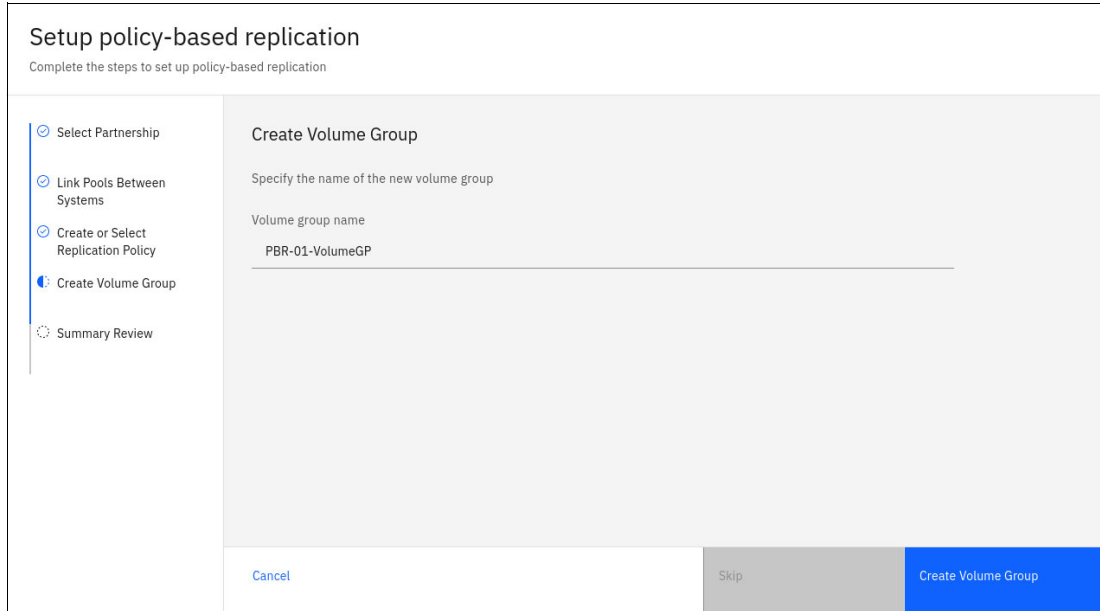


Figure 4-10 Create Volume Group

After the various steps in the Setup Policy-Based Replication wizard we now see a summary of the changes as shown in Figure 4-11. Click **Go to Volumes** to proceed.

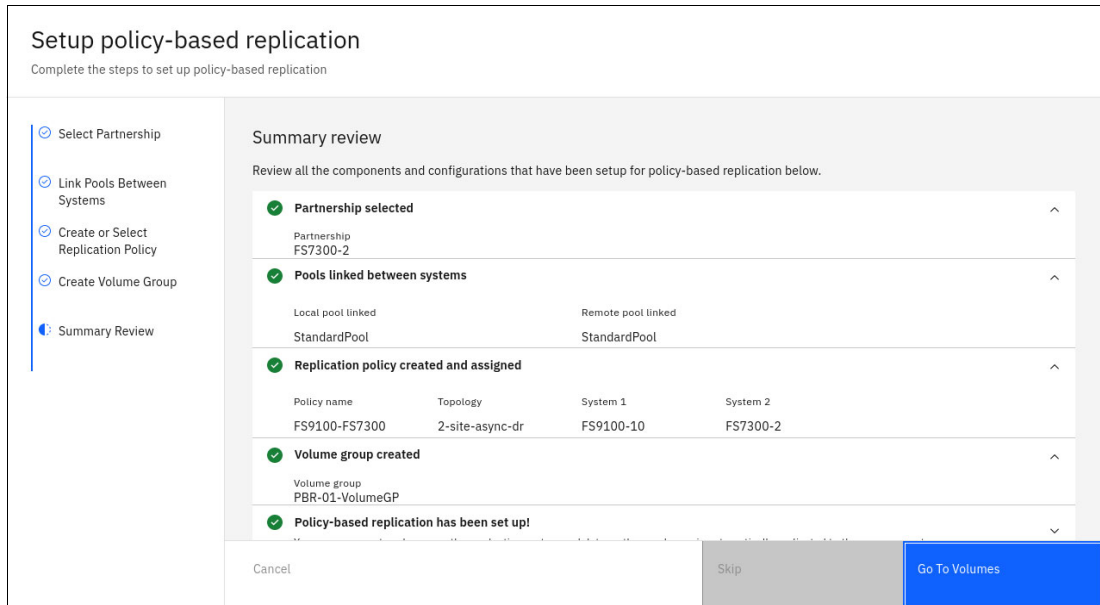


Figure 4-11 Policy-based replication create summary

The Setup Policy-Based Replication wizard completes and directs the user to the volumes menu.

Add volumes to volume group

Next step is to create volumes or add existing to our volume group. Figure 4-12 shows the newly created volume group, which is currently empty.

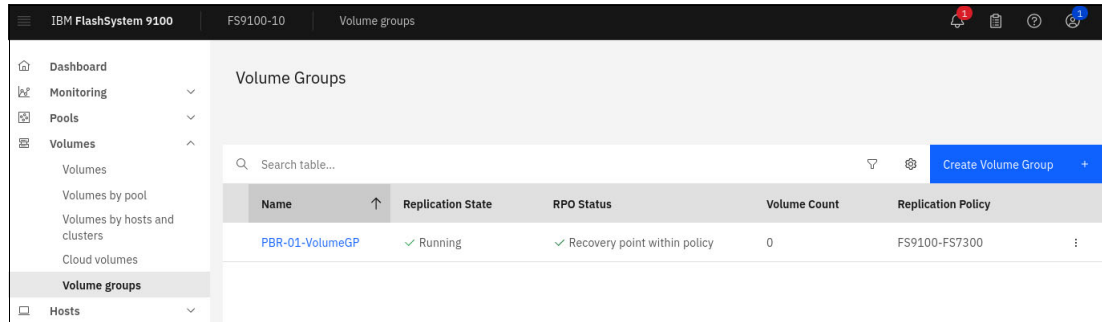


Figure 4-12 Volume Group created including no volumes

1. Click the newly created Volume Group and click **Actions** → **Create New Volumes** or **Add Existing Volumes**.
2. We select **Create New Volumes**, as shown in Figure 4-13 on page 57.

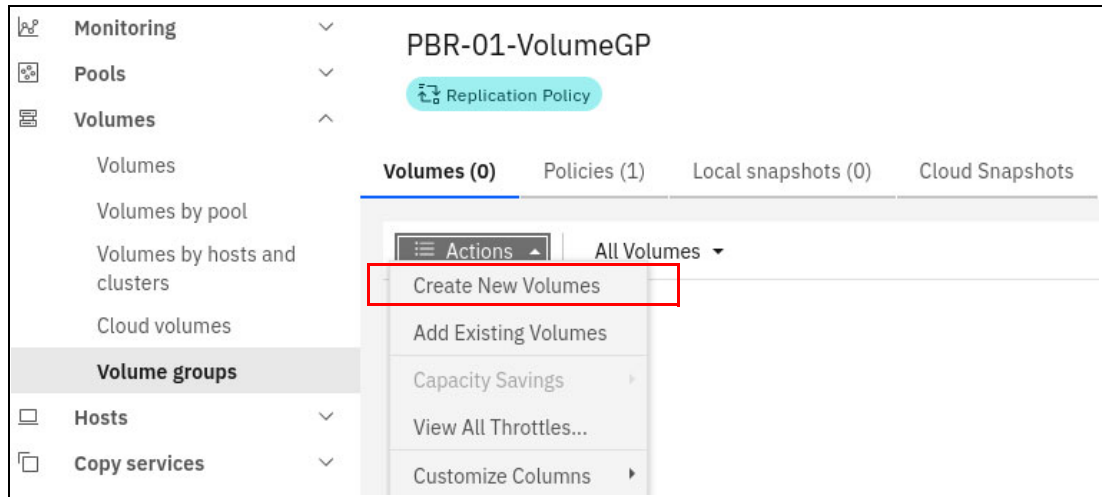


Figure 4-13 Create new volumes within Volume Group

3. Since we are creating new volumes a wizard to create volumes appears. From here we select in which storage pool the volumes should be created. We only have a single pool, **StandardPool**, to choose from. Figure 4-14 shows the Create Volumes wizard. Select the pool and click **Define Volume Properties**.

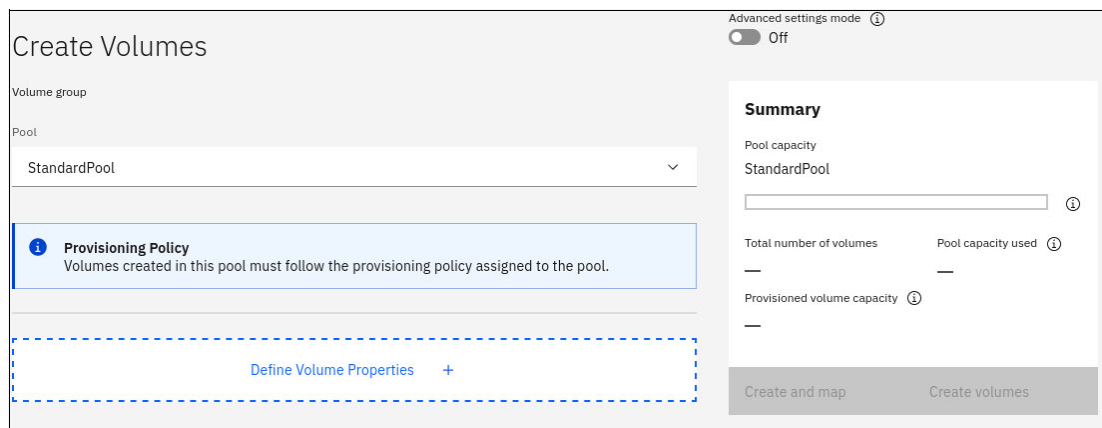


Figure 4-14 Create Volumes wizard begins

- For the purpose of this demonstration we create four small 50 GB volumes. We provide a name, number and size. The provisioning policy Capacity Optimized is active on the storage pool, so all volumes created in this storage pool will be thin provisioned.
- Figure 4-15 on page 58 shows the Define Volume Properties page. Click **Save** to continue.

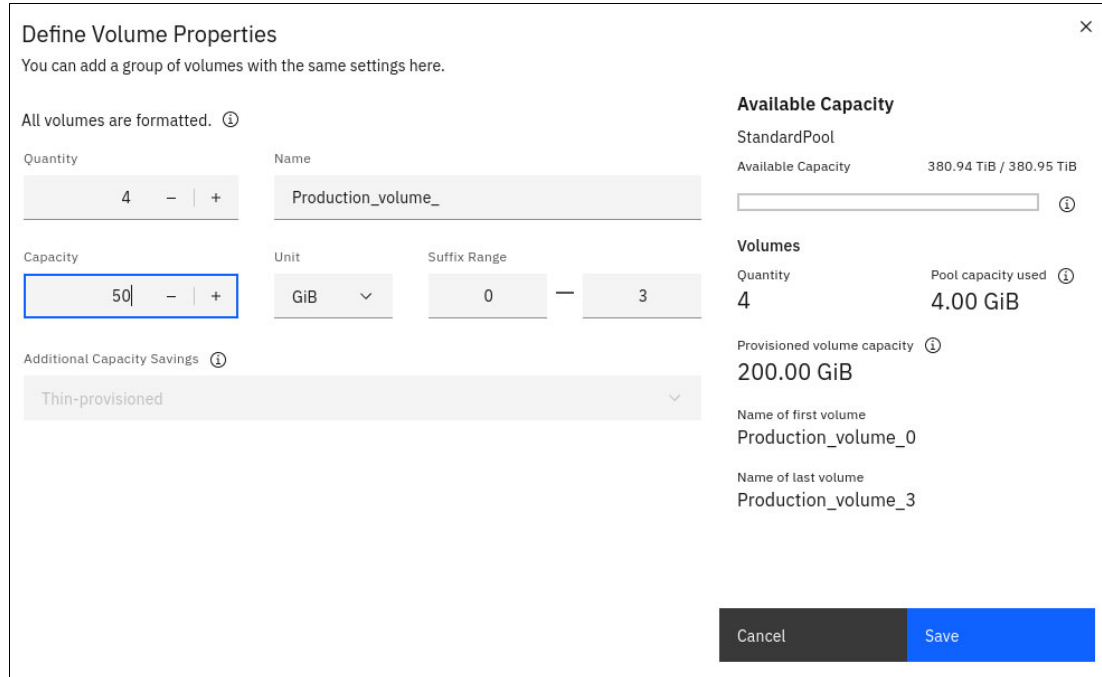


Figure 4-15 Create four volumes each 50 GB

We now have four volumes active in our volume group, and these are replicating to the recovery system.

The content of the volume group is shown in Figure 4-16.

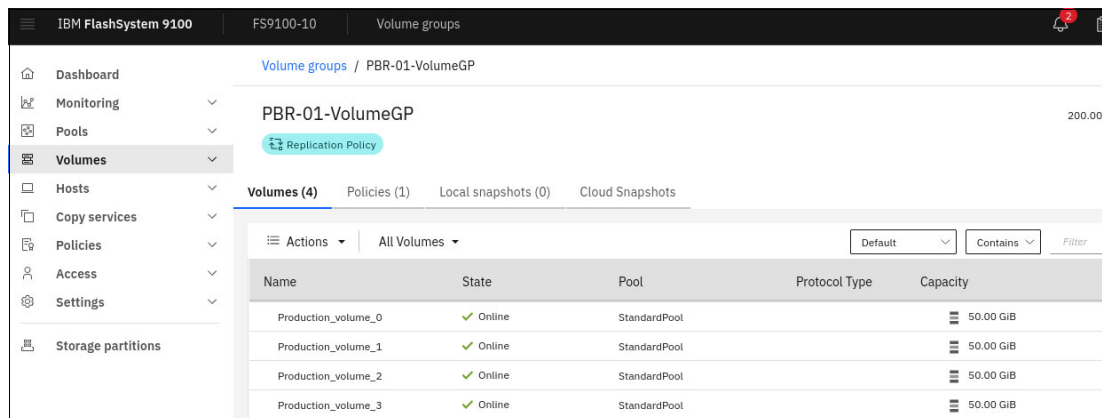


Figure 4-16 Four volumes created and added to the Volume group

Our configuration has been completed and volumes within the volume group PBR-01-VolumeGP are now copying from FS9100 to FS7300.

- To check the status on the recovery system go to the window **Volumes** → **Volume Groups**. We can see that the volume group created on the production system is now

replicated to the remote system. During initial replication, you can monitor the progress in Figure 4-17 on page 59.

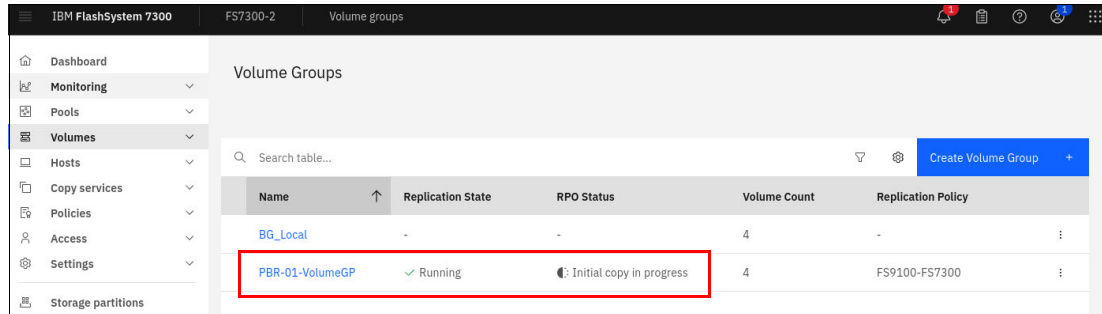


Figure 4-17 Initial copy ongoing

- By entering the defined volume group and clicking on the tab **Policies** we get to see a status of the current replication policy, as shown in Figure 4-18 on page 59.

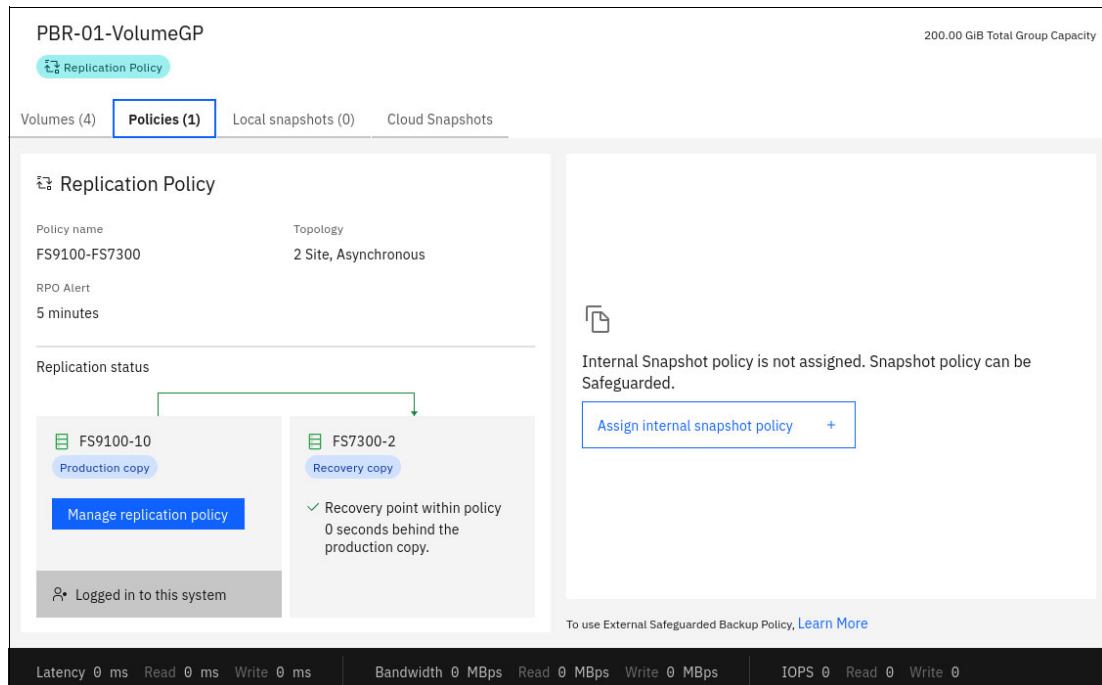


Figure 4-18 Volume group policy status

- You now have the option to click **Manage replication policy**. This is where you get the option to remove the replication policy and hence delete the current replication.
 If you are on the recovery system, the same window gives the option to enable access to the recovery volumes.

Note: The newly created volumes can be mapped to one or more hosts from the Volumes menu.

Using targets of replicated volumes

While the volumes on the recovery system can be mapped to hosts, it is important to note that these volumes remain in an offline state during the replication relation. However, there is an

option to **Enable independent access** from the secondary site. This can be done in the Policy tab of the volume group in the secondary site, as shown in Figure 4-19 on page 60.

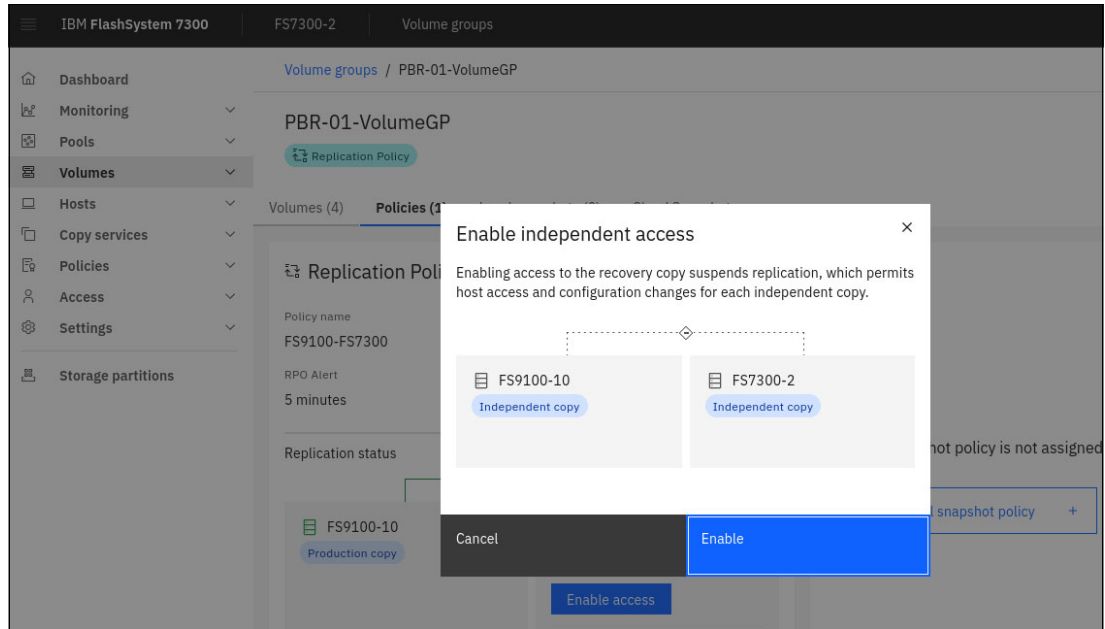


Figure 4-19 Enable volume access from recovery system

Enable access to the recovery copy suspends replication, which permits host access and configuration changes for each independent copy.

You may notice that you can now **Restart replication** and that changes made to this copy will not be replicated until replication is restarted. The replication status is shown in Figure 4-20 on page 60.

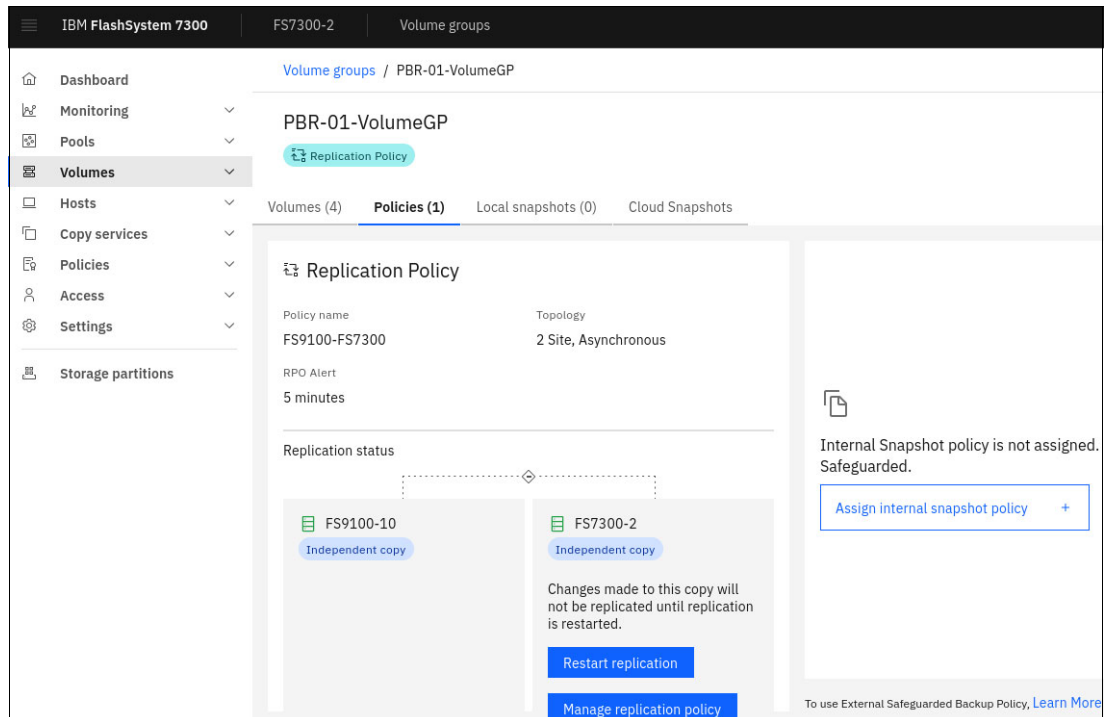


Figure 4-20 Replication stopped and access enabled to recovery volumes

Restart or reverse replication

After enabling access on the recovery system and suspending the replication, you can proceed to restart the replication while selecting the desired direction for the relation.

In Figure 4-20 on page 60, we click **Restart replication** to proceed.

Whether you want to restart the replication from the production site or to reverse the replication from the recovery site, which may now be the production site, depends on from which system you restart the replication.

Figure 4-21 shows how we restart the replication with the recovery site now as the production system.

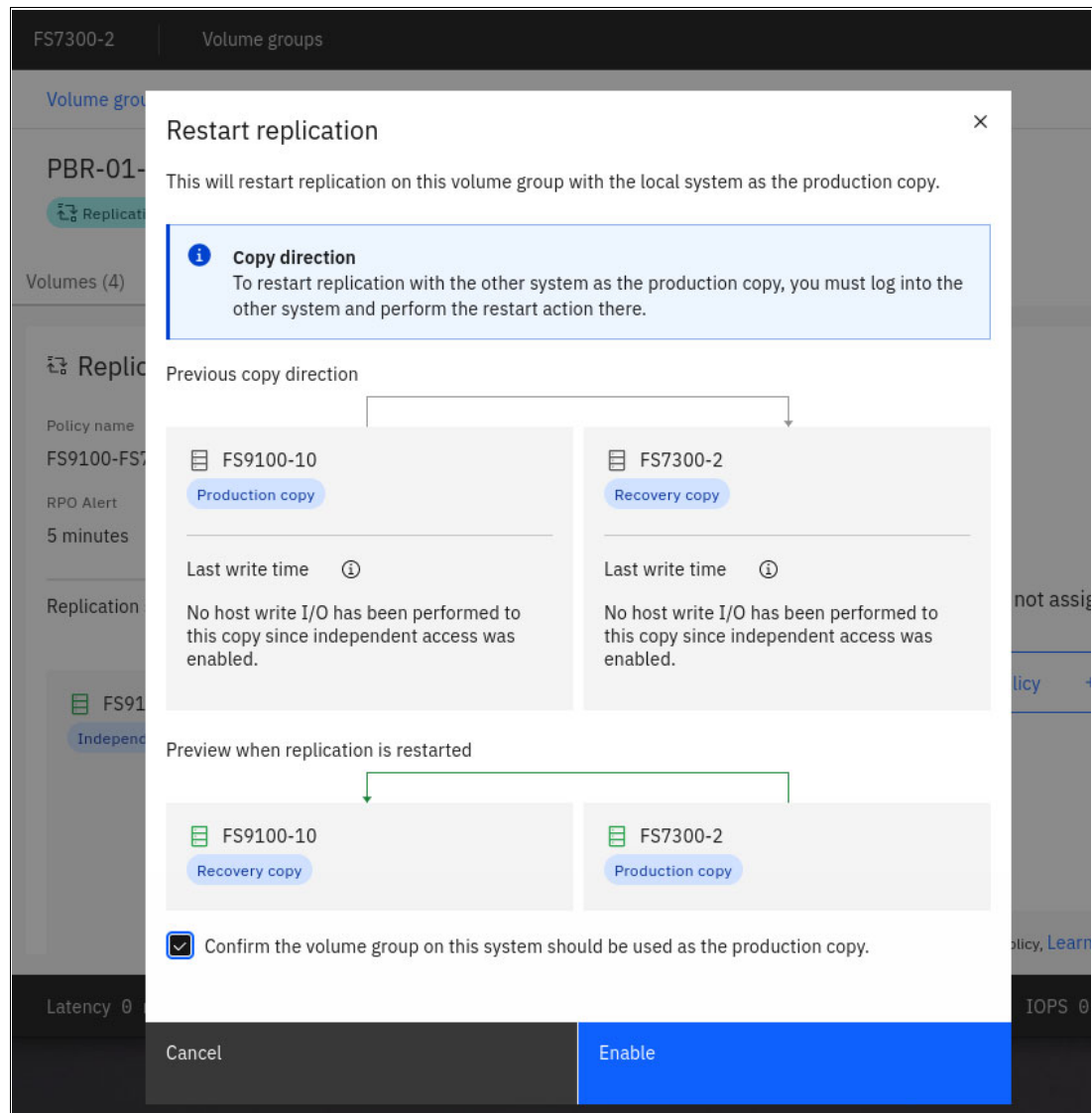


Figure 4-21 Restarting replication in reverse direction

The above action will overwrite the volumes on the FS9100-10 system which was once the production system. However the situation could be that this system has been down for some time and that the volumes on it are no longer current, because the recovery system is now functioning as the production system.

Figure 4-22 on page 62 shows that initial copy is ongoing from the FS7300-2 to the FS9100-10.

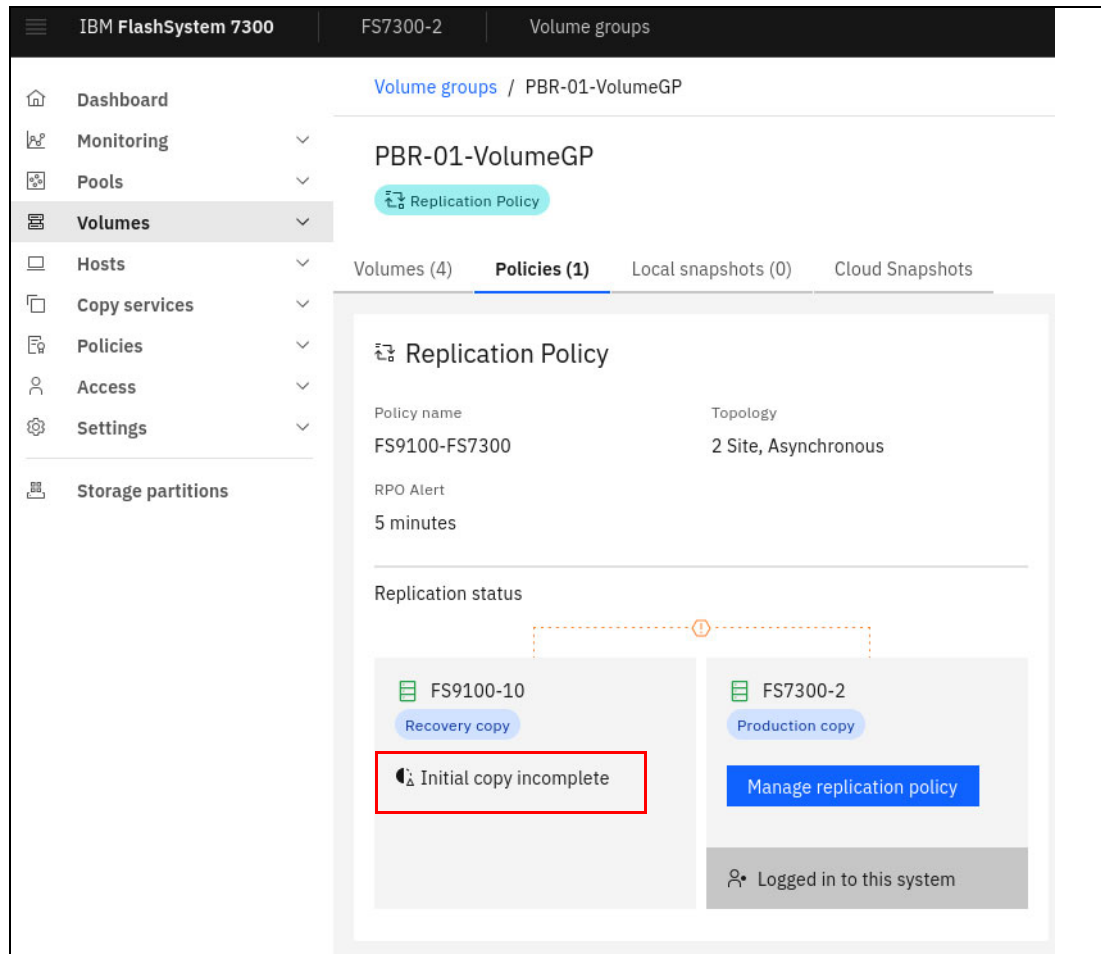


Figure 4-22 Copy direction reversed - initial copy ongoing

In the scenario above we were initially replicating from a FlashSystem 9100 to a FlashSystem 7300. In a failover situation then we might have to soon switch back to FS9100-10 being the primary because this is the normal production location.

Switching the direction of replication back to the FlashSystem 9100 would require the same actions once again, which is to log on to the FS9100-10 and click **Enable Access** on the Volume Group policy tab - just like we did in Figure 4-19 on page 60.

To switch copy direction in a safe way would require the hosts accessing the volumes, to be shut down while reversing the copy direction back to FS9100-10. So expect few minutes downtime for this.

4.2 Converting Global Mirror to policy-based replication

For IBM Storage Virtualize version 8.6.0 and 8.7.0, if you use Global Mirror for data replication between two partnered systems, it is possible to convert the existing setup to policy-based replication. During this conversion process, the remote-copy configuration can

be retained for a volume to ensure that a synchronized copy remains available on the disaster recovery (DR) system without any downtime.

Ensure that the following prerequisites are met before configuring a volume which is currently part of a Global Mirror relationship, to use policy-based replication:

- ▶ The relationship must be either Metro Mirror or Global Mirror.
- ▶ The volume being migrated must be the primary volume within the Metro Mirror or Global Mirror relationship.
- ▶ Volumes using Global Mirror can be manually migrated to use policy-based replication.
- ▶ Remote Copy features such as 3-site partnerships cannot be directly migrated to policy-based replication due to their more complex configuration requirements.
- ▶ No associated change volumes can be linked to the primary volume.
- ▶ During the migration process, do not change the direction of the remote-copy relationship or transform it into a secondary relationship.
- ▶ The primary volume cannot be designated as a recovery volume for policy-based replication.
- ▶ A volume within a Metro Mirror or Global Mirror relationship can have policy-based replication configured only if it belongs to the same I/O group specified in the replication policy. If it does not match, you can move the volume to the appropriate I/O group using the `movevdisk` command.
- ▶ If you keep a disaster recovery copy, ensure that you have enough resources available on the recovery system to accommodate both sets of copies.

Before running the `movevdisk` command to transfer a volume between I/O groups, several conditions must be satisfied:

- ▶ If the relationship is part of a consistency group, the volume cannot move between I/O groups when the policy-based replication is defined. Use the `movevdisk` command to move the volume as needed.
- ▶ The relationship state must be `consistent_synchronized`. The data in the source and target volumes of the relationship must be fully synchronized and consistent. Resolve any pending changes or discrepancies before proceeding with the volume movement.
- ▶ The relationship cannot be in a consistency group. Consistency groups are a collection of relationships that need to maintain data consistency as a group. If the relationship is part of a consistency group, it cannot be moved independently. Restrict the volume movement to relationships that are not associated with any consistency group.
- ▶ The relationship type must be Metro Mirror or Global Mirror. The `movevdisk` command is designed specifically for volumes involved in Metro Mirror or Global Mirror relationships. These relationship types enable synchronous or asynchronous replication of data between primary and secondary volumes. Only volumes associated with these types of relationships are eligible for the move operation.
- ▶ The relationship must not have a change volume associated with the primary volume. In some replication scenarios a change volume is used to track modifications made to the primary volume. If the relationship has an active change volume associated with the primary volume, the move operation cannot proceed. The change volume should be removed or detached before you start the volume transfer.
- ▶ The volume being moved must be the primary volume in the Metro Mirror or Global Mirror relationship. When you move a volume, the volume must be the primary volume in the Metro Mirror or Global Mirror relationship. The primary volume is the source volume where

the original data resides, while the secondary volume is the target for replication. Moving the primary volume ensures the appropriate replication of data to the new I/O group.

Convert Global Mirror to policy-based replication

To convert from Global Mirror (GM) replication to policy-based replication, follow these steps:

- ▶ Enable policy-based replication on the partnership:
 - Establish a mutual SSL certificate exchange between systems to enable REST API access.
 - Use the GUI for configuration of the existing partnership for policy-based replication.
- ▶ Establish linked pools between systems and assign necessary provisioning policies:
 - Linked pools can be established from either system.
 - The GUI provides a method for configuring pool links and assigning provisioning policies.
- ▶ Create replication policies:
 - Depending on the current configuration, it might be necessary to create multiple replication policies. For instance, if Global Mirror with Change Volumes (GMCV) is used with different cycling times, different recovery point objectives can be specified for various sets of volumes.
- ▶ Create volume groups and assign replication policies:
 - Ensure each consistency group has a corresponding volume group.
 - For policy-based replication, volumes must be placed within volume groups, so one or more volume groups must be created for any previously independent relationships.
- ▶ Move volumes into volume groups:
 - Prior to moving volumes into a volume group with a replication policy, remove the Remote Copy configuration from the volumes.
 - Move volumes into volume groups and allow sufficient time for the initial synchronization process to complete.

The above can be done manually or via the Setup Policy-Based Replication wizard as shown in Figure 4-26 on page 67.

Convert Global Mirror to policy-based replication using the GUI

To convert from Global Mirror replication to policy-based replication for replicated volumes using GUI, follow these steps:

1. Update the existing partnership to support policy-based replication. Access the local system's management interface and navigate to **Copy Services** → **Partnerships and Remote Copy**, as shown in Figure 4-23.

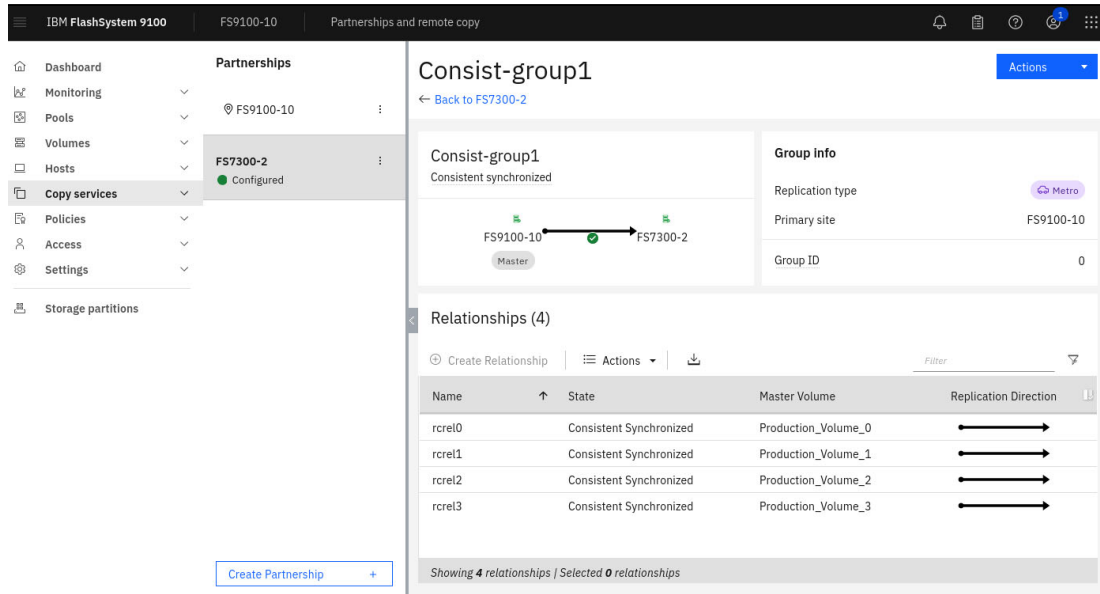


Figure 4-23 Update the existing partnership

- Choose the current Remote Copy partnership from the left navigation and select **Actions** → **Partnership Properties**, as shown in Figure 4-24.

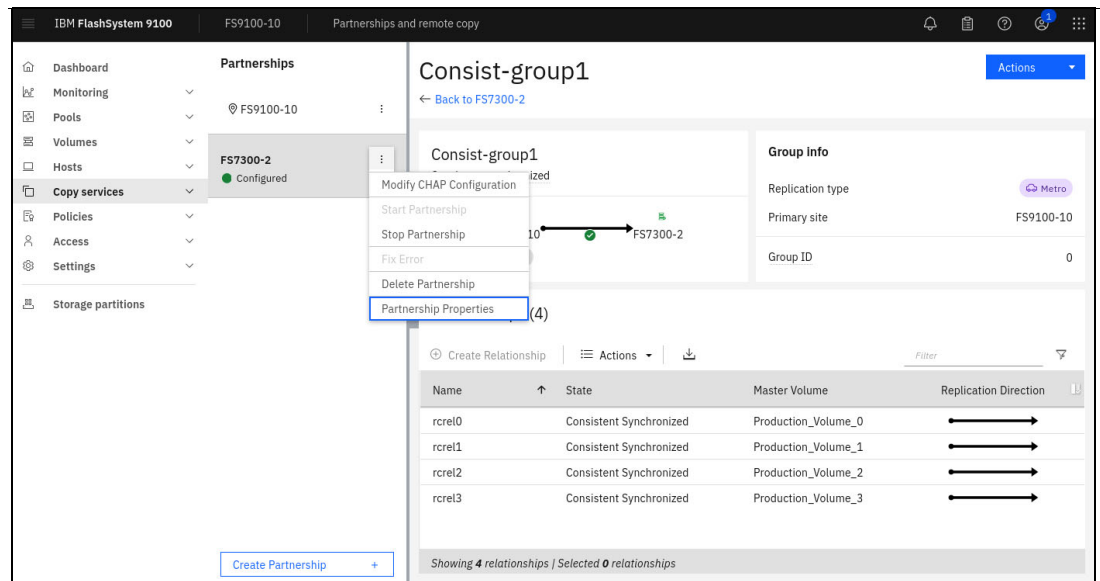


Figure 4-24 Partnership properties

- On the **Partnership Properties** page, enable policy-based replication to retrieve the certificate from the remote system. The management interface automatically creates a truststore on the local system to store the remote system's certificate.
 - View the remote system certificate to verify it.
 - Save the changes.
- Repeat steps 1 and 2 on the remote system within the partnership:
 - Enable policy-based replication on the remote system as well.

- b. Create a truststore for each system in the partnership if using the command line interface.
4. After the remote system is enabled for policy-based replication, confirm that the partnership can be configured to use policy-based replication.
 - a. On either of the systems, go to **Copy Services** → **Partnerships and Remote Copy** and select the respective partnership from the left navigation.
 - b. Verify the message This partnership is ready for use with policy-based replication to ensure configuration is successful.
 - c. Verify that the Global Mirror replication relationship is correctly converted to policy-based replication, as shown in Figure 4-25.

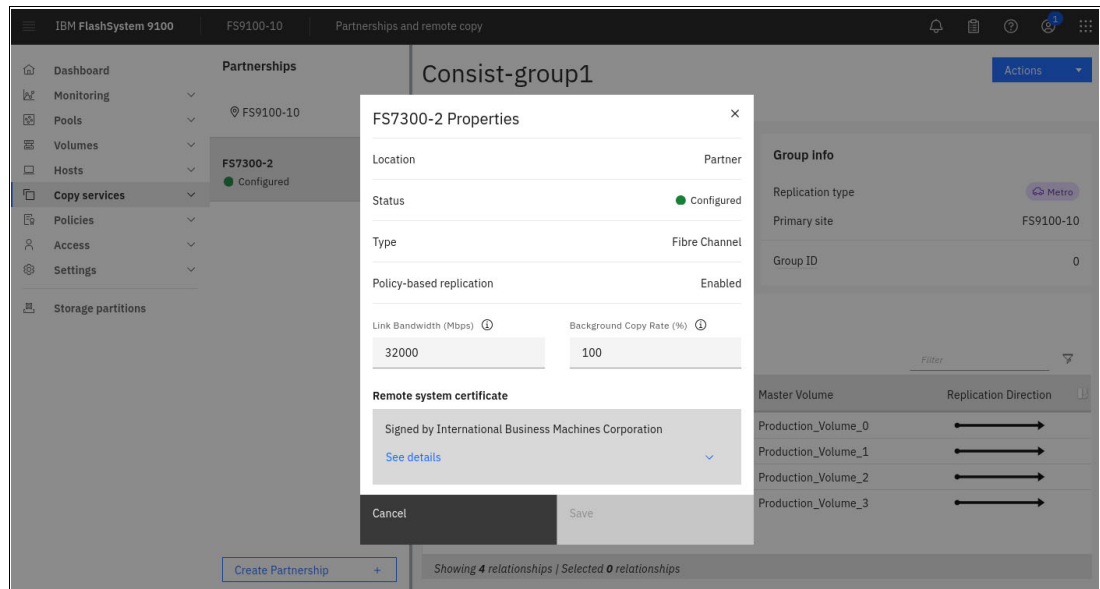


Figure 4-25 Verify that the partnership is ready for policy-based replication enabled

5. Configure a provisioning policy for each linked pool, create one or more replication policies, create an empty volume group for each consistency group and independent relationships and assign a replication policy as described in section “Setup policy-based replication wizard” on page 52. This Setup Policy-Based Replication wizard can be reached from menu **Copy Services** → **Partnerships and Remote Copy**, as shown in Figure 4-26 on page 67

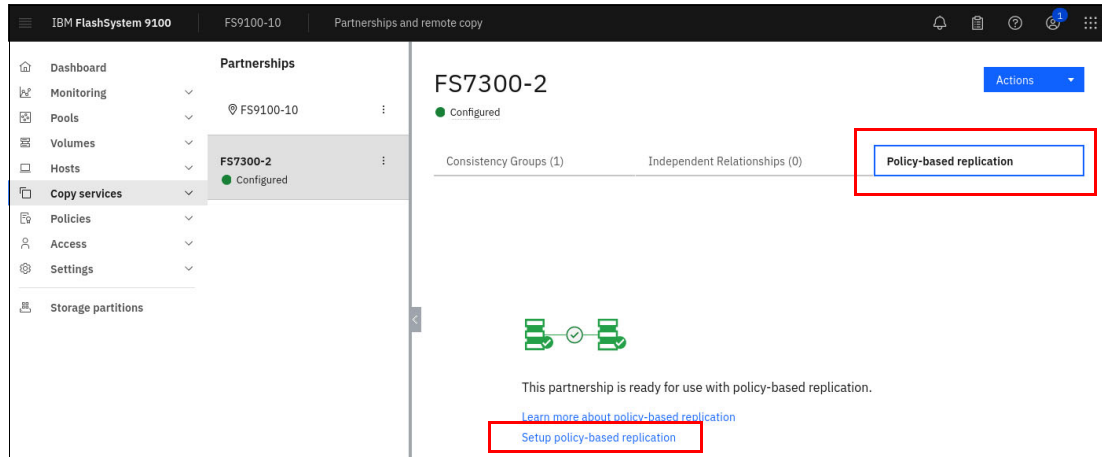


Figure 4-26 Setup policy-based replication wizard

Note: To convert Remote Copy volumes to policy-based replication no associated change volumes can be linked to the primary volume. If trying to move volumes with change volumes for a Remote Copy relationship into a PBR-enabled volume group, you will get error. Instead delete change volumes for existing relationships that might have such.

When the policy-based replication setup has completed we check the recovery system menu **Volumes** → **Volumes** to check that we now see two copies of the replicated production volumes as shown in Figure 4-27

Name	ID	State	Pool	Volume Group	Protocol Type
BG_Local1	4	Online	StandardPool	BG_Local	
BG_Local2	5	Online	StandardPool	BG_Local	
BG_Local3	6	Online	StandardPool	BG_Local	
BG_Local4	7	Online	StandardPool	BG_Local	
Production_Volume_0	0	Online	StandardPool		
Production_Volume_0_1	9	Offline (Recovery Cop...	StandardPool	PBR-VG-01	
Production_Volume_1	1	Online	StandardPool		
Production_Volume_1_1	10	Offline (Recovery Cop...	StandardPool	PBR-VG-01	
Production_Volume_2	2	Online	StandardPool		
Production_Volume_2_1	11	Offline (Recovery Cop...	StandardPool	PBR-VG-01	
Production_Volume_3	3	Online	StandardPool		
Production_Volume_3_1	8	Offline (Recovery Cop...	StandardPool	PBR-VG-01	

Figure 4-27 Volumes exist for Remote Copy as well as for policy-based replication

Remove Remote Copy configuration

When policy-based replication has been enabled on a Remote Copy configuration two copies of the target volumes exist on the recovery system; one for policy-based replication - these can quickly be identified because they belong to a volume group, and the Remote Copy ones. There may not be space enough on the recovery system for two copies, so below is how to remove the Remote Copy configuration and volumes.

Remove Remote Copy configuration via GUI

To remove the Remote Copy configuration, consider the following points and determine if you want to retain existing secondary volumes as a point-in-time copy for disaster recovery while establishing the new recovery copy with policy-based replication.

If you choose to keep the disaster recovery copy, ensure that sufficient capacity is available on the recovery system to accommodate both sets of copies.

Retaining existing volumes provides data protection in case of an outage and allows you to verify replicated data on the recovery system after configuring policy-based replication.

To remove the Remote Copy configuration, follow these steps:

1. Access the primary system's management interface and navigate to **Copy Services** → **Partnerships and Remote Copy**.
2. Select the **Consistency Groups** tab and verify that the current state of the consistency group is **Consistent Synchronized** for a Global Mirror consistency group or **Consistency Copying** for a Global Mirror consistency group with change volumes, as shown in Figure 4-28.

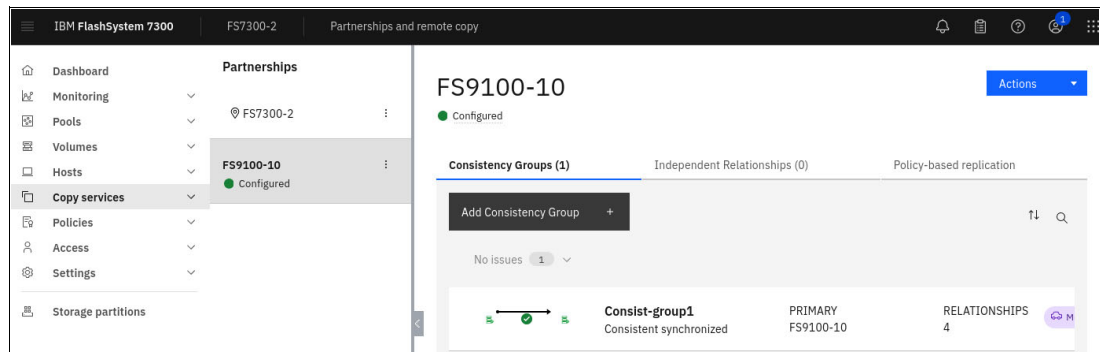


Figure 4-28 Verify the current state of the consistency group

3. Select **Actions** → **Stop Group**. On the **Stop Remote-Copy Consistency Group** page, choose the option **Allow secondary read/write access** to retain the secondary volumes as a disaster recovery copy, as shown in Figure 4-29 on page 68.

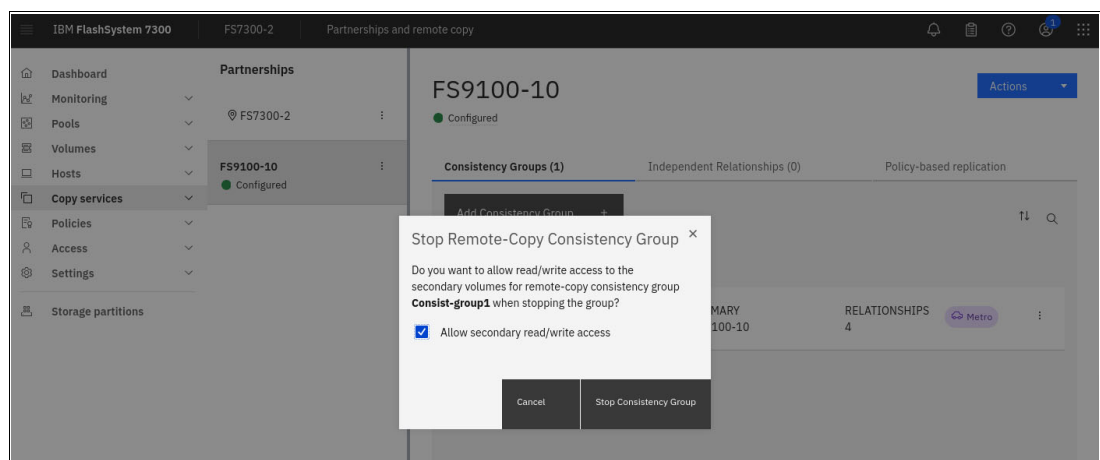


Figure 4-29 Stop Remote-Copy consistency group

- 4. Click **Stop Consistency Group**. The state of the consistency group changes to **Idling**, as shown in Figure 4-30.

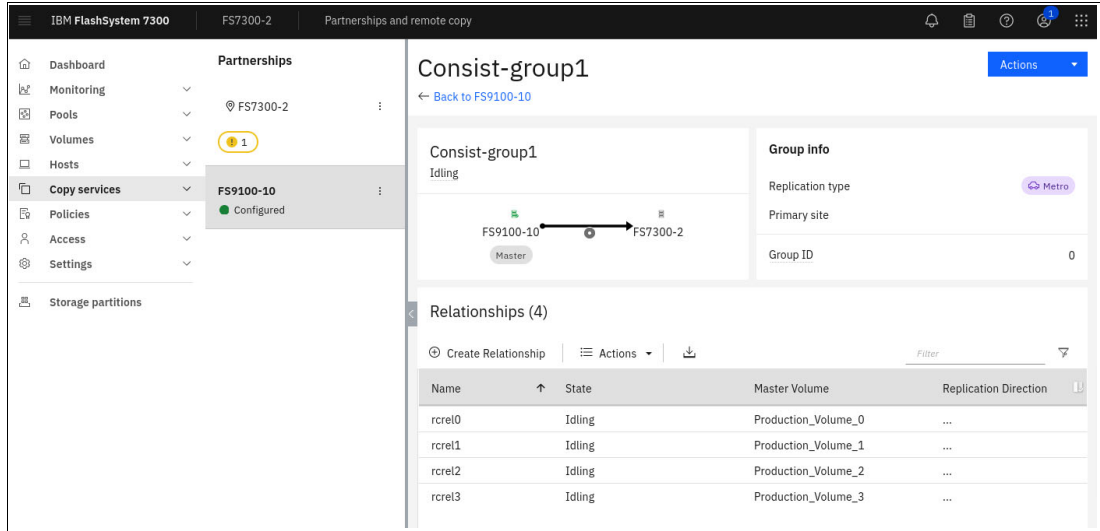


Figure 4-30 State of the consistency group to Idling

- 5. In the **Relationships** section, select all the relationships within the consistency group.
- 6. Right-click the selected relationships and choose **Delete**, as shown in Figure 4-31 on page 69.

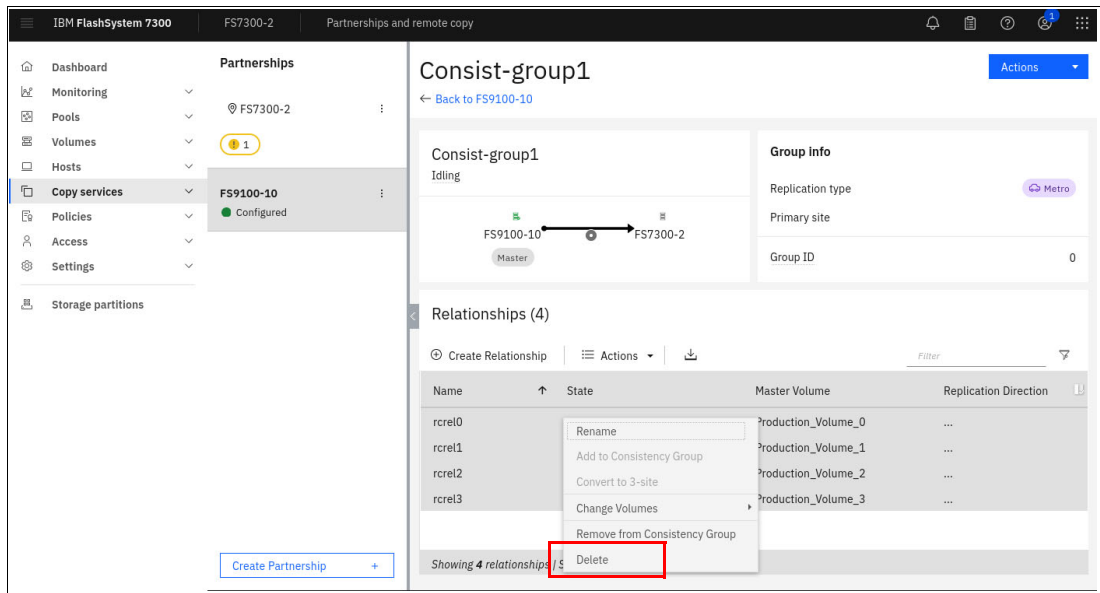


Figure 4-31 Delete relationship

- 7. On the **Delete Relationship** page, verify the number of relationships being deleted. Verify that the checkbox **Delete the relationship even when the data on the target system is not consistent** is cleared. This cleared checkbox allows the secondary volumes to be retained for disaster recovery until a new recovery point is established using policy-based replication, as shown in Figure 4-32.

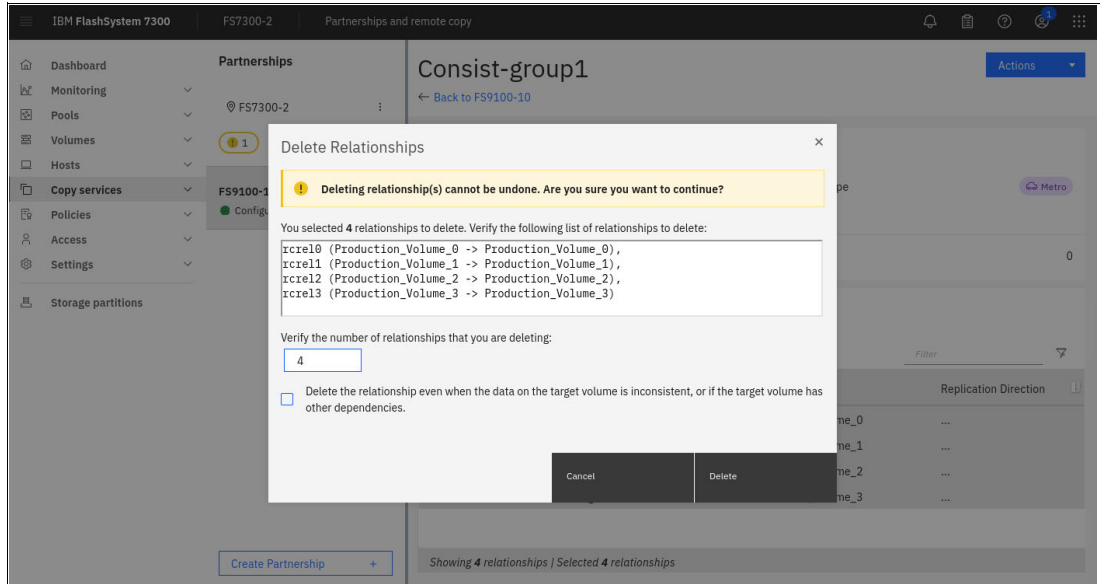


Figure 4-32 Confirm relationship deletion

8. Click **Delete**. After the relationships are deleted from the consistency group, select **Actions** → **Delete Group** to complete the removal process, as shown in Figure 4-33 on page 70.

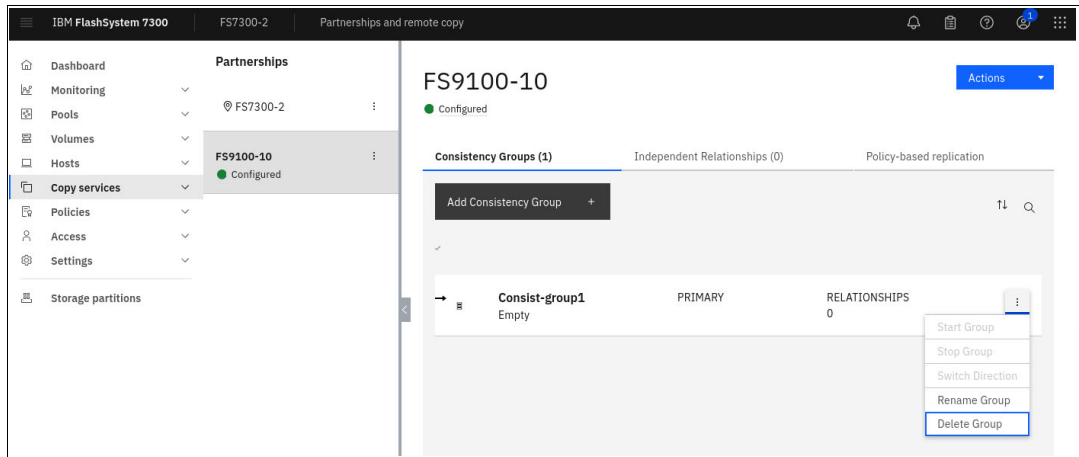


Figure 4-33 Delete group

The Remote Copy configuration is now deleted. The volumes from the Remote Copy relationship exist and is available for host mapping, or they can be deleted. Be aware that both copies Remote Copy and PBR volumes requires space, which requires free space in the storage pool.

The GUI will continue to show the Remote Copy features. This can be disabled from the menu **Settings** → **GUI Preferences** → **GUI Features** → **Remote Copy Functions in copy services**, as shown in Figure 4-34 on page 71.

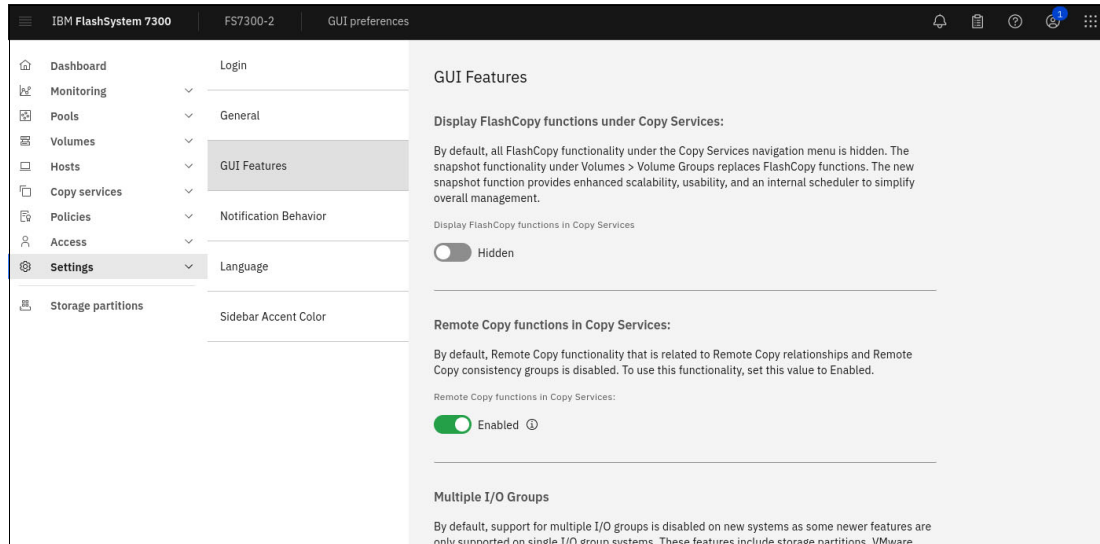


Figure 4-34 GUI preferences - remove Remote Copy features

Ransomware and cyberattacks

In parallel to replicating from one system to another, which helps in case of hardware or accessibility issues, the organization and the storage administrator may also want to implement a Safeguarded Copy environment. Safeguarded copies are immutable snapshots or cyber-resilient point-in-time copies of volumes that cannot be changed or deleted through user errors

For more information of Safeguarded Copy see the Redpaper *Data Resiliency Designs: A Deep Dive into IBM Storage Safeguarded Copy*, REDP-5737.



Managing policy-based replication

Managing policy-based replication is not limited to replication policies, but also partnerships, pool links, volume groups and policies.

Monitoring the RPO is a crucial aspect of business continuity and Storage Virtualize provides several ways of verifying if the objective of recovery points are reached and of receiving alerts if not.

In this chapter we cover the following topics:

- ▶ “Managing partnerships using the GUI” on page 74
- ▶ “Managing pool links” on page 76
- ▶ “Managing volume groups using the GUI” on page 78
- ▶ “Managing replication policies using the GUI” on page 87
- ▶ “Checking the RPO and the status of policy-based replication” on page 89

5.1 Managing partnerships using the GUI

When replicating volumes or volume groups, a partnership between systems must exist. The partnership defines the link between the systems (the type of network and bandwidth). To learn how to create partnerships, see Chapter 1, “Introduction” on page 1.

To manage existing partnerships in the management GUI, from the production system, select **Copy Services** → **Partnerships**. See Figure 5-1.

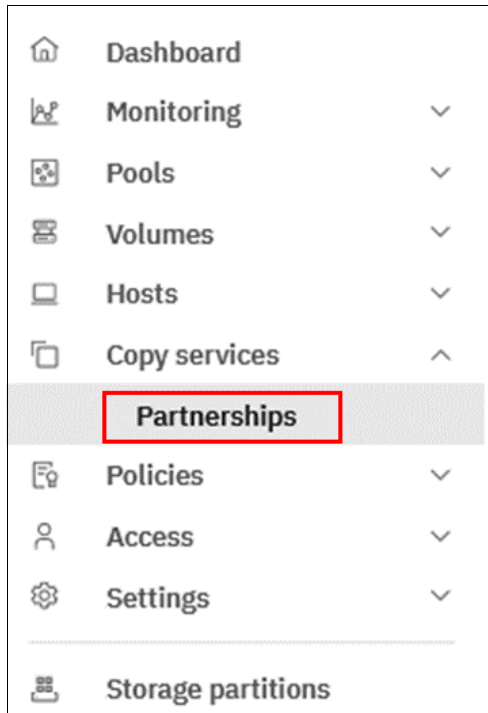


Figure 5-1 Partnerships menu

When partnerships are created on all systems, they appear as “Configured”. It is possible to create a policy for replication, if not already done, from that screen. See Figure 5-2.

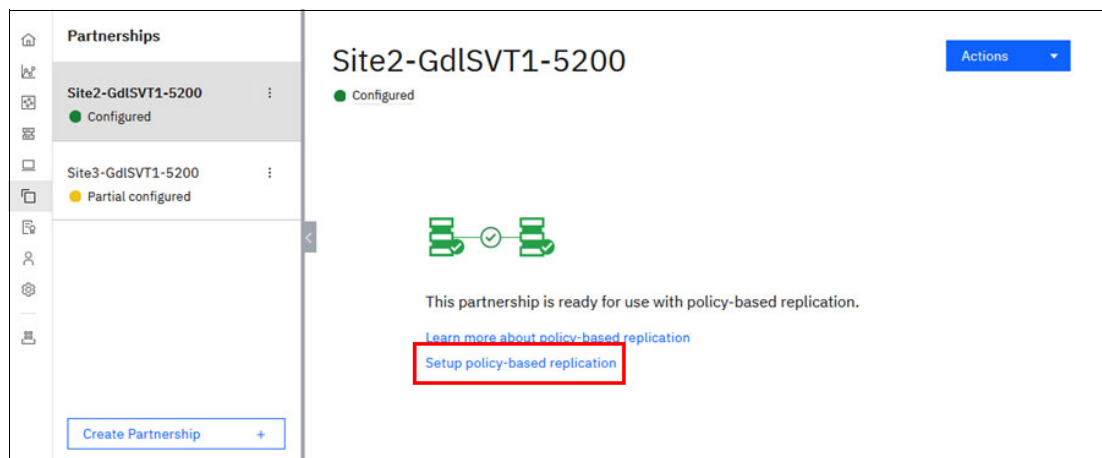


Figure 5-2 Fully configures partnership

Only one partnership can be defined between two systems, but a system can have multiple partnerships. A system can be partnered with up to three remote systems. No more than four systems can be in the same connected set. See Figure 5-3.

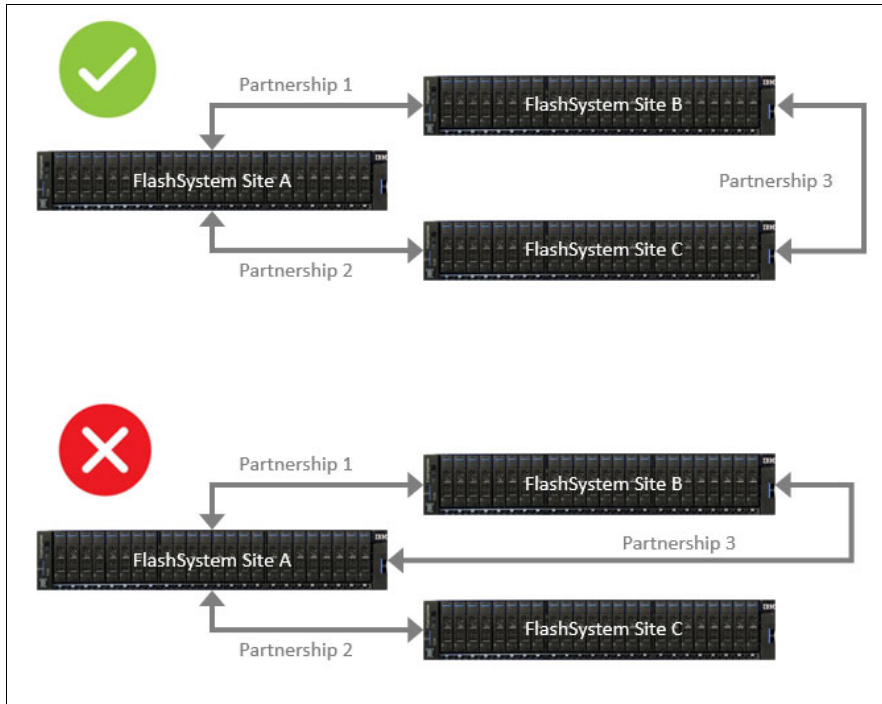


Figure 5-3 Supported multiple partnerships

5.1.1 Stopping a partnership

A partnership can be stopped from the partnership's actions menu. When stopped from a system (production or recovery), a partnership appears as “Local stopped” on the system where the action was taken and “Remote stopped” on the other system. It can only be restarted from the system where it was stopped. See Figure 5-4.

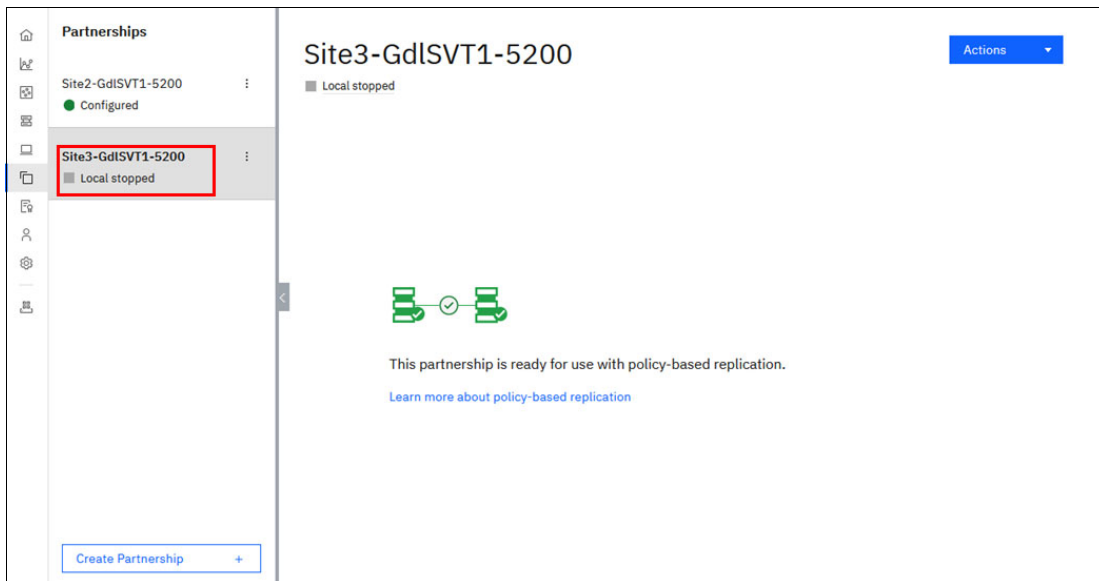


Figure 5-4 Stopping a partnership

When a partnership is stopped, all the volume groups replicating using this partnership are suspended. The Replications status shows, in the Volume Group page, under the **Policies** tab, a disconnected system.

Stopping a partnership can be a way to simulate an interruption of communication between the production system and the recovery system. On the recovery system, in the Volume Groups page, the last recovery point is given. Based on the RPO defined in the replication policy, the recovery volume group can be within or out of the RPO. See Figure 5-5.

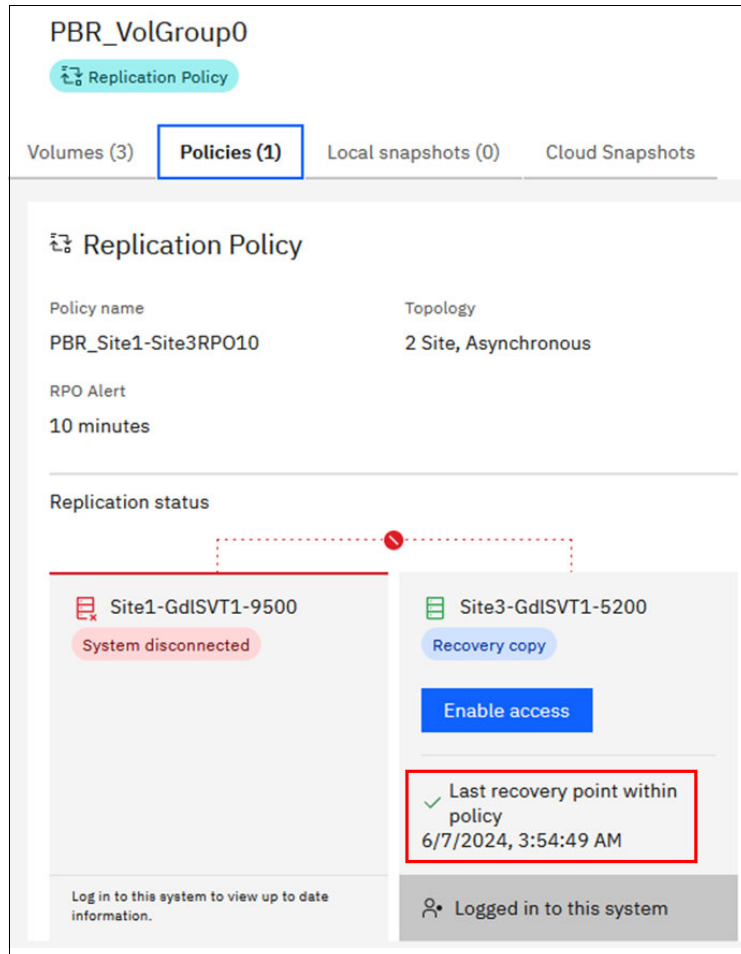


Figure 5-5 Last recovery point on stopped partnership

5.2 Managing pool links

Storage pool linking provides a mechanism to define which storage pool or child pool the system should create copies of a volume.

Use the Pools page in the management GUI to manage storage pools, and pool links between production and disaster recovery locations by navigating to **Pools** → **Pools** page. See Figure 5-6 on page 77.

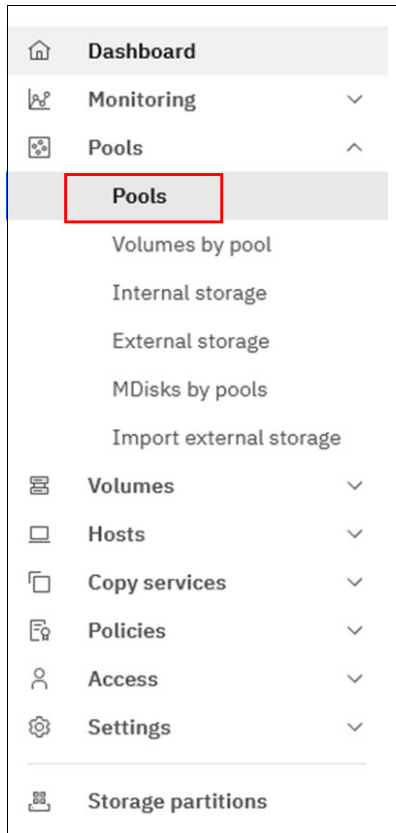


Figure 5-6 Pools management menu

If the storage pools exist on the production and recovery systems, you can add a link between the pools from either system. If a pool on one of the systems has existing links to another partnered system, you must add the link from the unlinked system. The existing link between pools for other partnerships is not affected. Alternatively, if child pools currently exist on the production system only, you can use the management GUI on the recovery system to create and link a child pool in a single step. The management GUI simplifies the process of creating a linked pool on the recovery system. The management GUI automatically displays the properties such as name, capacity, and provisioning policy from the production system. You can use these values to create the new linked child pool on the recovery system without logging in to the other system.

To create a link between storage pools, from the production system, right-click the pool to link and select **Add Pool Link for Replication**. The Add Pool Link page displays. See Figure 5-7 on page 78.

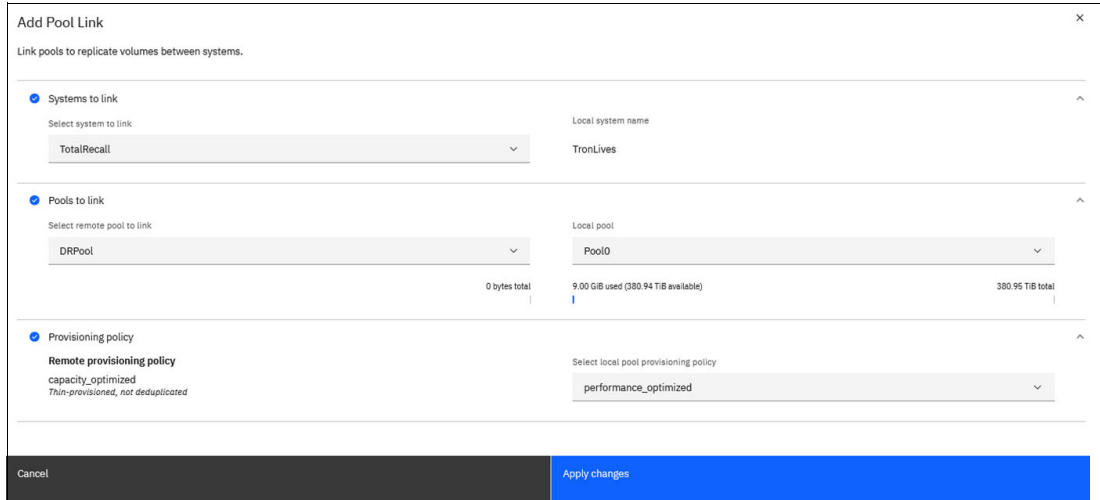


Figure 5-7 Adding a pool link

The Add Pool Link page displays options for the remote system on the left side of the page. Local system details are displayed on the right.

Select the **remote system** to link and select the **remote pool** to link from the left drop-down. Select the local pool from the right drop-down and check whether the provisioning policy is assigned to the remote pool. If a provisioning policy is already assigned to the remote pool, then select the provisioning policy for the local pool from the right drop-down.

To modify pool links between pools in production and disaster recovery locations, In the management GUI, select **Pools** → **Pools**, right-click the pool and select **Modify Pool Links for Replication**. On the Modify Pool Link page, select whether you want to unlink the selected pool from remote systems or move all links from the pool to another pool. See Figure 5-8.

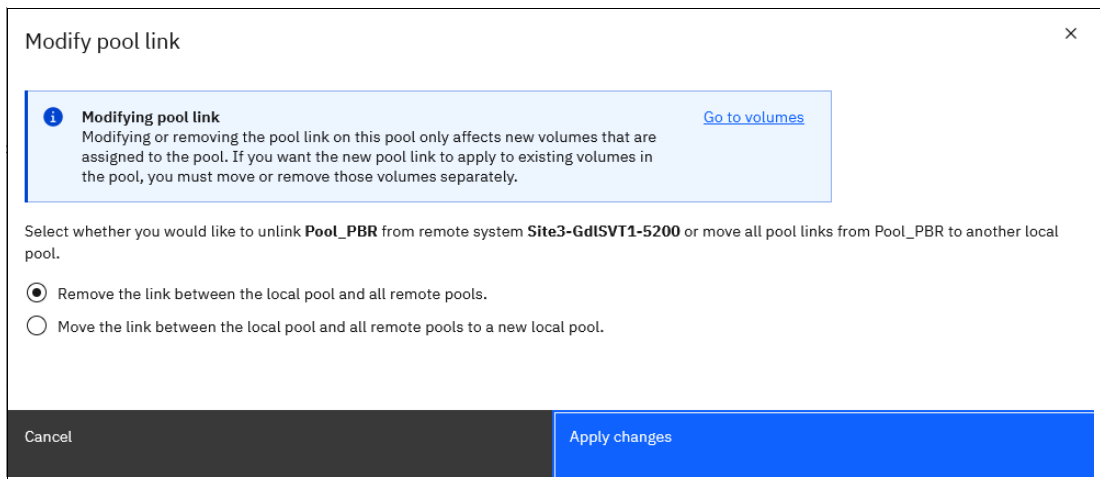


Figure 5-8 Modifying a pool link

5.3 Managing volume groups using the GUI

Volume groups provide a method for grouping volumes that are used by an application.

Replication policies apply to volume groups (and not stand-alone volumes).

To view existing volume groups in the management GUI, from the production system or the recovery system, select **Volumes** → **Volume Groups**. See Figure 5-9.

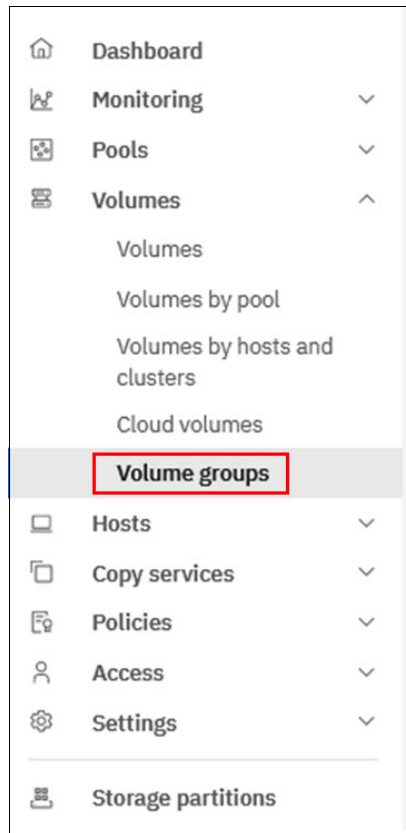


Figure 5-9 Volume Groups menu

Volume groups have multiple attributes among which the following can be changed:

- ▶ A name
- ▶ Volumes
- ▶ An optional replication policy
- ▶ An optional snapshot policy

The available actions on a volume group are renaming it, deleting it, adding or removing volumes, changing the replication policy, changing the snapshot policy, and manage local and cloud snapshots.

The name of a volume group cannot be changed while a replication policy is assigned.

The name of a volume cannot be changed while the volume is in a volume group with a replication policy assigned.

Volume groups can only have a single replication policy. A system can host multiple volume groups, each of them using a different replication policy, but a volume group cannot have multiple replication policies assigned. See Figure 5-10 on page 80 and Figure 5-11 on page 80.

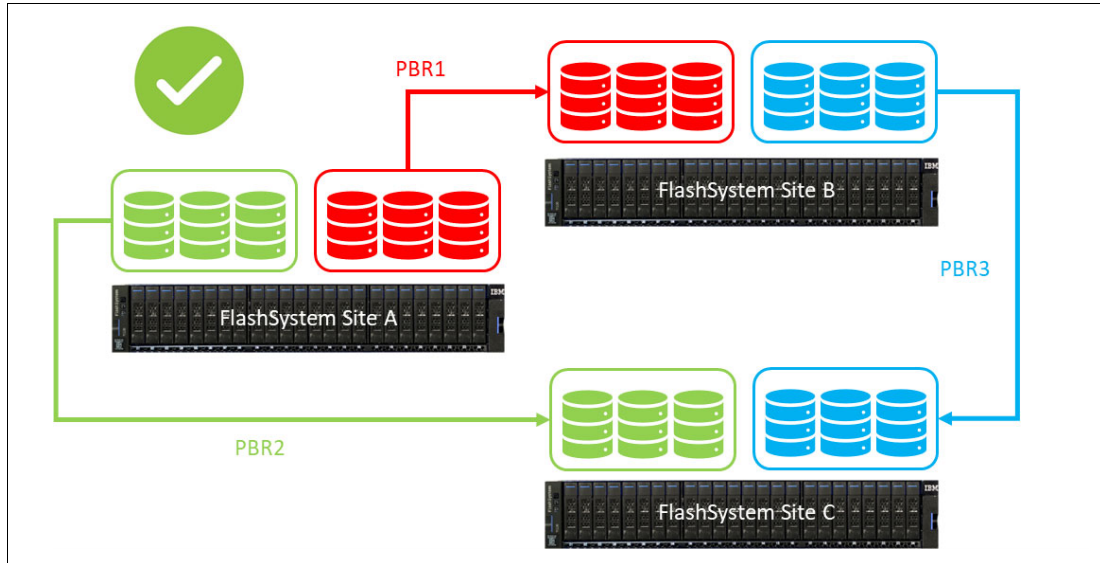


Figure 5-10 Supported multiple replication policies

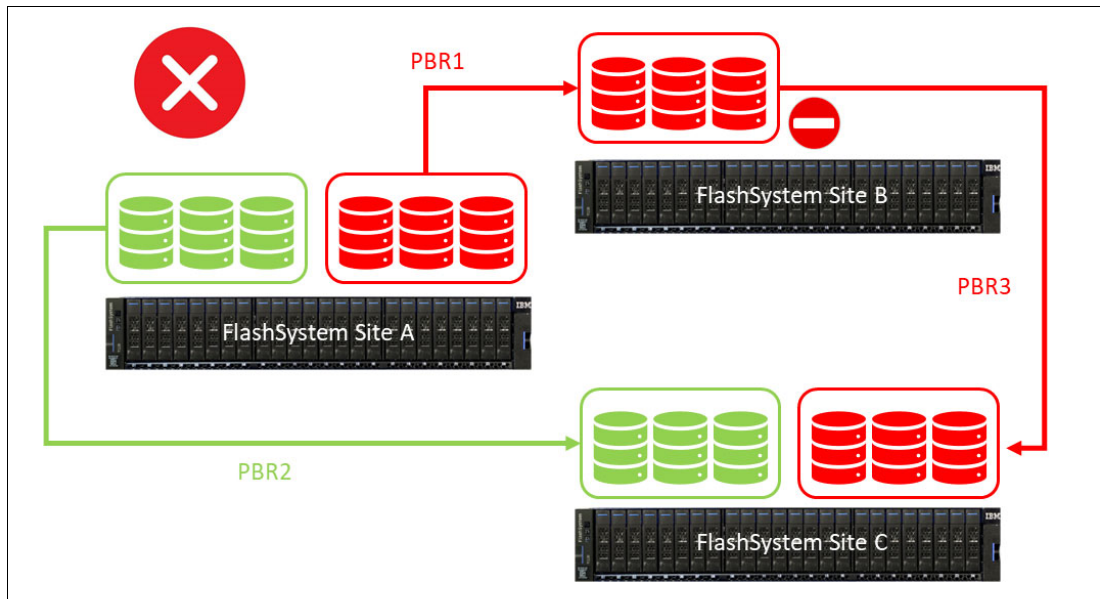


Figure 5-11 Unsupported multiple replication policies

The following actions cannot be performed on a volume while the volume is in a volume group with a replication policy assigned:

- ▶ Resize (expand or shrink)
- ▶ Migrate to image mode, or add an image mode copy
- ▶ Move to a different I/O group

5.3.1 Adding volumes to a volume group

To add volumes in an existing volume group, select the **Volumes** → **Volumes** menu and, in the list of volumes, right-click the ones to be added, then select **Add to Volume Group** and

choose the volume group where to add the volume. **CTRL** and **SHIFT** keys can be used for selection.

Volumes that already belong to a volume group cannot be added to another volume group. See Figure 5-12.

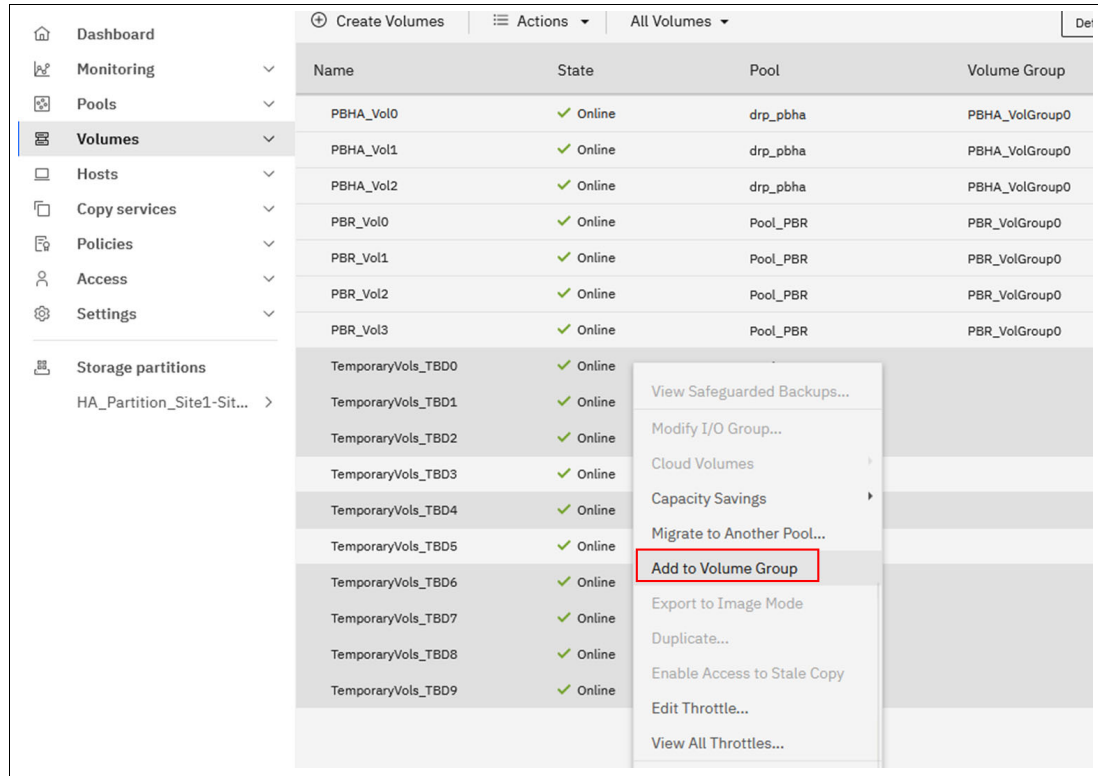


Figure 5-12 Adding volumes to a Volume group

Volume groups are not exclusively used for replication and can be managed on a standalone system for local copies, for instance. Volumes from a volume group, *that are not* associated with a replication policy, can be moved to another volume group. To move volumes from one volume group to another, select the Volumes menu and, in the list of volumes, right-click the ones you want to move. Then, select **Move to Volume Group** and choose the volume group where you want to add the volume.

5.3.2 Removing volumes from a volume group

To remove volumes from a volume group, navigate to **Volumes** → **Volume Groups**, select the volume group you want to remove volumes from, then select the volumes to be removed, right-click and select **Remove from Volume Group** action. See Figure 5-13 on page 82.

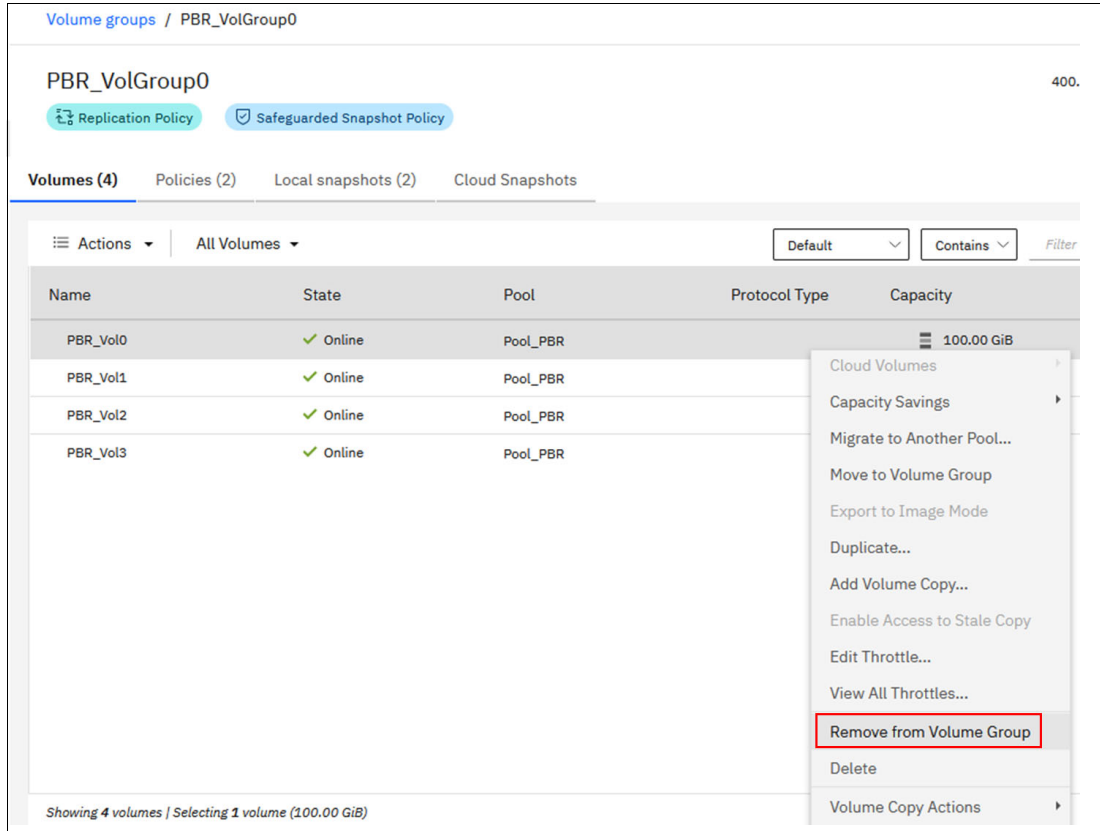


Figure 5-13 Removing volumes from a volume group

Removing a volume from a volume group is a configuration change and can only be performed on the production system. It reflects on the recovery system where the volume is also deleted from the volume group, once it is no longer part of the recovery point.

If a local snapshot was taken for the volume group from which the volumes are deleted, before the deletion, it cannot be restored as the number of volumes are not the same anymore. However, a clone of the volume group can be made to restore the volumes.

Deleting a volume in a volume group is again a configuration change and reflects on the recovery system.

5.3.3 Taking snapshots of volume groups

Volume groups can be copied on local systems by taking snapshots. Snapshots of Volume groups can be made instantly or can be scheduled for regular copies.

To take an instant snapshot of a volume group using the GUI, navigate to **Volumes** → **Volume Groups**, select the volume group to be copied, select the **Local Snapshots** tab and click the **Take Snapshot** button. Instant snapshots cannot be safeguarded, they have no expiration date and can be restored, cloned, or deleted anytime. See Figure 5-14 on page 83.

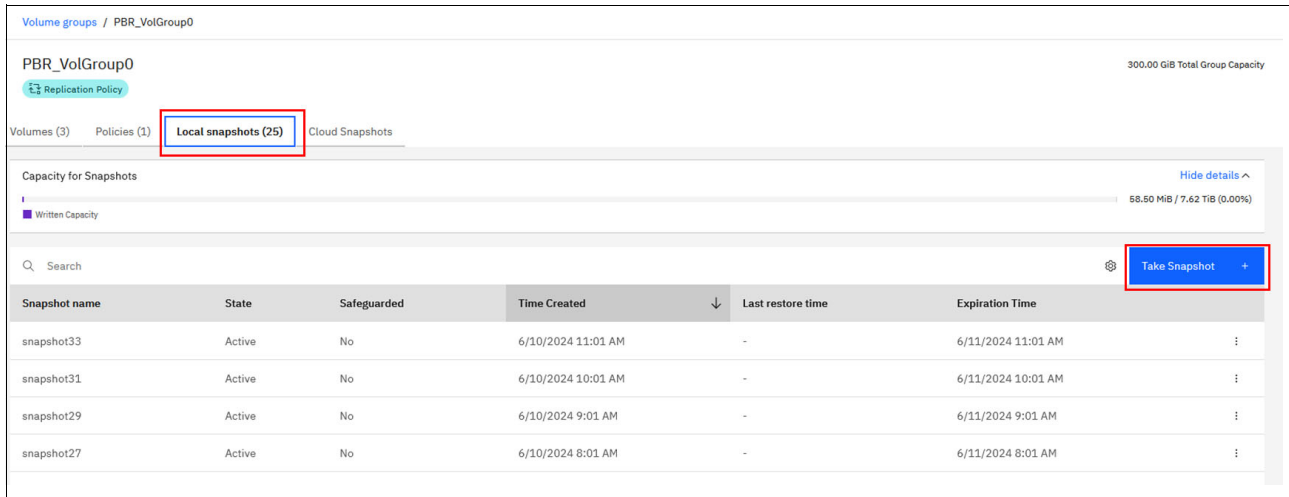


Figure 5-14 Taking volume groups snapshot

A snapshot policy can be assigned to a volume group. To manage snapshot policies, navigate to **Policies** → **Snapshot Policies**. A snapshot policy is defined by a frequency, time, day of the week, day of the month and retention period for snapshots.

To assign a policy to a volume group using the management GUI, navigate to **Volumes** → **Volume Groups**, select the volume group for which to assign a policy, click the **Assign internal snapshot policy** button. See Figure 5-15.

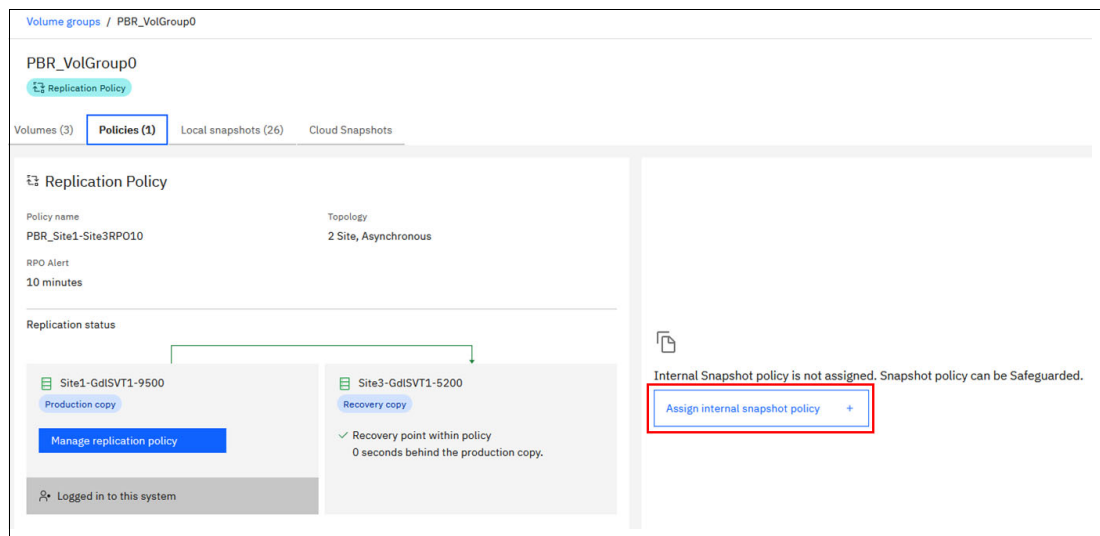


Figure 5-15 Assigning an internal snapshot policy to a volume group

Select the policy to assign to the volume group and specify the start date/time. You can make Safeguarded snapshots of the volume group. Safeguarded snapshots created from the selected volume group are backups that cannot be changed or assigned to hosts. See Figure 5-16 on page 84.

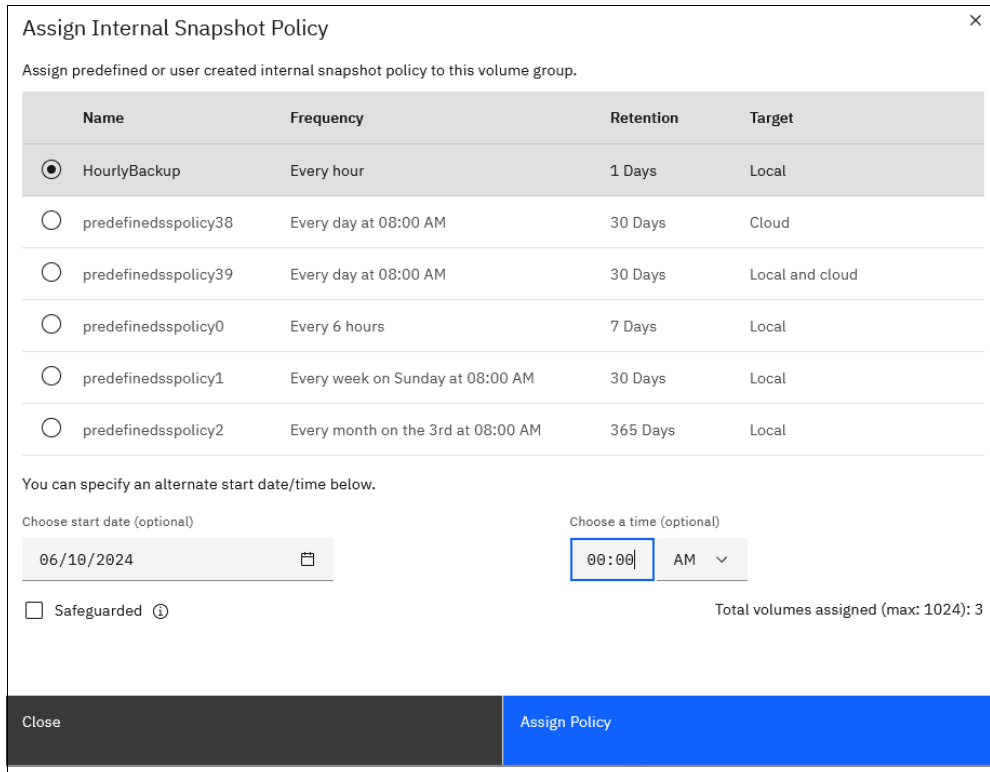


Figure 5-16 Selecting a snapshot policy for a volume group

You can assign only one local snapshot policy to a volume group. You can assign a local snapshot policy and a replication policy to a volume group.

It is possible to orchestrate the snapshots of volume groups through an external tool (like IBM Storage Copy Data Management (CDM) or IBM Copy Services Manager (CSM) to make Safeguarded Copies. If the option to assign an external policy is not visible in the Volume Group page, it might be hidden by the system GUI settings. Go to the **Settings** → **GUI preferences** → **GUI Features** page to make the option visible. See Figure 5-17 on page 85.

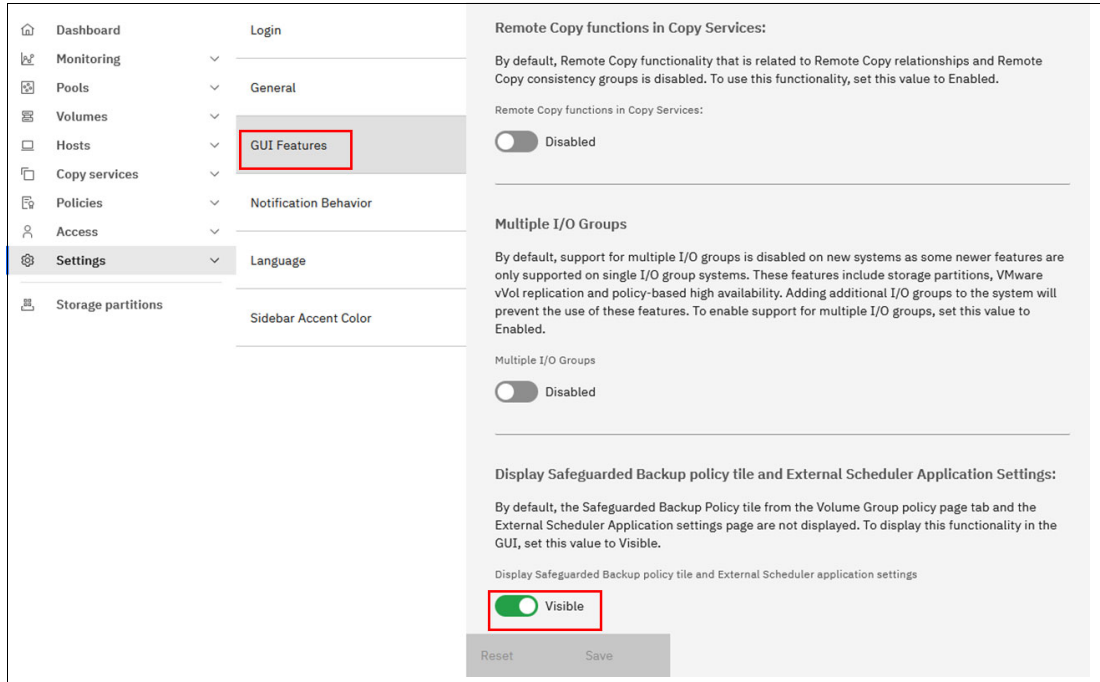


Figure 5-17 Displaying external Safeguarded backup policy in volume groups

A new panel is then available in the volume group properties. See Figure 5-18.

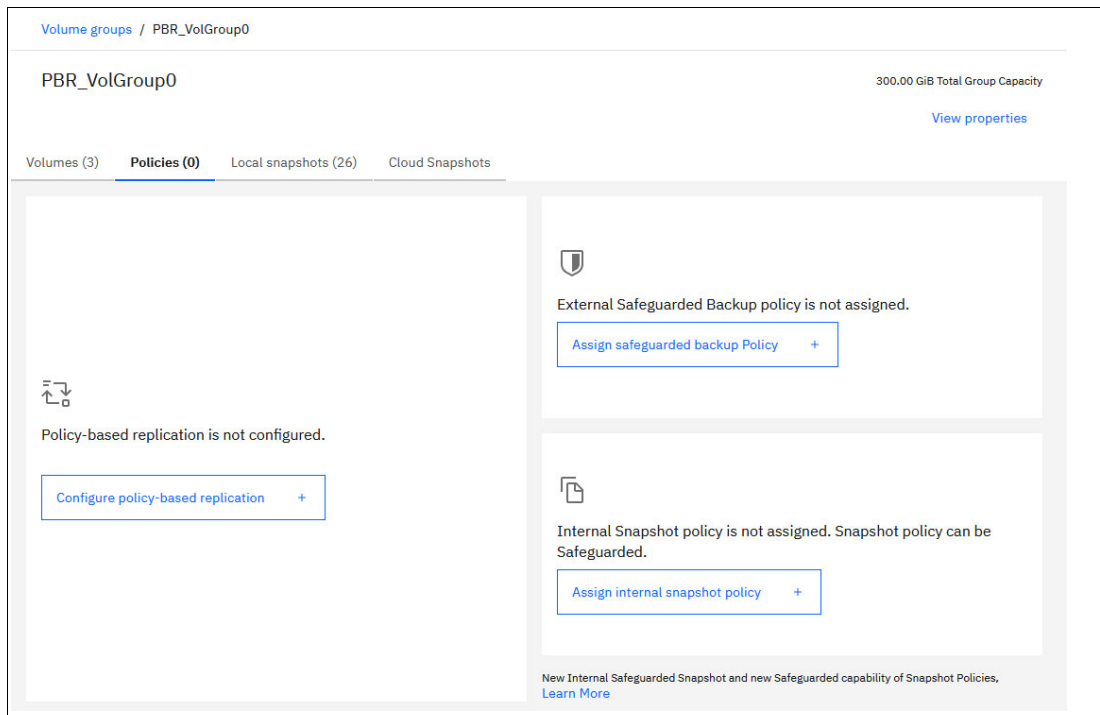


Figure 5-18 A Volume Group page with external Safeguarded backup policies available

5.3.4 Restoring a volume group from snapshots

Volume groups can be restored from one of their snapshots.

To restore a volume group from snapshots, the following requirements must be met:

- ▶ When requesting a restore operation, the volume group specified by the volume group parameter must be the same parent group from which the snapshot was originally created.
- ▶ The composition of the volume group must be the same at the time of the restore as it was at the time the snapshot was taken.
- ▶ If volumes have been added to or removed from the volume group in the time between the snapshot being taken and the restore being requested. Then those volumes must be removed from or added back into the volume group before the restore can be performed.

Note: The volume group on "FlashSystem Site B" may not be mapped to any host preventing any change on data. The last replicated state from "FlashSystem Site A" before enabling independent access (last recovery point) would then be restored.

- ▶ If a volume has been deleted after the snapshot was taken, it will have been removed from the volume group and will be in the deleting state. By requesting a snapshot restore, assuming all other prerequisites have been met and the restore operation proceeds. The volume will be added back to the volume group and put into the active state.
- ▶ If the volumes have been expanded between the time that the snapshot was taken and the restore requested, then the restore will fail. The user will have to shrink the parent volumes back to the size they were when the snapshot was taken before a restore will go ahead.
- ▶ This has implications if new snapshots have been taken since a volume was expanded, as you will not be able to shrink the parent volumes without cleaning out any new snapshots, and their dependent volume groups.

To restore a volume group from snapshots, use the **Volumes** → **Volume Groups** panel in management GUI. Select the volume group that needs to be associated with the snapshots and under the Snapshots tab, select the **Restore** option on the overflow menu for the snapshot being used for the restore. There is an option to either restore the entire volume group or a subset of volumes within that group. Select the applicable option and the subset of volumes if applicable and select **Restore**. See Figure 5-19.

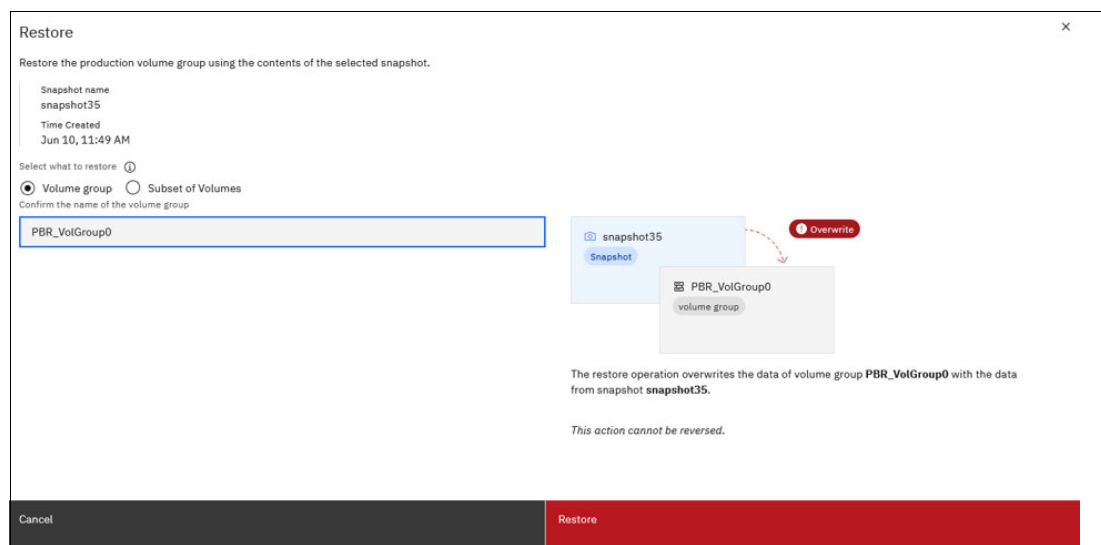


Figure 5-19 Restoring a volume group snapshot

A snapshot restore overwrites the data on the target volume group. It is critical to identify the right snapshot to restore and the target of the restoration.

A volume group in an active replication policy cannot be restored from a snapshot. the replication policy must first be unassigned from the volume group.

To see when a volume group was last restored from a snapshot, go to the **Properties** option on the **Volumes** → **Volume Groups** panel.

5.4 Managing replication policies using the GUI

To manage replication policies in the management GUI, select **Policies** → **Replication policies**. See Figure 5-20.

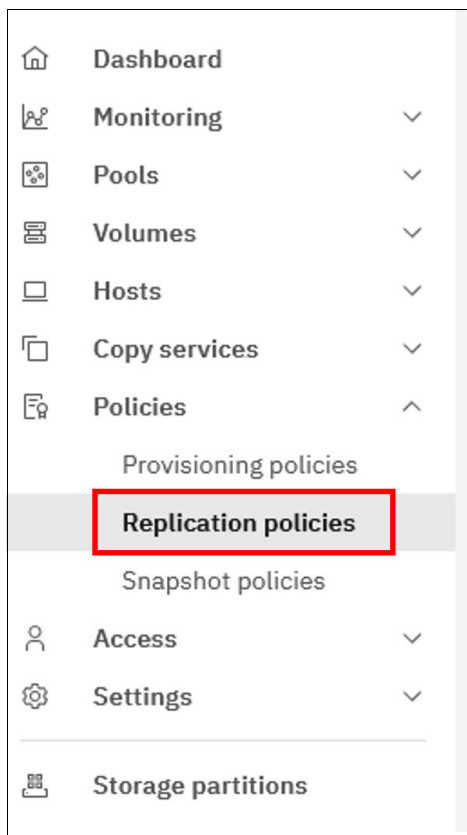


Figure 5-20 Replication policies menu

A list of existing policies is displayed in a table, if any. The table lists the existing replication policies with their name, Location 1 system, Location 2 system and number of volume groups using the policy.

To create a new policy, click the **Create replication policy** blue button. The necessary information to enter are the name of the new policy and the requested RPO (the minimum RPO is 1 minute). See Figure 5-21 on page 88.

Note: In Storage Virtualize version 8.7 (as of this writing), only two-site configurations are supported for both replication (using asynchronous technology) and high availability (HA). This means you can only select two locations, designated as location 1 and location 2.

Create replication policy

You can create a replication policy to define how volume groups are replicated between systems. When you create a replication policy on this system, the policy will automatically be created on the other system.

Replication Policy
A replication policy cannot be changed after it is created. If you want to use different settings in a policy, you must create a new replication policy and assign the new policy to your volume groups.

Name
TwoSitesAsync

Topology
2 Site, Asynchronous

Location 1
System
TronLives

Location 2
System
TotalRecall

Recovery point objective (RPO)
Specify the desired recovery point objective for the policy. An alert will be sent if the recovery point exceeds this value.

Send an alert if data on the recovery copy is older than: 1 min

Cancel Create

Figure 5-21 Creating a replication policy

When the replication policy is created, you can assign volume groups to it by clicking the **Overflow** menu button and selecting **Assign to volume groups**. See Figure 5-22.

Name	Location 1 System	Location 2 System	Volume group count
TwoSitesAsync	TronLives	TotalRecall	0

Assign to Volume Group
Delete

Figure 5-22 Assigning a replication policy to a volume group

The system where the policy is assigned is the production system, and the other system in the policy is the recovery system for the volume group. See Figure 5-23 on page 89.

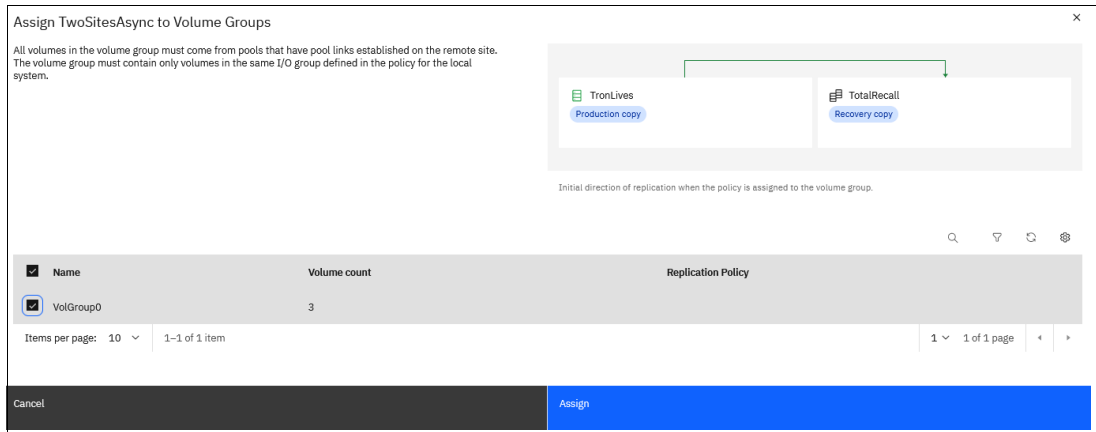


Figure 5-23 Assigning a replication policy to a volume group

Note that if a volume group is already associated to a replication policy, it will not appear in the list of available volume groups.

When a volume group is associated a replication policy, the synchronization of the volumes starts. The system creates the recovery copy of the volume group and volumes on the remote system automatically. There is no need to create them on the remote system.

The first synchronization of the volumes is done at the speed of the available partnership's link bandwidth. The *background copy rate* setting for that partnership is not used if there is only policy-based replication.

To manage replication policies in the management GUI that are assigned to existing volume groups, select **Volumes** → **Volume Groups**. Select the volume group and select **Policies**.

5.5 Checking the RPO and the status of policy-based replication

To guarantee business continuity, it is crucial to regularly monitor the Recovery Point Objective (RPO) and status of your FlashSystem replication. An alert is automatically triggered when a replication falls outside its defined RPO or if the status of the replication link changes. For advanced monitoring needs, some users may prefer to leverage third-party tools that utilize RESTful APIs to track RPO and replication status within their FlashSystem environment.

5.5.1 Checking the RPO and status using the management GUI

To view the replication status for a volume group, navigate to the **Policies** tab within the management GUI.

In the Volumes section, select **Volume Groups** and choose the specific group you want to monitor. On the Volume Groups page, click the **Policies** tab. The RPO status will be displayed under the recovery copy illustration. See Figure 5-24 on page 90.

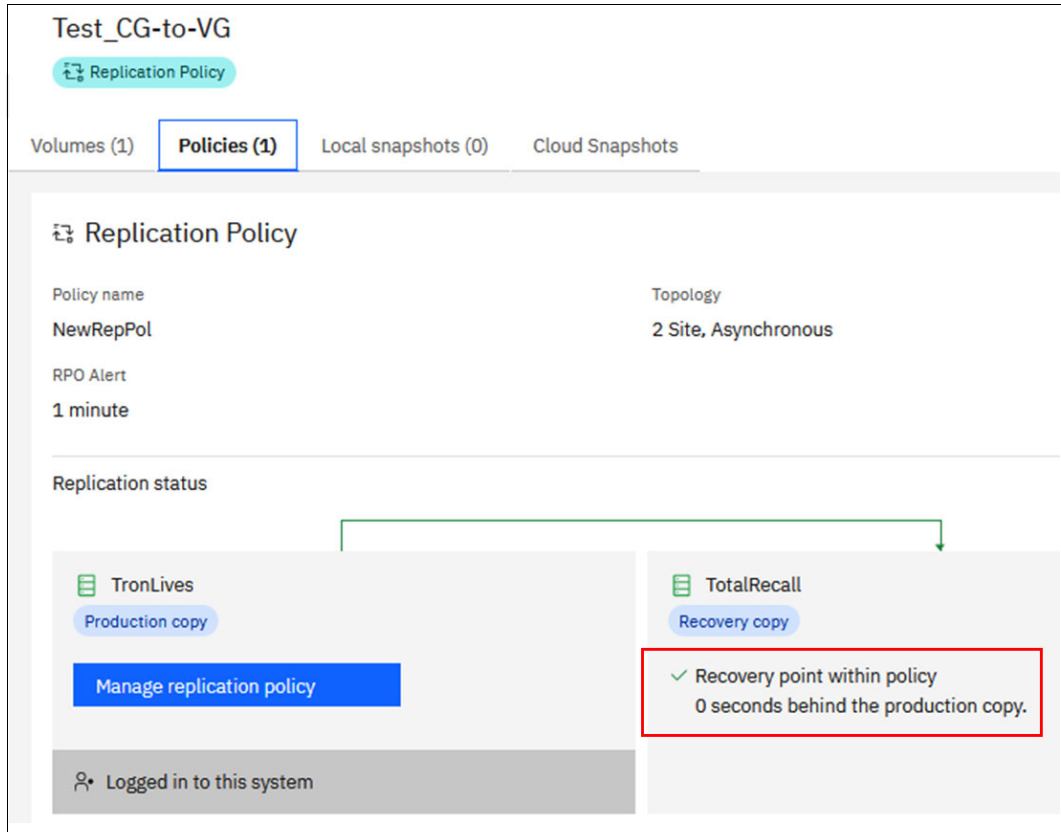


Figure 5-24 Volume group RPO and status

In the Volume Groups list page, the RPO is also displayed for every replicating volume group. See Figure 5-25.

Name	Replication State	RPO Status	Volume Count	Replication Policy	ID
Test_CG-to-VG	✓ Running	✓ Recovery point within policy	1	NewRepPol	1
VolGroup0	✓ Running	⚠ Initial copy in progress	2	NewRepPol	0

Figure 5-25 Volume groups and their RPO

The RPO can have the following statuses listed in Table 5-1.

Table 5-1 RPO status

RPO status	Description
Within policy's RPO	Indicates that replication is within the RPO value set in the policy.
Outside policy's RPO	Indicates that the data on the recovery copy is outside the RPO value set in the policy.
Initial copy in progress	Indicates that the replication is in progress for the first time.
Initial copy incomplete	Indicates that the replication is incomplete, and replication is suspended.

If the recovery point is outside policy's definition, then an alert is triggered and logged in the event log. It is automatically fixed when the recovery point returns within the policy's definition. See Figure 5-26.

Status	Description	Object Type	Object ID	Object Name	Action
Alert	The recovery point objective (RPO) for the volume group has been exceeded	volume_group	1	Test_CG-to-VG	

Figure 5-26 An outside policy RPO alert

The replication status is also displayed and illustrated between the production and recovery copies. The status of the replication can be one of the following (Table 5-2).

Table 5-2 Replication status

Replication state	Description	Action required
Replication running	Indicates that data is currently being replicated between systems.	No action required
Independent access	Indicates that replication is stopped and each copy of the volume group is accessible for I/O.	To resume replication, choose the system you would like to use the data and configuration from. Make this system the production copy.
Replication suspended	Indicates that replication is suspended due to an error on one of the systems. Replication will automatically resume when all errors are resolved.	Review the event log and address errors.
System disconnected	Indicates that the connection between systems is unavailable.	Restore connectivity between the systems.

5.5.2 Checking the RPO and status using REST APIs

You can check the RPO status of replicating volume groups using a tool capable of communicating with the RESTful API server of FlashSystem.

The HTTPS server requires authentication of a valid username and password for each API session. The /auth endpoint uses the POST method for the authentication request. See Example 5-1.

Example 5-1 Authenticating and getting a token

```
curl --location --request POST
'https://<your_flashsystem_ip>:7443/rest/v1/auth' \
--header 'X-Auth-Username: <your_userID>' \
--header 'X-Auth-Password: <your_password>' \
-d ''
```

In the example above, <your_flashsystem_ip> should be replaced with your FlashSystem's IP address, <your_userID> should be replaced by your user ID and <your_password> should be replaced with your password.

The response to this request is a token (a string) which must be used for further requests. It will be referenced in the next example as <your_token>.

Once the authentication is successful (response status is 200) and you have stored the token, use the command `lsvolumegroupreplication`.

In Example 5-2, we use the token that was given in the previous authentication request.

Example 5-2 Requesting volume groups' replication status

```
curl --location --request POST
'https://<your_flashsystem_ip>:7443/rest/v1/lsvolumegroupreplication' \
--header 'X-Auth-Token: <your_token>' \
--data ''
```

The response to this request, if successful (status is 200), is returned and can be exploited by the third-party tool. Example 5-3 is the response to our previous request. It gives us the status of the two volume groups defined on our system.

Example 5-3 Viewing the volume groups and their replication status

```
{
  "id": "1",
  "name": "Test_CG-to-VG",
  "replication_policy_id": "0",
  "replication_policy_name": "NewRepPol",
  "ha_replication_policy_id": "",
  "ha_replication_policy_name": "",
  "location1_system_name": "TronLives",
  "location1_replication_mode": "production",
  "location1_within_rpo": "",
  "location2_system_name": "TotalRecall",
  "location2_replication_mode": "recovery",
  "location2_within_rpo": "yes",
  "link1_status": "running",
  "partition_id": "",
  "partition_name": "",
  "recovery_test_active": "no",
  "draft_partition_id": "",
  "draft_partition_name": ""
}
```

The attributes “location1_within_rpo” or “location2_within_rpo” (depending on the direction of the copy) indicate the replication is within the RPO or not. In the example above, the volume group with id 1 has a recovery point within the policy.

More details can be gathered if the id of the volume group is specified in the request. For example, Example 5-4 shows the response for the request `https://<your_flashsystem_ip>:7443/rest/v1/lsvolumegroupreplication/1`.

Example 5-4 Details of the replication status for a given volume group

```
{
  "id": "1",
  "name": "Test_CG-to-VG",
  "replication_policy_id": "0",
  "replication_policy_name": "NewRepPol",
  "ha_replication_policy_id": "",
  "ha_replication_policy_name": "",
  "local_location": "1",
  "location1_system_id": "0000020421A086D8",
  "location1_system_name": "TronLives",
```

```

"location1_replication_mode": "production",
"location1_status": "healthy",
"location1_running_recovery_point": "",
"location1_fixed_recovery_point": "",
"location1_within_rpo": "",
"location1_volumegroup_id": "1",
"location1_sync_required": "",
"location1_sync_remaining": "",
"location1_previous_replication_mode": "",
"location1_last_write_time": "",
"location2_system_id": "0000020420C082FA",
"location2_system_name": "TotalRecall",
"location2_replication_mode": "recovery",
"location2_status": "healthy",
"location2_running_recovery_point": "0",
"location2_fixed_recovery_point": "",
"location2_within_rpo": "yes",
"location2_volumegroup_id": "0",
"location2_sync_required": "",
"location2_sync_remaining": "",
"location2_previous_replication_mode": "",
"location2_last_write_time": "",
"link1_status": "running",
"partition_id": "",
"partition_name": "",
"checkpoint_achieved": "yes",
"recovery_test_active": "no",
"draft_partition_id": "",
"draft_partition_name": ""
}

```

You can also check the event log for exceeding RPO events. In Example 5-5, after authenticating and retrieving a token, we list the alerts from the event log, with event ID 052004 which is “The recovery point objective (RPO) for the volume group has been exceeded” event.

Example 5-5 Requesting RPO alerts list from the event log

```

curl --location 'https://<your_flashsystem_ip>:7443/rest/lseventlog' \
--header 'X-Auth-Token: <your_token>' \
--header 'Content-Type: application/json' \
--data '{
  "filtervalue": "event_id=052004",
  "fixed": "yes",
  "order": "severity"
}'

```

The response to this request provides a list of all alerts related to the event “The recovery point objective (RPO) for the volume group has been exceeded.” Each alert includes the latest timestamp, the name and ID of the affected volume group, and a sequence number for future reference (to get more details about the event). See Example 5-6.

Example 5-6 Eventlog with exceeded RPO events only

```

{
  "sequence_number": "208",
  "last_timestamp": "240527051415",

```

```

    "object_type": "volume_group",
    "object_id": "1",
    "object_name": "Test_CG-to-VG",
    "copy_id": "",
    "status": "alert",
    "fixed": "yes",
    "event_id": "052004",
    "error_code": "",
    "description": "The recovery point objective (RPO) for the volume group
has been exceeded"
  },
  {
    "sequence_number": "212",
    "last_timestamp": "240529102925",
    "object_type": "volume_group",
    "object_id": "1",
    "object_name": "Test_CG-to-VG",
    "copy_id": "",
    "status": "alert",
    "fixed": "yes",
    "event_id": "052004",
    "error_code": "",
    "description": "The recovery point objective (RPO) for the volume group
has been exceeded"
  }
}

```

Further details for a specific event can be displayed by specifying the sequence number in the request URL. In Example 5-7, we request more details for the sequence number 208. We retrieved the sequence number from previous response.

Example 5-7 Example of specific event details

```

curl --location --request POST
'https://<your_flashsysiem_ip>:7443/rest/1seventlog/208' \
--header 'X-Auth-Token: <your_token>' \
--data ''

```

The returned response gives us details on the event, which can be exploited for further optimization (number of occurrence, duration of the event, and so forth.). See Example 5-8.

Example 5-8 Example of details for a specific exceeded RPO alert

```

{
  "sequence_number": "208",
  "first_timestamp": "240527051415",
  "first_timestamp_epoch": "1716801255",
  "last_timestamp": "240527051415",
  "last_timestamp_epoch": "1716801255",
  "object_type": "volume_group",
  "object_id": "1",
  "object_name": "Test_CG-to-VG",
  "copy_id": "",
  "reporting_node_id": "1",
  "reporting_node_name": "node1",
  "root_sequence_number": "",
  "event_count": "1",

```

```
    "status": "alert",
    "fixed": "yes",
    "auto_fixed": "yes",
    "notification_type": "warning",
    "event_id": "052004",
    "event_id_text": "The recovery point objective (RPO) for the volume group
has been exceeded",
    "error_code": "",
    "error_code_text": "",
    "machine_type": "9846AG8",
    "serial_number": "78E316M",
    "FRU": "None , None , None , None ",
    "fixed_timestamp": "240527051415",
    "fixed_timestamp_epoch": "1716801255",
    "callhome_type": "none",
    "sense1": "00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00",
    "sense2": "00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00",
    "sense3": "00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00",
    "sense4": "00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00",
    "sense5": "00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00",
    "sense6": "00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00",
    "sense7": "00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00",
    "sense8": "00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00"
}
```



6

Implementing policy-based HA

This chapter guides you through implementing policy-based high availability (policy-based-HA) for IBM Storage Virtualize 8.7.

This chapter has the following sections:

- ▶ 6.1, “Implementing policy-based high availability” on page 98
- ▶ 6.2, “Policy-based HA versus HyperSwap” on page 98
- ▶ 6.3, “Configuring policy-based HA” on page 99
- ▶ 6.4, “Migrating storage partitions between systems” on page 113

6.1 Implementing policy-based high availability

Before you implement policy-based HA, ensure your environment meets the requirements for it and that you understand the concepts of the solution see [Planning High Availability](#).

This guide explains how to configure a policy-based HA solution. You can set up policy-based HA using the management GUI or the CLI.

6.1.1 Storage partitions

Storage partitions are vital components in policy-based HA. Storage partitions are:

- ▶ Management units that contain volume groups, volumes, hosts ports, and volume-to-host mappings.
- ▶ Are associated with a HA replication policy.
- ▶ Used to implement policy-based HA available since V8.6.1.
- ▶ Used to implement Storage Partition Mobility available since V8.6.3.
- ▶ Used to implemented Flash Grid which is a feature in V8.7.0

Within a storage partition:

- ▶ All volumes are in volume groups.
- ▶ Host mappings can only be created between volumes and hosts in the same partition.

In storage partitions configured for HA replication, there are two important properties: the preferred management system and the active management system

The preferred management system is configured by the storage admin. Certain error situations can failover the management system to the HA-partner. The system will automatically failback to the preferred management system when it is able to. The preferred management system can also be changed by the user.

All configuration actions on a storage partition must be performed on the active management system. The storage partition can be monitored on either system.

You can configure additional volumes, volume groups, hosts, and host-to-volume mappings at any time, either by adding to an existing partition or by creating a new one. A partition must include all volumes mapped to any hosts included in the partition.

6.2 Policy-based HA versus HyperSwap

Policy-based HA replaces HyperSwap as the high availability solution for Storage Virtualize. Existing systems with HyperSwap can be code updated to version 8.7, but cannot be updated beyond 8.7. HyperSwap can be created on compatible systems on code version 8.7, but policy-based HA provides several advantages over HyperSwap such as:

- ▶ Simplified management.
- ▶ Better performance.
- ▶ Migration options.
- ▶ Non-mirrored volumes remain accessible in the case of connectivity issues.
- ▶ Hardware on the two sites do not need to be compatible.

6.2.1 Migrating from HyperSwap to policy-based HA

One major difference from HyperSwap to policy-based HA is that with HyperSwap two storage enclosures are clustered together as a two-IO-group system acting as a single managed entity. For policy-based HA the two storage enclosures are individual enclosures not clustered together. High availability is managed via hosts and volumes in storage partitions that spans two individual storage enclosures.

The steps involved in converting a two-storage-enclosure HyperSwap configuration to a policy-based HA setup are listed below:

1. **Unmirror HA-volumes:** Convert all HyperSwap volumes to basic volumes leaving a volume copy only on the enclosure (IO-group) to remain.
2. **Migrate non-HA-volumes:** Migrate all non-HA volumes to the enclosure (IO-group) to remain.
3. **Deconfigure HyperSwap:** Change topology to standard.
4. **Remove enclosure:** Given IO-group 0 is the enclosure to remain remove the nodes from I/O group 1. The nodes in IO-group 1 are now in state *Candidate*.
5. **Initialize a new system:** Create a new system using the nodes from I/O group 1.
6. **Configure policy-based HA:** Once you have two independent storage systems, you can configure policy-based HA between them for high availability.
7. **Configure policy-based HA:** On the original system follow instructions in 6.3, “Configuring policy-based HA” on page 99 to create a new storage partition and configure policy-based HA.

Important: There is currently no method to migrate from HyperSwap to policy-based HA while maintaining high availability during migration.

6.3 Configuring policy-based HA

The environment consists of a FlashSystem 9100 and a FlashSystem 7300 interconnected via SAN switches. These switches employ dedicated Inter-Switch Link (ISL) connections, with one dedicated ISL per switch fabric:

- ▶ **Public traffic ISL:** Carries data traffic between the storage systems and the hosts.
- ▶ **Private traffic ISL:** Facilitates communication between the storage system nodes for node to node operations.

Figure 6-1 shows the topology for the systems we are configuring.

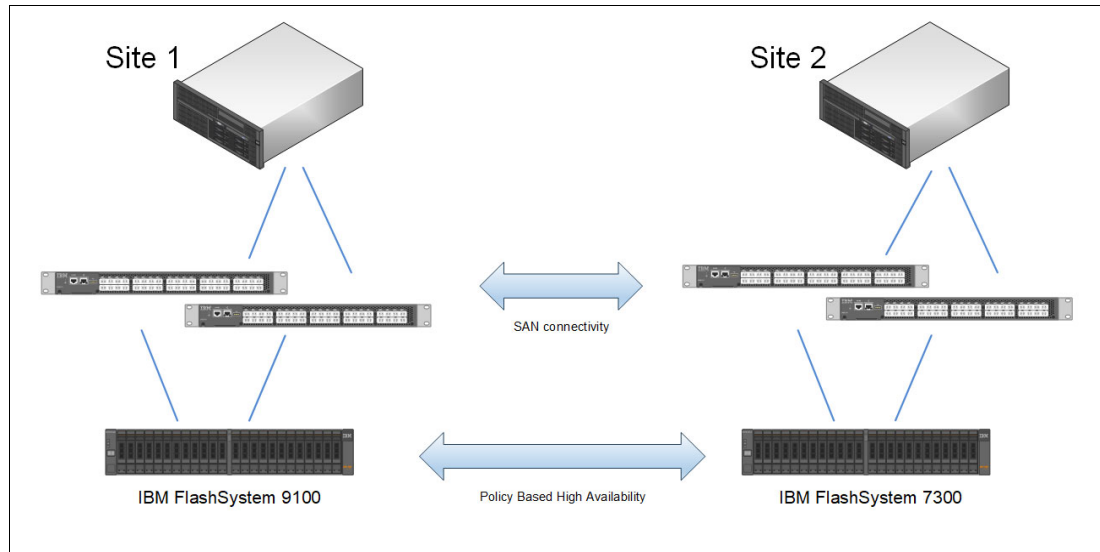


Figure 6-1 The topology of the systems we are configuring

Create policy-based HA using GUI

High availability should be configured from the **Storage Partitions** or the **Copy Services** → **Partnerships** menu in the management GUI.

Follow these steps to configure policy-based HA, by using the management GUI:

1. Check code levels on both connecting systems. Use the GUI to verify that you have Storage Virtualize version 8.7.0 or later on both systems. This is shown in “Check code levels on both connecting systems” on page 49
2. If you do not already have a partnership, create partnership with the HA-partner. This is shown in “Define partnership for replication” on page 49.
3. From the panel **Copy Services** → **Partnerships** select a partnership that is ready for use with policy-based replication, and select **Setup policy-based replication**. This starts the Setup policy-based replication wizard. Select the **HA-partner** and choose **High-availability replication** as shown in Figure 6-2 on page 101. Click **Continue** to proceed.

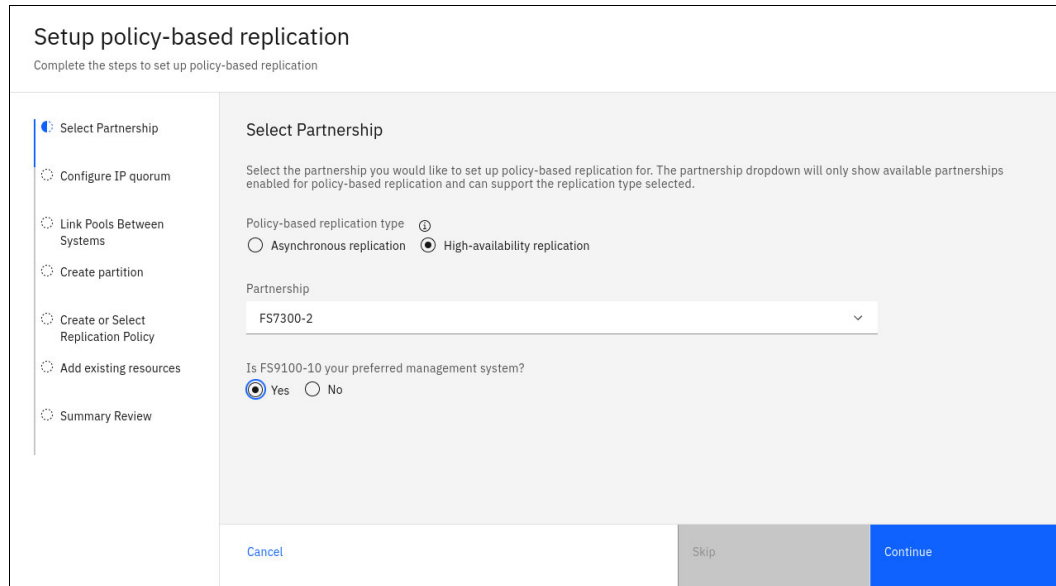


Figure 6-2 The Setup policy-based replication wizard begins

4. Next step is to configure an IP quorum application. Click **Download IPv4 Application** as shown in Figure 6-3. The IP quorum application downloads to the local system and can be executed either from here, or distributed to a host dedicated to running IP-quorum. Click **Continue** to proceed.

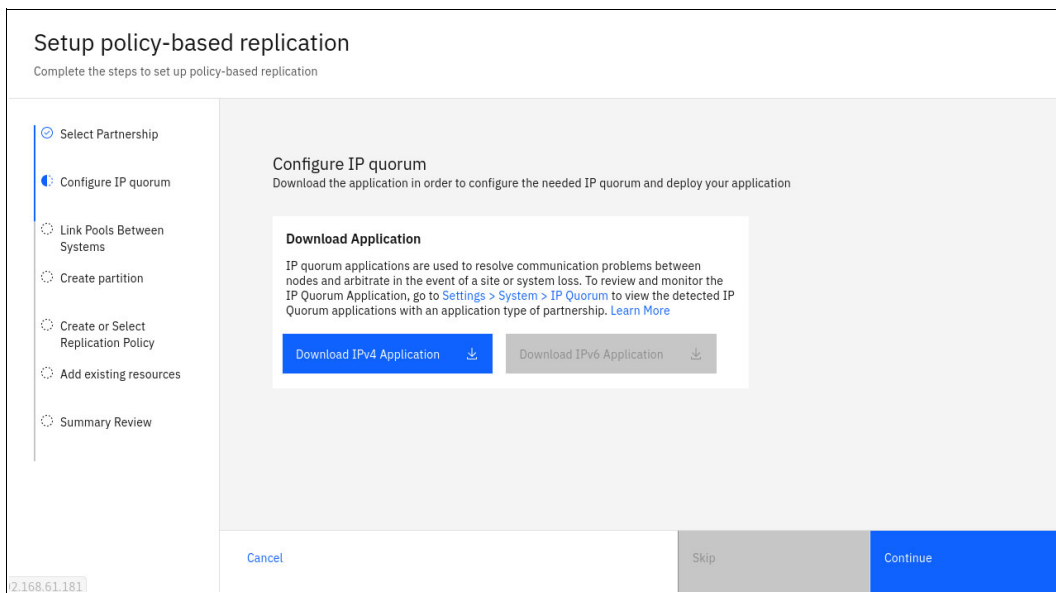


Figure 6-3 Download IP quorum

Note: You need to have connectivity from the *all* servers that are running an IP quorum application to the service IP addresses of *all* nodes or node canisters.

For a full list of IP quorum requirements see [IP quorum requirements web page](#).

For more information on how to use IP quorum see [IP quorum application web page](#).

- The wizard proceeds to Link storage pools between systems. Verify systems to link, and select the storage pools to link on the two systems. We choose StandardPool on both systems. Also choose a provisioning policy. We choose **capacity_optimized** to get thin provisioning for the volumes in these pools. Click **Link Pools** as shown in Figure 6-4 on page 102.

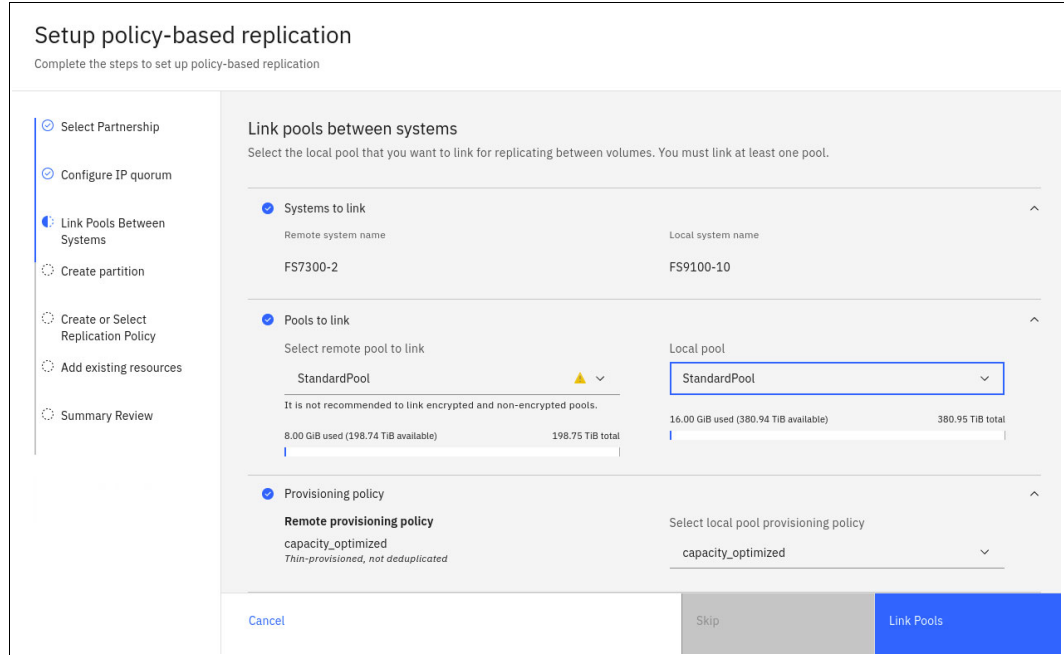


Figure 6-4 Link pools between systems

- Create a storage partition and give it a name. Click **Create partition** to proceed as shown in Figure 6-5.

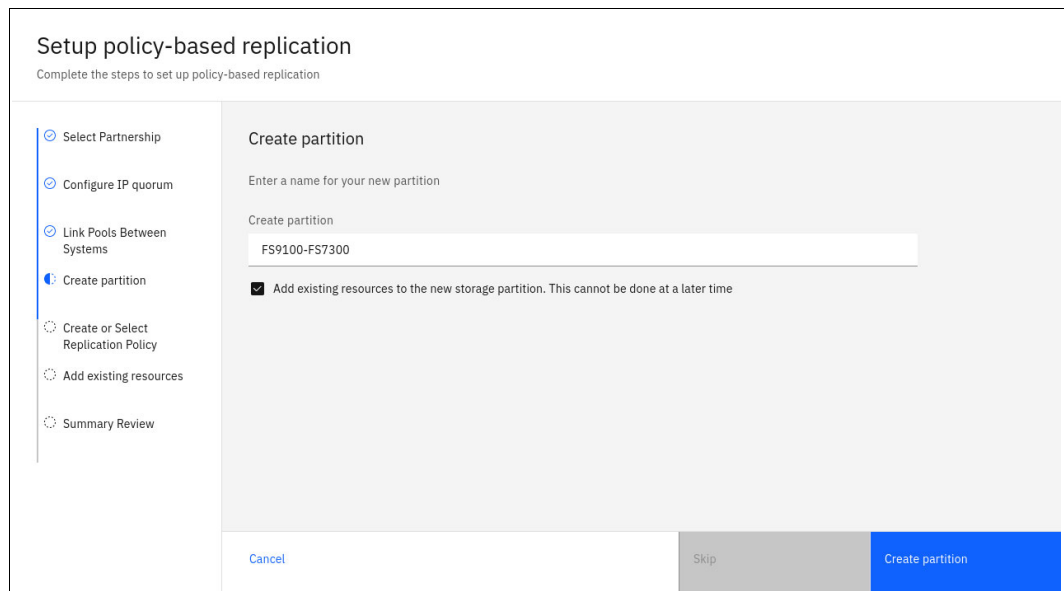


Figure 6-5 Create partition

- 7. Create an HA replication policy and give it a name. We name our partition FS9100-FS7300-HA and click **Create replication policy**, as shown in Figure 6-6 on page 103.

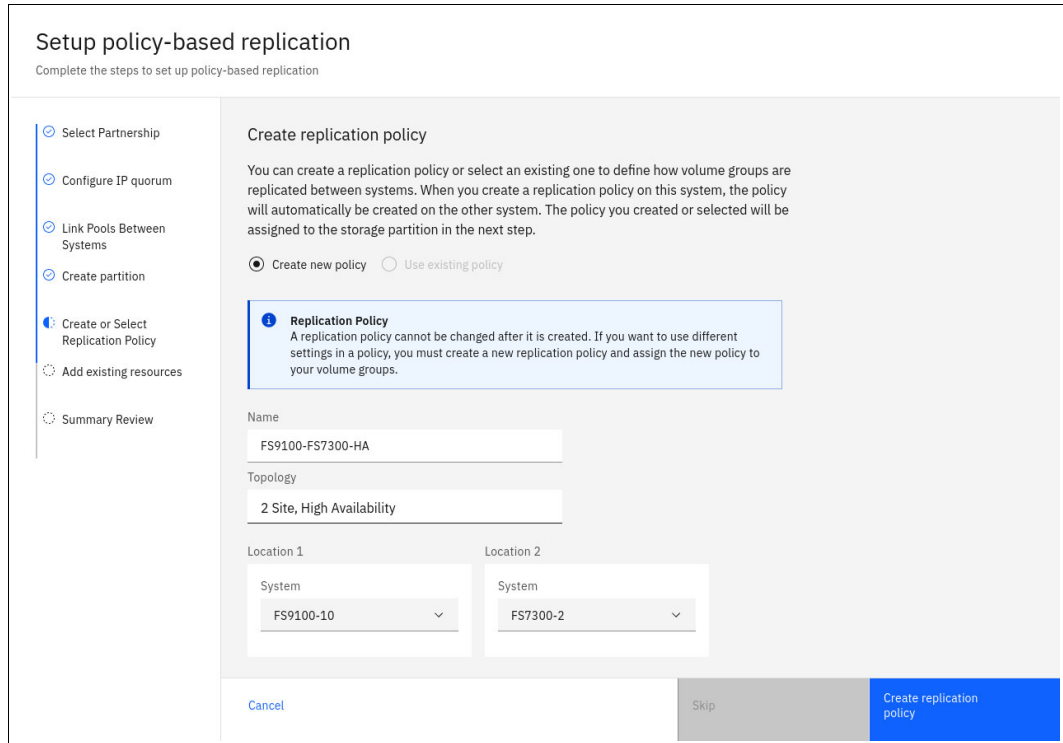


Figure 6-6 Create replication policy

- 8. The Setup policy-based replication wizard prompts for selecting volume groups to add to the newly created storage partition. Volumes in the volume group being selected will be synchronously mirrored between the two partnered systems. The volume group or groups can be selected from a list already created volume groups or this step can be skipped. Click **Select volume groups** as shown in Figure 6-7 on page 103.

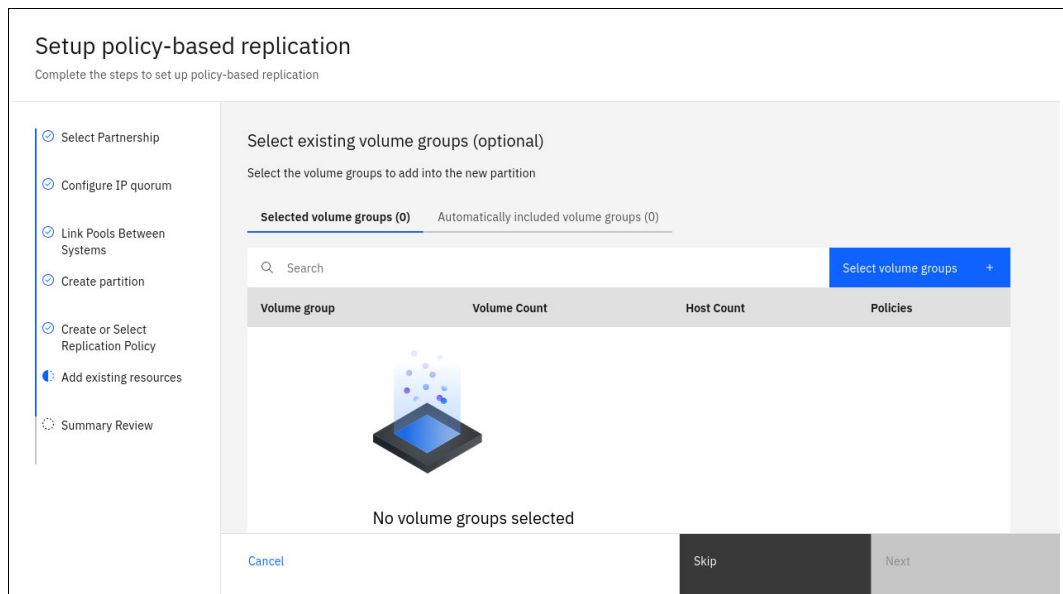


Figure 6-7 Select volume groups

9. We select a volume group already created and containing four volumes. The resulting window displaying this selection is shown in Figure 6-8 on page 104. Click **Next** to proceed.

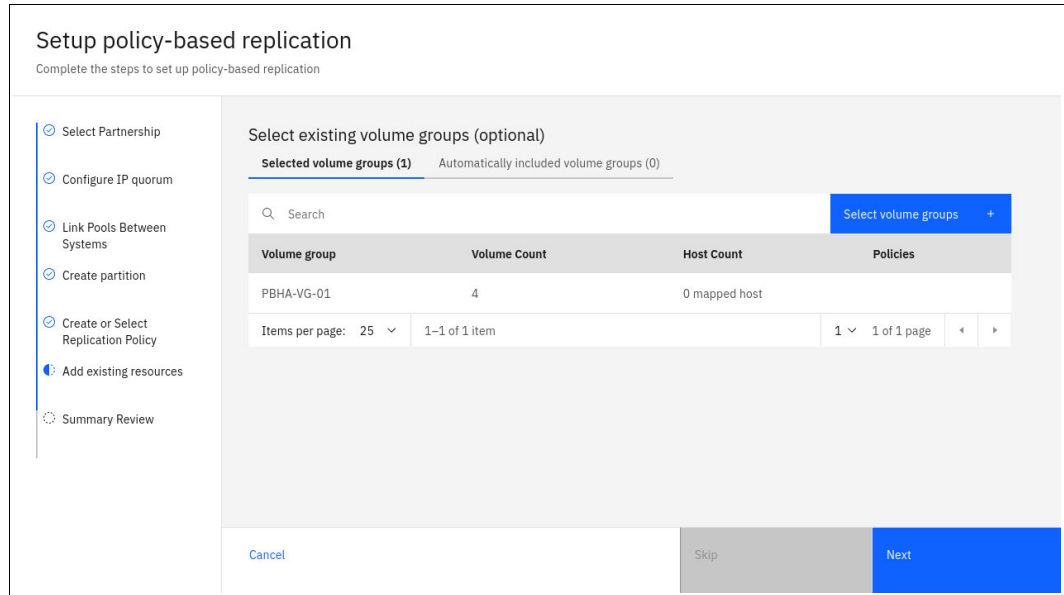


Figure 6-8 Volume group selected

Note: While we have volumes configured, we have not added any hosts yet. We will cover host creation and configuration in “Create hosts in policy-based HA and map volumes” on page 108.

10. The Setup policy-based replication wizard finalizes and shows Summary Review page. We have not enabled IP quorum yet, which is causing a warning message. Policy-based HA can be enabled without quorum, but quorum should be enabled before going into production. Click **Close** to exit as shown in Figure 6-9 on page 105.

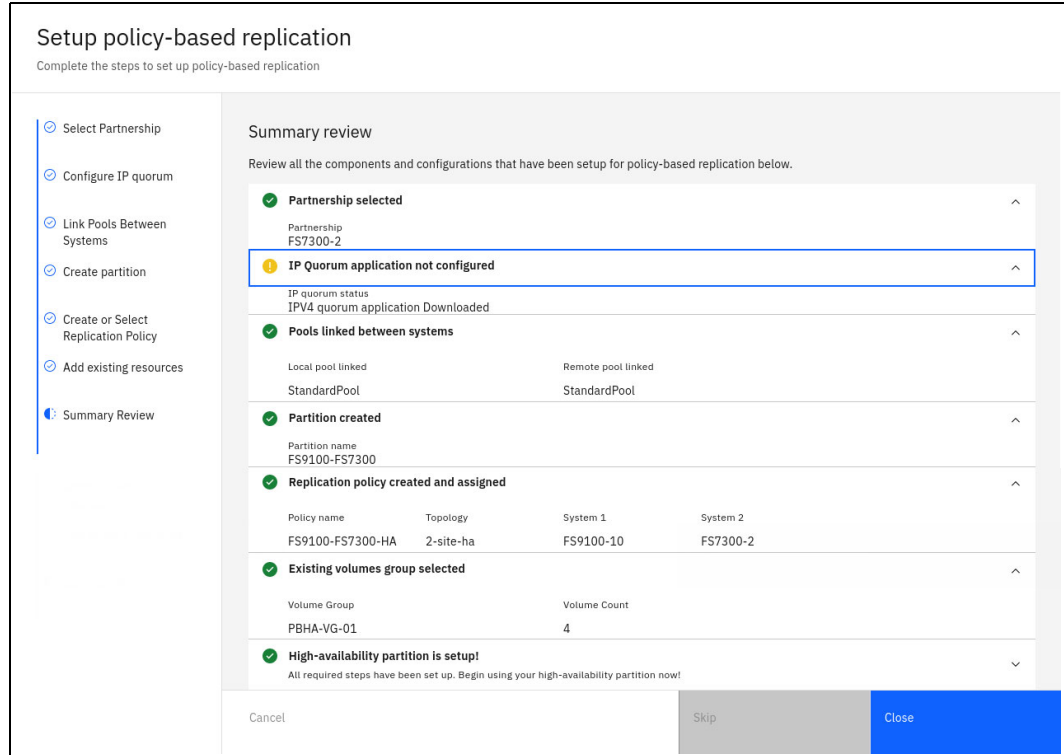


Figure 6-9 Summary review

11. The wizard exits to the Storage Partition view, as shown in Figure 6-10 on page 105.

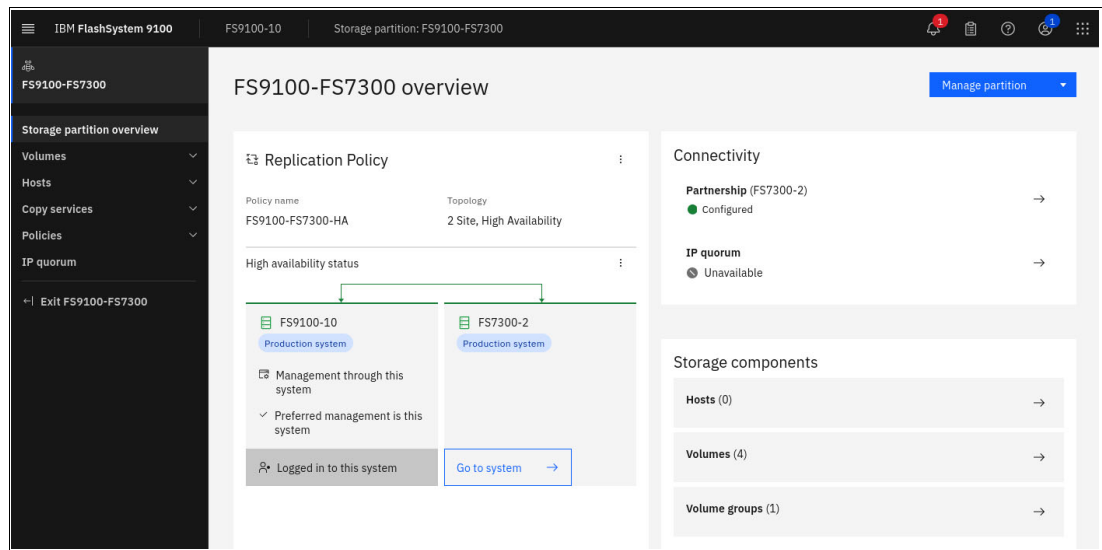


Figure 6-10 Storage Partition view

The Storage Partition overview shows that the FS9100-10 and FS7300-2 are active in a highly available relationship where the FS9100-10 is the preferred management system.

In the scenario above a FlashSystem 9100 and a FlashSystem 7300 are in a highly available relationship together. Depending on the workload, the smaller system should be able to handle all workload which would be a best practice.

The next step involves creating new components on both FlashSystem storage systems:

- ▶ **New hosts:** Define new hosts in the management software.
- ▶ **Volume groups:** Create volume groups to manage related volumes for easier administration.
- ▶ **Host-to-volume mappings:** Establish mappings between the newly created hosts and the volumes they will access.

During host creation, the storage administrator can specify a location preference. This ensures that local FlashSystem volumes on the same site as the host are prioritized for access. This approach optimizes performance and minimizes network traffic.

Use the Storage Partition Overview panel to monitor connectivity between the two systems and the IP quorum applications, and the health of the hosts and volumes associated with the partition.

Create volumes in policy-based HA

Earlier, we established a storage partition and included existing volumes from a volume group. Now, let us demonstrate how to create new volumes within this storage partition.

These new volumes will be created on the primary management system, which is the FlashSystem 9100 (FS9100) in our policy-based HA configuration. Thanks to the replication policy applied to the storage partition, any volumes created here will be automatically mirrored to the partner system, ensuring data redundancy and high availability.

Follow these steps to configure volumes to be used for storage partitions in policy-based HA.

1. From the **Volumes** menu within the storage partition click **Create Volumes**, as shown in Figure 6-11 on page 106

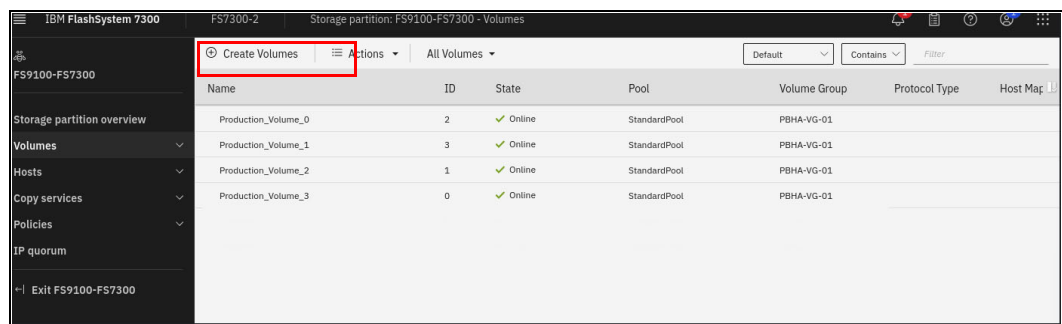


Figure 6-11 Create volumes from within the storage partition

2. Click **Define Volume Properties** on the first window and get to the Define volume properties menu. We provide name and capacity for our new volumes and click **Save**, as shown in Figure 6-12 on page 107

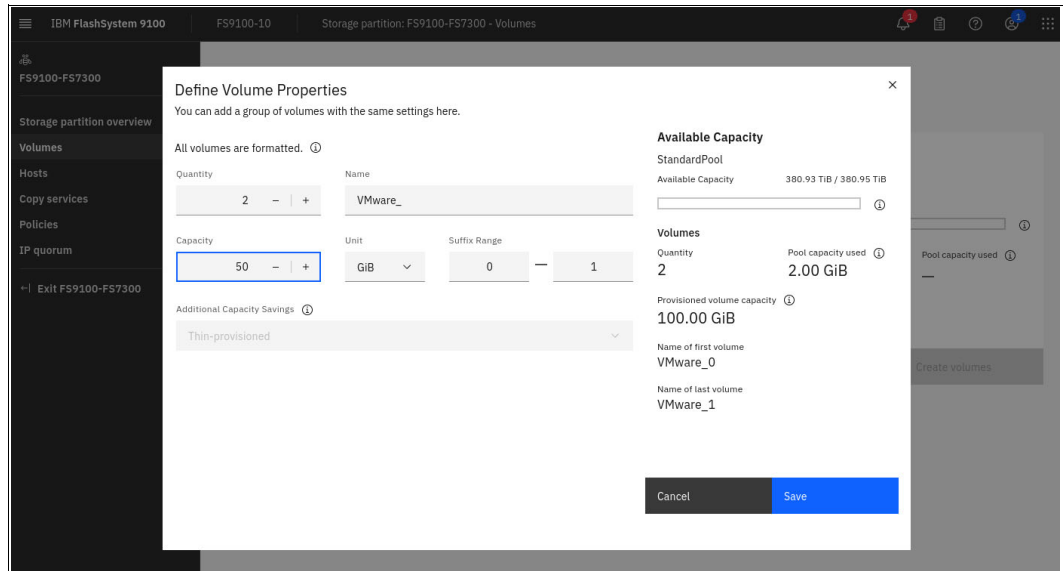


Figure 6-12 Define volume properties

3. Back at the Create Volumes menu we see a summary of changes. Click **Create volumes** to proceed Figure 6-13 on page 107

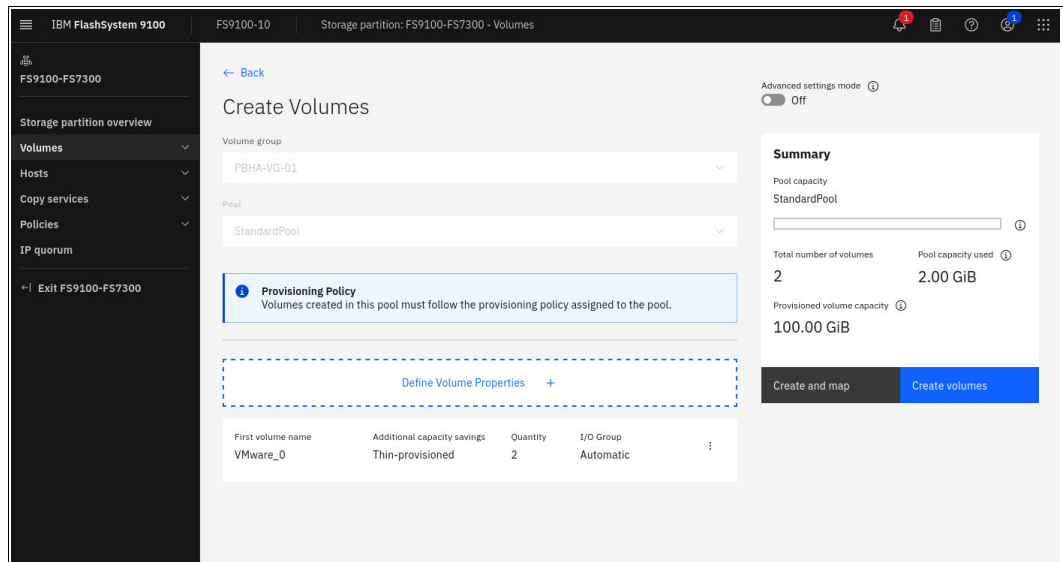


Figure 6-13 Create volumes wizard

4. On the primary management system the FS9100 we now see the new volumes and we see that they belong to a volume group to which a replication policy is enabled, as shown in Figure 6-14 on page 108. These volumes are now replicating to the FS7300.

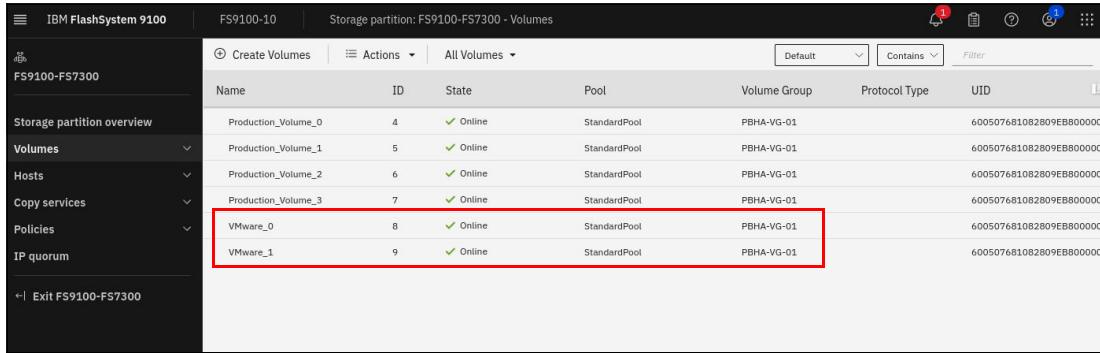


Figure 6-14 Volumes in the storage partition of the FS9100 and FS7300

Volumes which are not required to be mirrored in a policy-based HA relationship can be created from the **Volumes** → **Volumes** menu outside of the Storage partition section of the GUI.

By checking the **Volumes** → **Volumes** menu on the FS7300 HA-partner we see that the new volumes also exist on the FS7300, as shown in Figure 6-15.

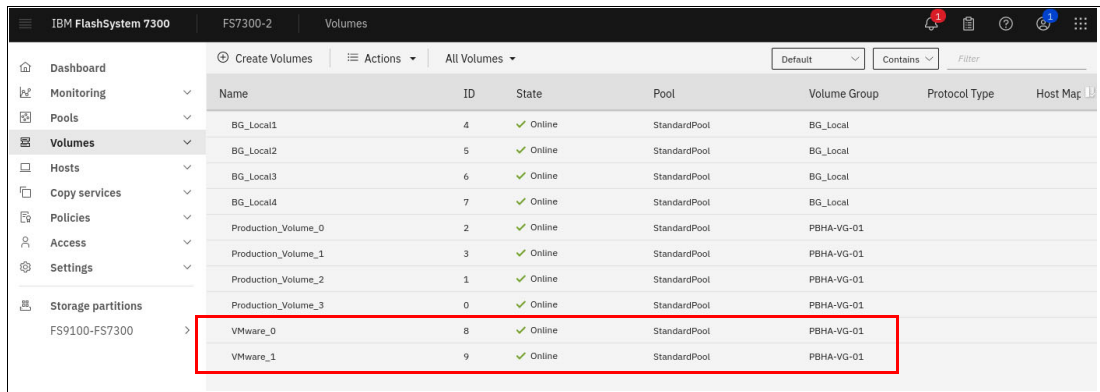


Figure 6-15 Volumes on the FS7300

The volumes you see in Figure 6-15 are accessible from outside the Storage Partitions menu for convenience. These same volumes will also be listed within the Storage Partitions menu.

Since these volumes belong to a volume group with an enabled replication policy, they are mirrored to the HA partner system for redundancy and failover capabilities. This ensures that data remains available even if the primary system encounters an issue.

You will not be able to create volumes attached to the replicating volume group from the FS7300 because it is not the preferred management system. You will be able to create non-replicated volumes outside of the Storage Partitions menu on the FS7300 system.

Create hosts in policy-based HA and map volumes

In the following section we demonstrate how to create hosts and map volumes to it. Follow these steps to map volumes to hosts in policy-based HA:

1. Enter the storage partition FS9100-FS7300 and go to the **Hosts** menu, as shown in Figure 6-16. We currently have no hosts. Click **Add Host** to proceed.

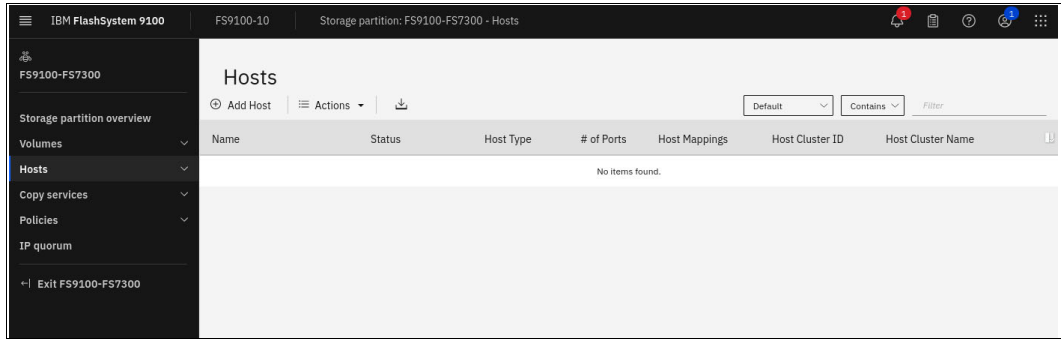


Figure 6-16 Hosts menu in policy-based HA

2. The Add Host wizard opens. Ensure that **Assign location** checkbox is checked. Then, select the preferred location of the host, as shown in Figure 6-17 on page 109. The location name is the hostname of the storage device.

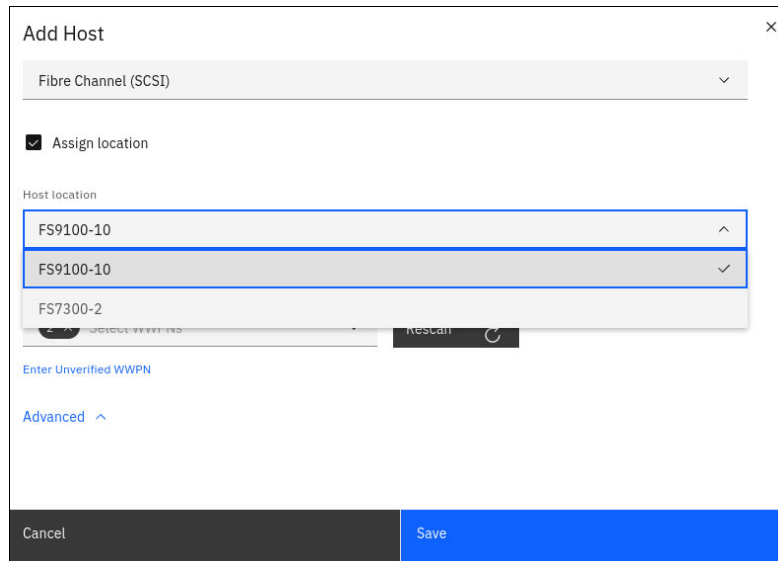


Figure 6-17 Select host location

3. Under Host port (WWPN) select the WWPNs that belong to the host being added, as shown in Figure 6-18 on page 110.

Add Host [X]

Fibre Channel (SCSI) [v]

Assign location

Host location
FS9100-10 [v]

Host port (WWPN)

2 X Select WWPNs [v] [Rescan] [↻]

- 10008C7CFF2E0900
- 10008C7CFF2E0901
- 2100001B3286DB63
- 2101001B32A6DB63

Cancel [Save]

Figure 6-18 Select host WWNs

4. Review the settings and click **Save** to proceed. See Figure 6-19.

Add Host [X]

NPIV Enabled
Because NPIV is enabled on this system, host traffic is only allowed over the storage system's virtual ports. Ensure that SAN zoning allows connectivity between virtual ports and the host.

Name
VMware-1

Host connections
Fibre Channel (SCSI) [v]

Assign location

Host location
FS9100-10 [v]

Host port (WWPN)

2 X Select WWPNs [v] [Rescan] [↻]

[Enter Unverified WWPN](#)

[Advanced](#) [^]

Cancel [Save]

Figure 6-19 Review hosts settings

5. The host is now created and its ports are online, as shown in Figure 6-20.

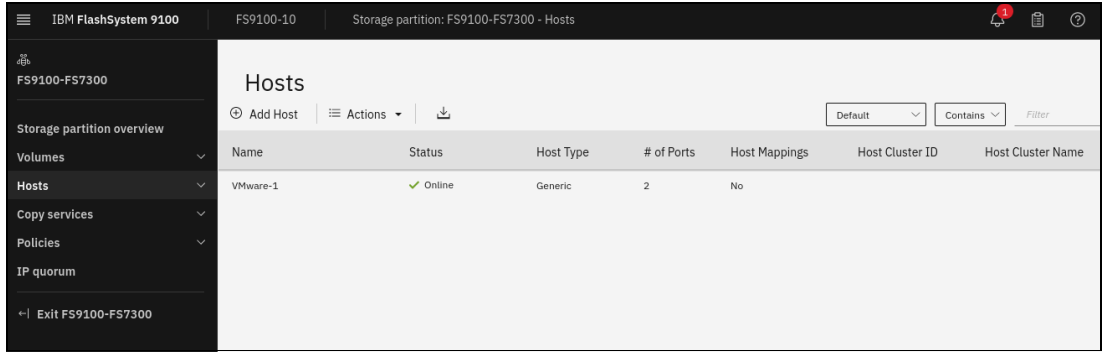


Figure 6-20 Host created

Map volumes to host

We have successfully added the host. Now, let us move on to defining its storage access by creating volume mappings in the following section.

1. From the Volumes menu select the volumes which you are mapping to hosts and right click. Then, click **Map to Host or Host Cluster**, as shown in Figure 6-21 on page 111.

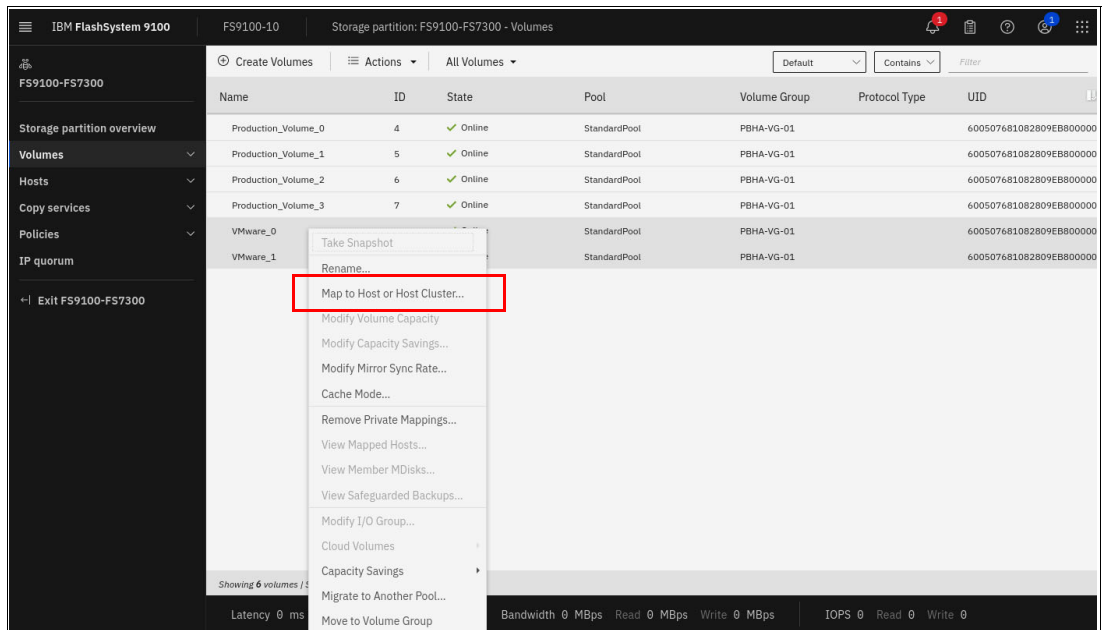


Figure 6-21 Create volume mappings to host

2. The Create mapping wizard opens. Select your preferred options and the host or host cluster to which you are mapping volumes and click **Next** as shown in Figure 6-22.

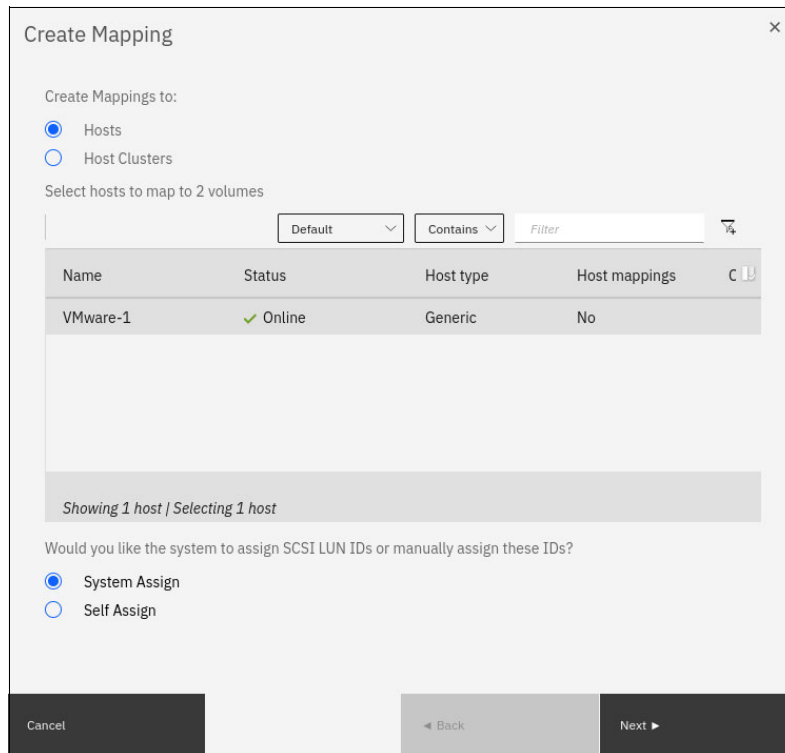


Figure 6-22 Create mapping wizard

- Review the volumes to be mapped and click **Map Volumes** as shown in Figure 6-23 on page 112.

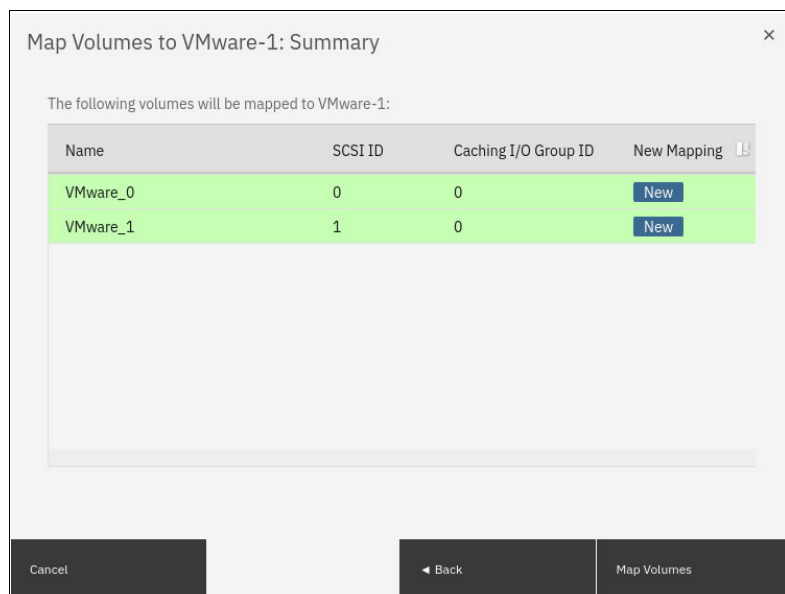


Figure 6-23 Select volumes to map

- The Storage Partition overview page now shows hosts online, as shown in Figure 6-24

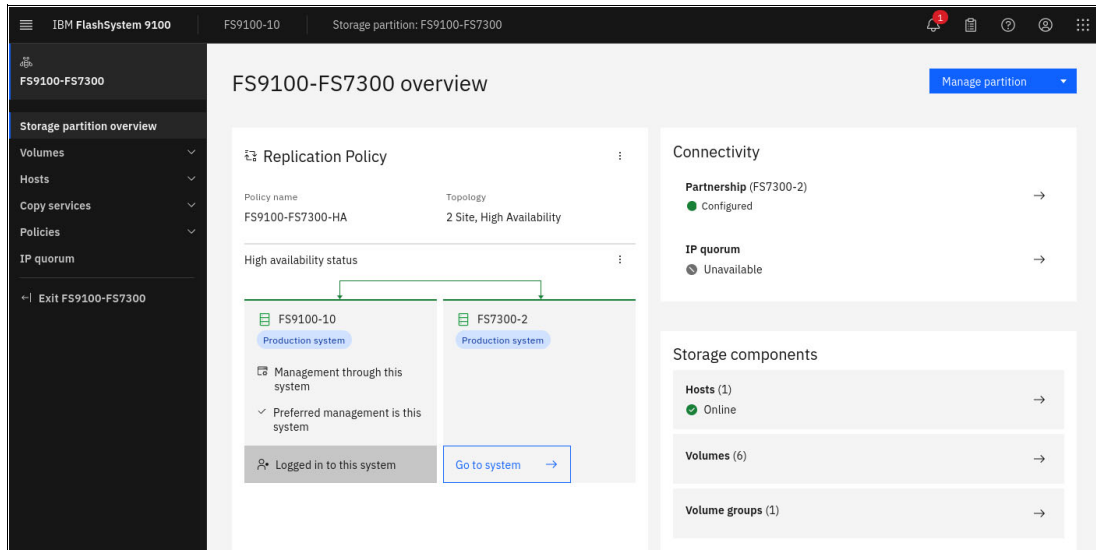


Figure 6-24 Storage Partition overview page

6.4 Migrating storage partitions between systems

IBM Storage Virtualize supports the non-disruptive migration of storage partitions between systems. This enables a partition to be migrated from one Storage Virtualize system to another without any application downtime. Migrating a storage partition requires both Fibre Channel and IP connectivity between the source and target storage systems.

Storage partitions bring a new level of flexibility to Storage Virtualize by enabling the migration of both the frontend and backend storage components. With a single command, you can seamlessly migrate all underlying storage, along with associated hosts, volumes, and host-to-volume mappings, to a new system.

This process involves updating storage paths for attached hosts. Fortunately, multipathing drivers handle this automatically, ensuring a non-disruptive and transparent migration with no impact on applications or users. This highly simplifies the decommissioning of old equipment.

Note: Migrating storage partitions requires appropriate SAN-zoning between systems to be migrated.

Storage partition migration use cases

The following are examples of use cases of storage partition migration:

- ▶ Non-disruptive upgrade of system hardware from older systems to newer regardless of model. For example, upgrading from FS5200 to FS7300.
- ▶ Load-balancing by migrating storage partitions from overloaded systems to other systems.

Migration process overview

Migrating a storage partition is a very simple two step migration process consisting of:

1. Exchange certificates between the two systems and create storage pool links
2. Run single command to *change* partition to the new system.

The CLI `svctask chpartition` is the sole CLI-command to trigger a migration. This makes initiating the migration simple and uncomplicated and direct.

Storage Virtualize provides event driven confirmation steps when user validation is needed during migration. These events are provided on the migration source as well as on the migration target system during the migration process. Examples of such events are:

- ▶ Check multipath on hosts before final path switch.
- ▶ Confirm deletion of original copy.

When data synchronization has completed, the hosts will see a new set of paths to identical volumes, now on a different and new storage system.

CLI procedure to migrate storage partitions

Flash grid is new in Storage Virtualize version 8.7.0, however storage partitions have been available since 8.6.3. Migration of storage partitions has also been available since 8.6.3.

For Storage Virtualize version 8.7.0 partitions can be migrated between systems that are members of the same Flash Grid, or can be migrated between systems that are not configured in a Flash Grid. Migration of storage partitions in 8.7.0 is only supported on systems that support Flash Grid.

Migration enables relocation of storage partitions from a source system to a different system location. As a consequence of which, the following flow of events occurs:

- ▶ All the objects that are associated with the storage partition are moved to the migration target storage system.
- ▶ Host I/O is served from the migration target storage system once the migration completes.
- ▶ At the end of the migration process, the storage partition and all of its objects are removed from the source system.

Note: Storage Virtualize 8.7's initial release does not include Storage Partition Migration functionality within the graphical user interface (GUI). This feature is expected to be available in a future update. You can still use CLI for migration tasks.

Prerequisites

Before you can use non-disruptive Storage Partition Migration function, ensure that the following prerequisites are met:

- ▶ Review the high availability requirements to ensure that the storage partition supports being migrated and the host operating systems support this feature, see [Planning High Availability](#)
- ▶ Confirm that both systems are members of the same Flash Grid, or that neither system is a member of a Flash Grid, and that both systems meet the requirements for Flash Grid.
- ▶ Use the `lpartnershipcandidate` command and make sure that both source and target systems are correctly zoned and are visible to each other.
- ▶ SAN-zoning requirements are the same as for HyperSwap configurations which require dedicated ISL-connections for public traffic (hosts communication) and for private traffic (node-node communication).
- ▶ Although the systems are visible to each other, a partner system can have multiple storage pools that are able to host the migrated storage partition. In that case establish suitable storage pools between systems on the source and target systems. For more information, see `chmdisk` command.

- ▶ Use the [lspartnership](#) command and ensure whether the source and a specific system are already in partnership.
- ▶ Make sure that both systems have their certificates added to the truststore of each other, with the REST API usage enabled. For more information, see [mktruststore](#) command.

Limits and restrictions

Refer to IBM Documentation for the current limits and restrictions exist for automated storage partition migrations.

Procedure

Perform the following steps:

1. Run the [chpartition](#) command with the `-location` option to migrate the storage partition to its required system location.
2. The following example initiates a migration of storage partition to the designated location system:

```
chpartition -location <remote_system> mypartition1
```

3. To check the migration status, run [lspartition](#).
4. After successfully migrating the storage partition's data and configuration to the target system, an event will notify the storage administrator. This event verifies that the affected hosts have established new paths to the volumes on the target system. Once you confirm this by fixing the event, host I/O operations for the storage partition will automatically switch to using the paths on the target storage system. This event is raised and fixed at source storage system.
5. To finalize the migration process, another event will prompt the storage administrator to remove the copy of the data and configuration from the source storage system.

Important: Before confirming this event by fixing it, it is crucial to verify the performance of the storage partition on the target system. This ensures a smooth transition and optimal performance after migration. This event is raised and fixed at the target storage system.

6. An informational event at the target storage system marks the completion of the storage partition migration. This event is raised at target storage system.

You can monitor the progress of the migration, including the amount of data remaining to be copied, using the [lsvolumegroupreplication](#) command.

You can monitor the migration using `migration_status` field that is shown by [lspartition](#) command indicating that there is no migration activity active or queued for that storage partition.

For more information, see [chpartition](#) command to automate the procedure of migration of storage partitions. This automates intermediate steps, such as setting up Fibre Channel partnerships between the systems.

A ongoing storage partition migration can be aborted by specifying a new migration location that uses `-override` option. The migration to the new location gets queued behind existing queued migrations, if any.

For more information of the storage partition migration procedure see [IBM Documentation web page](#)



Managing policy-based high availability

In this chapter we describe how to manage and monitor a policy-based HA environment.

We begin by examining a pre-existing policy-based HA configuration to gain insights into the available settings and current replication status. Next, we delve into common operations for volumes, hosts (including optimized data path management for long distances), and partition management. We explore various migration options, such as migrating existing data to a policy-based HA environment or migrating policy-based HA protected data to a different storage system. We conclude by exploring the unique considerations for snapshots within a policy-based HA environment.

This chapter has the following sections:

- ▶ “Evaluate the current status of policy-based HA” on page 118
- ▶ “Volume management” on page 123
- ▶ “Host management” on page 134
- ▶ “Partition management” on page 136
- ▶ “Migration options for policy-based HA partitions” on page 139
- ▶ “Snapshots and policy-based HA” on page 140

7.1 Evaluate the current status of policy-based HA

To gain a preliminary understanding of which policy-based replication services are configured on your system, you can utilize either the GUI or CLI interface. Both options are presented below.

7.1.1 Evaluate the current environment using GUI

Replication partnerships define the relationships between this storage system and others. To view these partnerships, navigate to **Copy Services** → **Partnerships** on the storage system. This will display all existing replication partnerships for this particular system. Figure 7-1 depicts a single configured partnership with a FS9100 system.

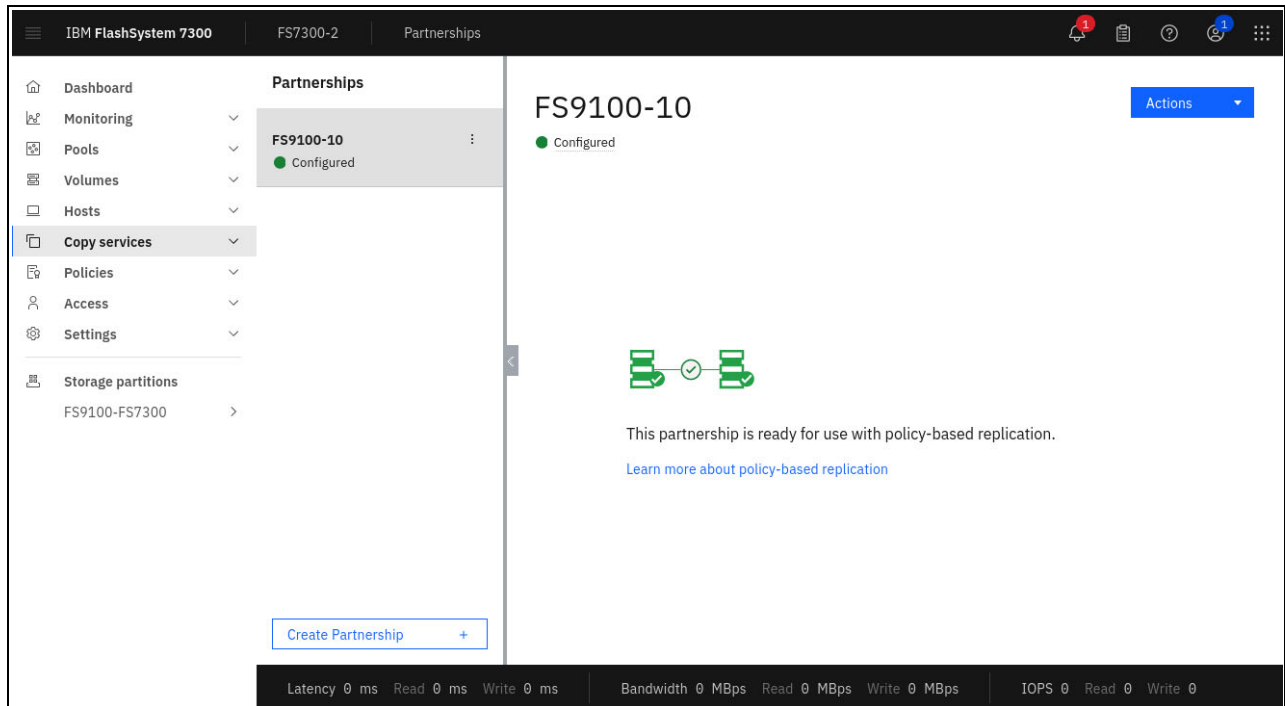


Figure 7-1 Replication partnership

Replication policies define replication details, including source and target system IDs, location names and IDs, RPO alerts (if any), IO group selections, and the replication topology.

These policies can be verified by opening **Policies** → **Replication Policies** as shown in Figure 7-2 on page 119.

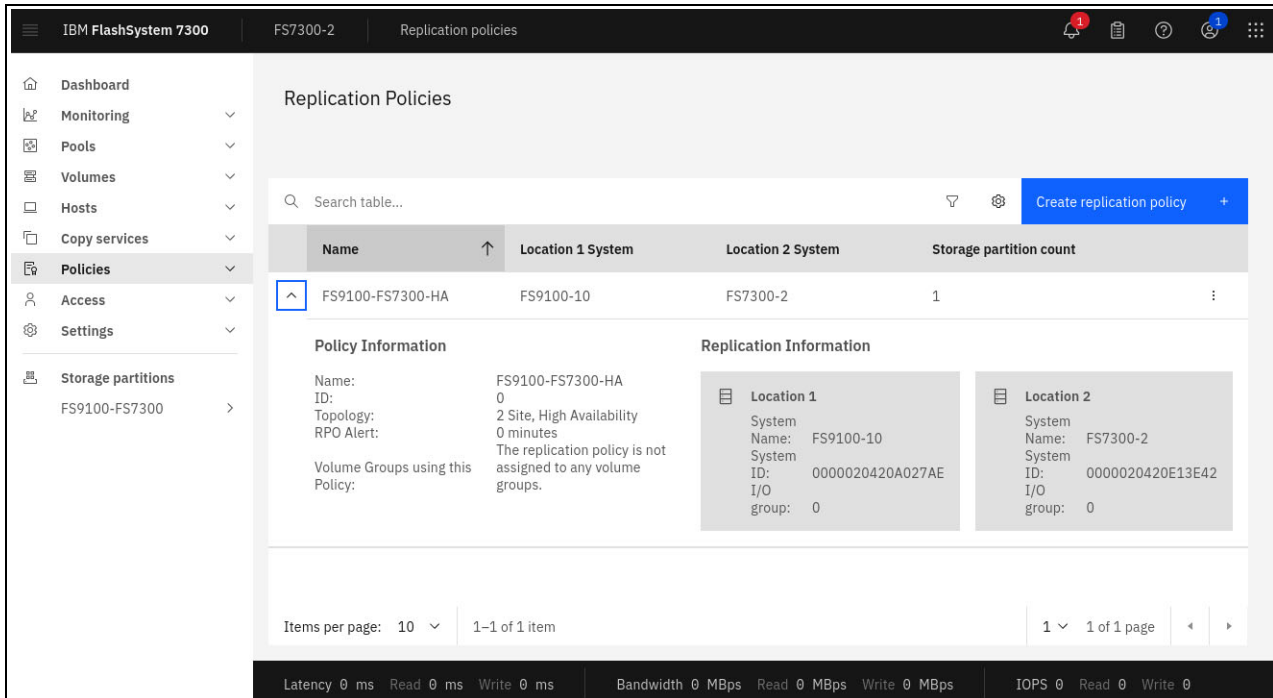


Figure 7-2 Replication policy for 2-site high availability (policy-based HA)

Partitions are used in policy-based HA to set a common management environment for hosts, volumes, volume groups, and related snapshots. The data copy direction between both site a partition setting. Partitions are available under the **Storage partitions** menu item as shown in Figure 7-3.

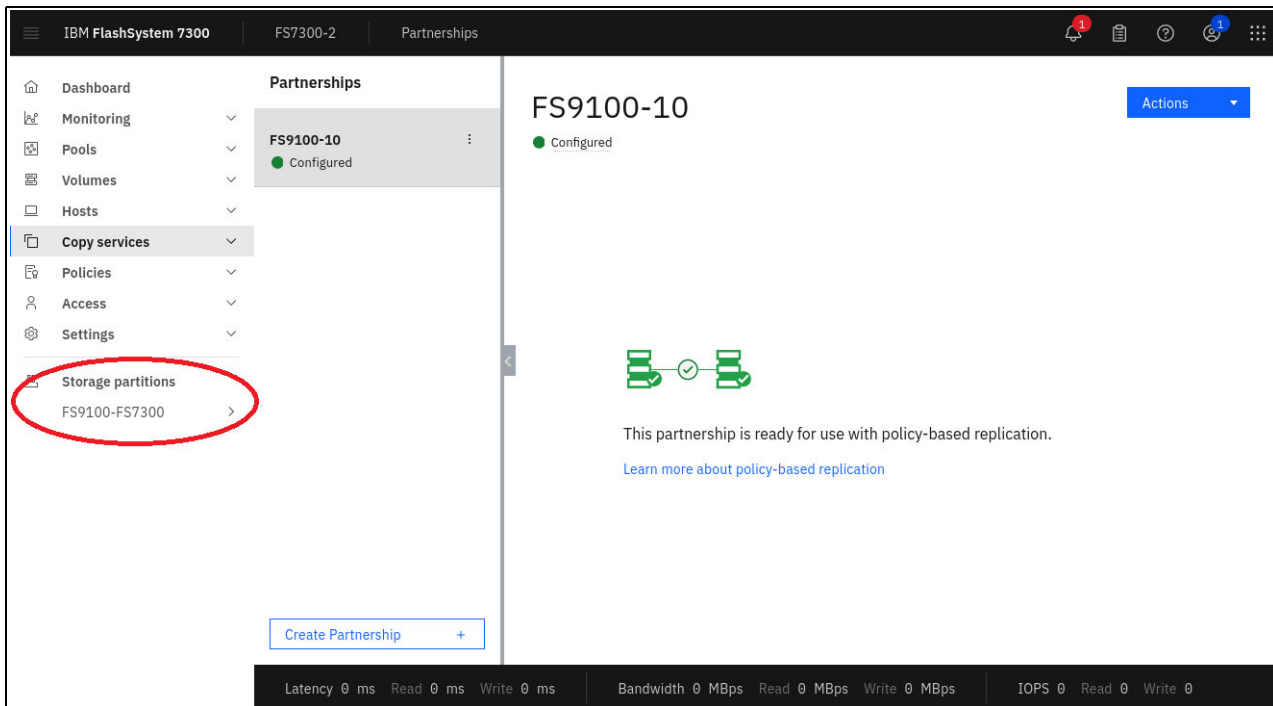


Figure 7-3 Configuration entry point for partitions

Partition details like the following are shown (Figure 1-4 on page 6):

- ▶ Assigned replication policy
- ▶ Current replication status
- ▶ Active management system (which sets the current copy direction)
- ▶ Hosts and volumes
- ▶ Quorum device status

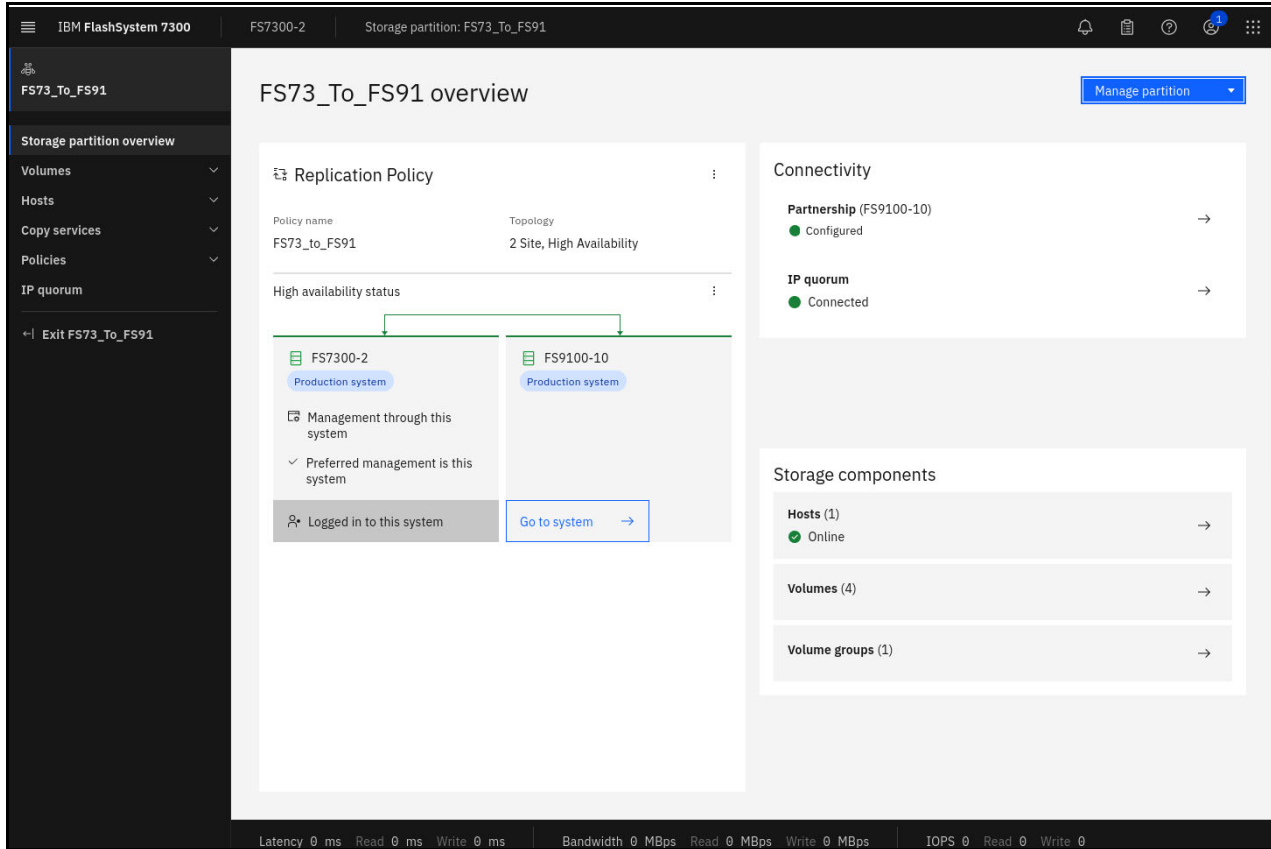


Figure 7-4 Partition configuration: Overview

For all partition configuration tasks, use the **Partitions** menu item located on the left side of the screen. Standard menu options for general operations reappear upon selecting **Exit** from the left-hand side menu.

7.1.2 Evaluate the current environment using CLI

Check the status of the partnership between the storage systems and verify that the partnership is fully configured as shown in Example 7-1.

Example 7-1 Verify remote copy partnerships

```
BM_FlashSystem:FS7300-2:Team4>lspartnership
id          name      location partnership  type cluster_ip
event_log_sequence link1 link2 link1_ip_id link2_ip_id
0000020420E13E42 FS7300-2 local
0000020420A027AE FS9100-10 remote  fully_configured fc
```

All policy-based replication activities require a policy, which defines the general replication settings. Identify the existing policies and check for the number of partitions, using those policies as shown in Example 7-2.

Example 7-2 List available replication policies

```

IBM_FlashSystem:FS7300-2:Team4>lsreplicationpolicy
id name          rpo_alert topology  volume_group_count location1_system_name
location1_iogrp_id location2_system_name location2_iogrp_id partition_count
0 FS73_to_FS91 0          2-site-ha 0          FS7300-2          0
FS9100-10      0          1

```

The definition for source and target systems, the topology and the number of partitions using this policy are shown in the policy details in Example 7-3.

Example 7-3 Replication policy in detail

```

IBM_FlashSystem:FS7300-2:Team4>lsreplicationpolicy 0
id 0
name FS73_to_FS91
rpo_alert 0
topology 2-site-ha
volume_group_count 0
location1_system_id 0000020420E13E42
location1_system_name FS7300-2
location1_iogrp_id 0
location2_system_id 0000020420A027AE
location2_system_name FS9100-10
location2_iogrp_id 0
partition_count 1
IBM_FlashSystem:FS7300-2:Team4>

```

The partitions are managed by the policies. To verify this, check the policies used and the current status of the partitions as shown in Example 7-4. Note the setting for the active management system. This system coordinates policy-based HA replication within the partition. In our example, the FS7300-2 is the active management system.

Example 7-4 List of partition and one partition in detail

```

IBM_FlashSystem:FS7300-2:Team4>lspartition
id name          preferred_management_system_name active_management_system_name
replication_policy_id replication_policy_name location1_system_name
location1_status location2_system_name location2_status host_count
volume_group_count ha_status link_status desired_location_system_name
migration_status draft draft_volume_group_count draft_host_count uuid
0 FS73_To_FS91    FS7300-2          FS7300-2          healthy
0          FS73_to_FS91      FS7300-2          established
FS9100-10    healthy          1          1
synchronized no 0
0          476A7EFA-25B1-573E-92F0-5A01FD73841F
1 Non-HA_Partition
1          1
no 0          0
D6C0672B-4141-5CC9-B9E3-A443A17D8E74
IBM_FlashSystem:FS7300-2:Team4>lspartition 0
id 0
name FS73_To_FS91
preferred_management_system_id 0000020420E13E42
preferred_management_system_name FS7300-2
active_management_system_id 0000020420E13E42

```

```

active_management_system_name FS7300-2
replication_policy_id 0
replication_policy_name FS73_to_FS91
location1_system_id 0000020420E13E42
location1_system_name FS7300-2
location1_status healthy
location2_system_id 0000020420A027AE
location2_system_name FS9100-10
location2_status healthy
host_count 1
host_offline_count 0
volume_group_count 1
volume_group_synchronized_count 1
volume_group_synchronizing_count 0
volume_group_stopped_count 0
ha_status established
link_status synchronized
desired_location_system_id
desired_location_system_name
migration_status
draft no
draft_volume_group_count 0
draft_host_count 0
location1_total_object_count 6
location2_total_object_count 6
merge_target_partition_id
merge_source_partition_id
uuid 476A7EFA-25B1-573E-92F0-5A01FD73841F
user_action_sequence_number
user_action_type
IBM_FlashSystem:FS7300-2:Team4>

```

Optimized data path management for long distances relies on host location settings. If host locations are not defined, the replication is originated from the system that receives the write from the host, and all host traffic (read and write) are managed by the active management system. In this scenario, the copy source volumes reside at the same site as the active management system. As shown in Example 7-5 on page 122, the FS7300-2 will be the default access point for all hosts without a location setting within this partition.

Example 7-5 Partition: Identification of active management system

```

IBM_FlashSystem:FS7300-2:Team4>lspartition 0 | grep active_management
active_management_system_id 0000020420E13E42
active_management_system_name FS7300-2
IBM_FlashSystem:FS7300-2:Team4>

```

Changing the active management system immediately reverses the copy direction between the storage systems and designates the new system as the default storage system for host access, as illustrated in Example 7-6.

Example 7-6 Partition: Change active management system

```

IBM_FlashSystem:FS7300-2:Team4>chpartition -preferredmanagementsystem FS9100-10 0
IBM_FlashSystem:FS7300-2:Team4>lspartition 0 | grep active_management
active_management_system_id 0000020420A027AE
active_management_system_name FS9100-10

```



```
IBM_FlashSystem:FS7300-2:Team4>
```

Policy-based HA configurations can leverage optimized data paths through host-level "location" settings. In the absence of a location setting, a host will always access the active management system for this partition, regardless of its physical location, for all read and write I/O. Changing the active management system will impact both the copy direction for all partition volumes and the default access point for hosts without a location setting. Public ISLs will be used for this redirected traffic.

With a defined host location, read and write I/O are performed (if possible) directly with the storage system assigned to your location. This system replicates written data to the remote site, and the active management system acknowledges both copies, ensuring data consistency. Consequently, all data access can be done locally, minimizing additional data traffic. The private ISL is used for this efficient replication process.

Check the host configuration and the location parameter as shown in Example 7-7.

Example 7-7 Identify storage location and verify the host location

```
IBM_FlashSystem:FS9100-10:Team4>1shost
id name      port_count iogrp_count status site_id site_name host_cluster_id
host_cluster_name protocol owner_id owner_name portset_id portset_name
partition_id partition_name draft_partition_id draft_partition_name
ungrouped_volume_mapping location_system_name
0 PB_HA_1 1      4          online
scsi          64      portset64  0          FS9100-FS7300
no           FS9100-10
1 PB_HA_2 1      4          online
scsi          64      portset64  0          FS9100-FS7300
no           FS7300-2
IBM_FlashSystem:FS9100-10:Team4>1shost 0 | grep location
location_system_name FS9100-10
IBM_FlashSystem:FS9100-10:Team4>
```

7.2 Volume management

In this section we discuss how to manage volumes.

7.2.1 Create a new volume in a partition

You can easily create new, empty volumes within a partition. The system will automatically create corresponding policy-based HA volumes at the remote site, including all host assignments. The steps are as follows:

1. **Navigate to Volumes:** Within your partition, select **Volumes** → **Volumes**.
2. **Create volumes:** Click the **Create Volumes** button.
3. **Configure volume details:** Follow the wizard to create one or multiple volumes according to your needs. Be sure to select the appropriate volume group.
4. **Assign host mapping:** In a second step, assign the host mapping for your newly created volume(s).

The same steps are also available in the CLI. In our example, as shown in Example 7-8, we create a single new volume and assign the volume to the appropriate volume group.

Example 7-8 Create new policy-based HA volume

```
IBM_2145:SVC_SA2:superuser>svctask mkvolume -name PB_HA_Add_12 -pool 0 -size
42949672960 -unit b -volumegroup 0
```

As already discussed, we must assign the new volume to both appropriate hosts. See Example 7-9.

Example 7-9 Assign policy-based HA volume to both hosts

```
IBM_2145:SVC_SA2:superuser>svctask mkvdiskhostmap -force -host 2 4
IBM_2145:SVC_SA2:superuser>svctask mkvdiskhostmap -force -host 3 4
```

7.2.2 Delete a volume in a partition

Deleting a policy-based HA volume permanently removes the volume definition and all associated data on both storage systems. This action also removes any host mappings assigned to the volume. The process for deleting a policy-based HA volume is similar to deleting a traditional volume within the partition as shown in Figure 7-5.

Note: A volume deletion for a volume with active snapshots will not be physically removed from the storage system, as long as a dependent snapshot exists.

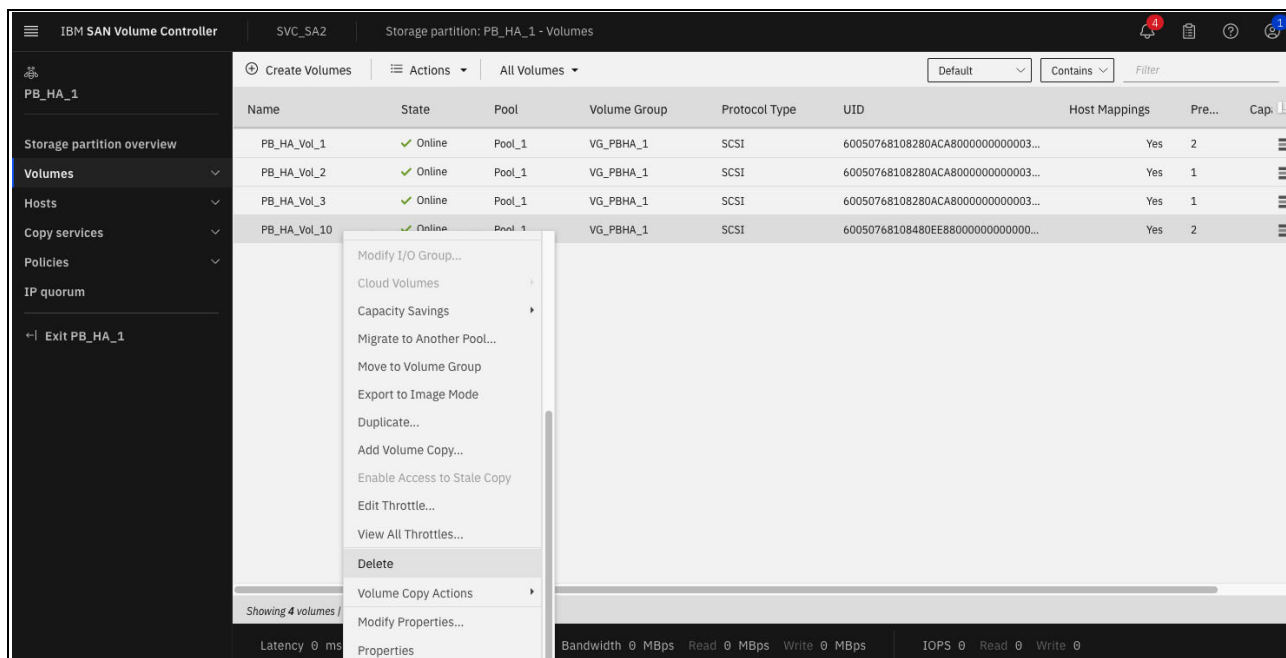


Figure 7-5 Delete a policy-based HA volume

Follow the deletion process, select the removal although there are already host assignments in place as shown in Figure 7-6.

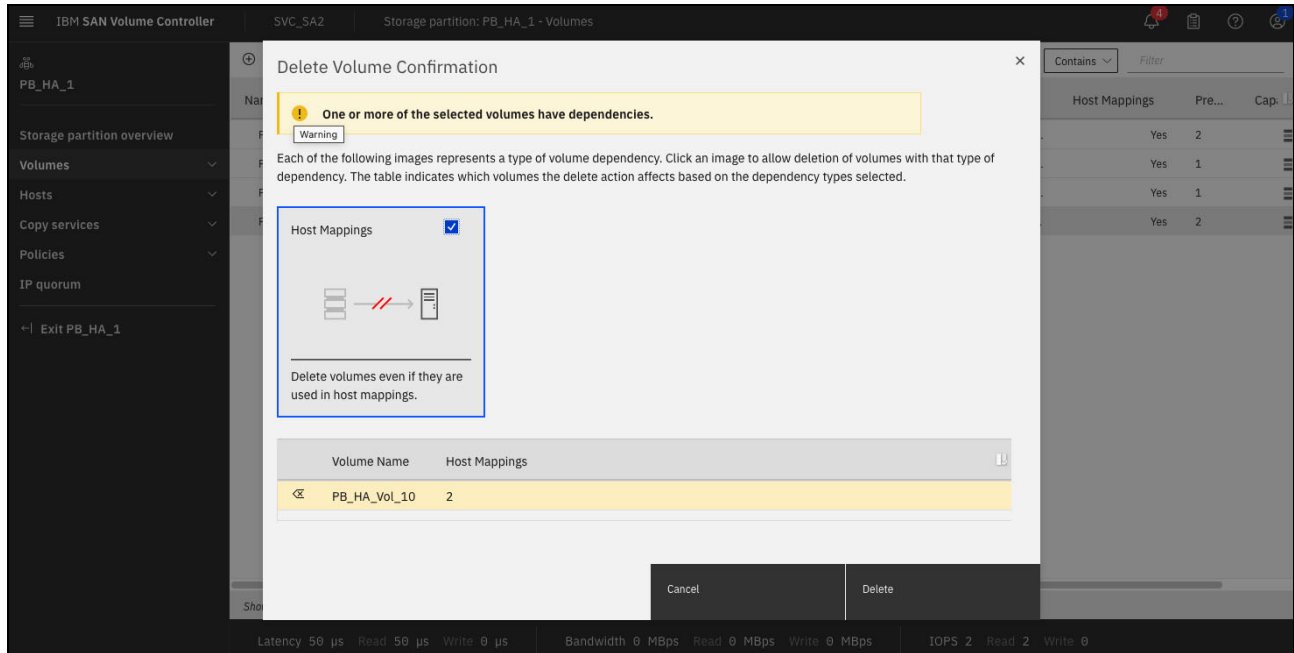


Figure 7-6 Delete a policy-based HA volume and confirm host mapping removal

The volume and data deletion and the host unassignment will be executed on both storage systems without additional checks.

The volume deletion can also be done using the CLI as shown in Example 7-10 on page 125.

Example 7-10 Policy-based HA volume deletion

```

IBM_2145:SVC_SA2:superuser>lsvdisk | grep VG_PBHA_Add
0 PB_HA_Vol_10 0 io_grp0 online 0 Pool_1
40.00GB striped many many 60050768108480EE880000000000028 2
...
4 PB_HA_Vol_12 0 io_grp0 online 0 Pool_1
40.00GB striped many many 60050768108480EE880000000000033 2
1 not_empty 1 no 0 0
Pool_1 no no 4
PB_HA_Vol_12 0 VG_PBHA_Add scsi no 0
no 0 no
IBM_2145:SVC_SA2:superuser>rmvolume -removehostmappings 4

```

The single command successfully removed all host definitions, all host mappings and all data for this volume at both sites.

7.2.3 Add data volumes to a partition and merge partitions

Adding volumes with existing data to an already existing partition is a multi-step approach. The process is supported by a GUI wizard, so it makes life easier to perform the initial process using the GUI. However, *it is important to note that these volumes must first be migrated to a temporary partition with identical properties as the target partition.* This ensures a smooth merge using the partition merging feature.

So, the existing volumes must be summarized in a temporary partition using exactly the same properties as the already existing partition. The partition merge combines volume groups and

hosts from both partitions into a single entity. Importantly, existing volumes, their assignments within volume groups, host definitions, and host-volume access remain unchanged during the merge process.

There are multiple prerequisites:

- ▶ The appropriate zoning must be configured correctly for both partitions.
- ▶ Volume and host names can only be used in one partition.

In our example, we will use only the GUI method, as the GUI significantly simplifies the overall process.

1. We already have a single policy-based HA partition with three volumes and two hosts as shown in Figure 7-7 on page 126. All data is replicated to the remote site.

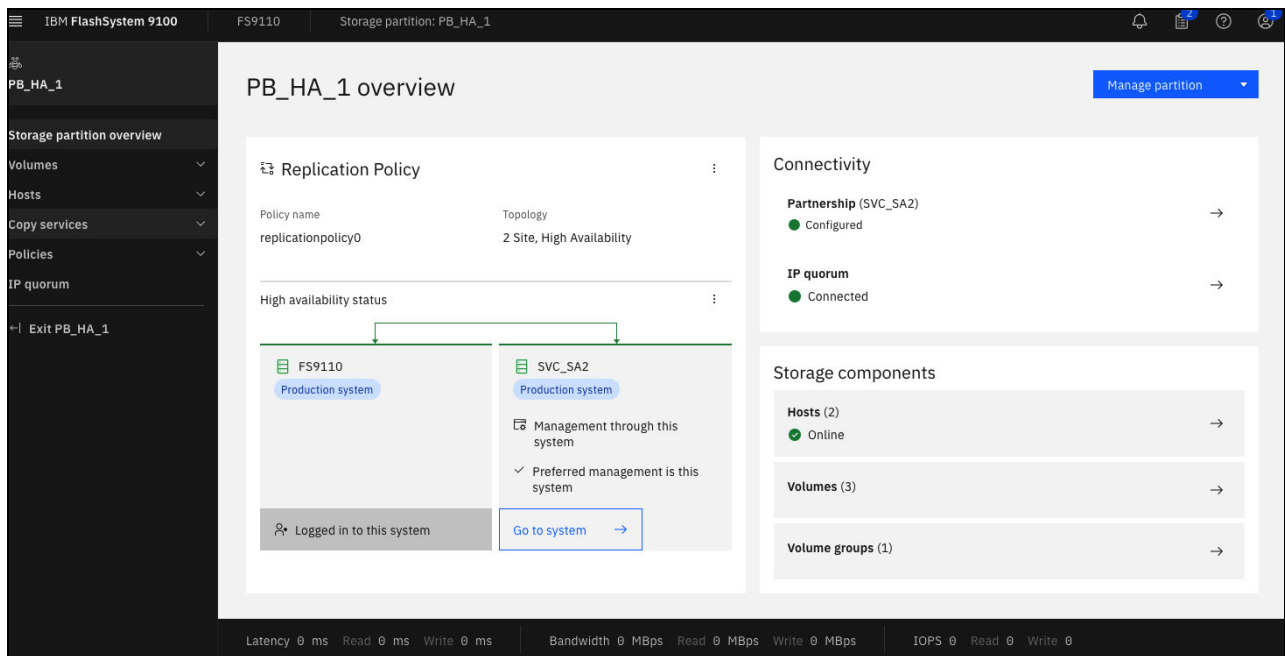


Figure 7-7 Original storage partition, replicated to the remote site

2. There are three active volumes, already replicated to the remote site as shown in Figure 7-8.

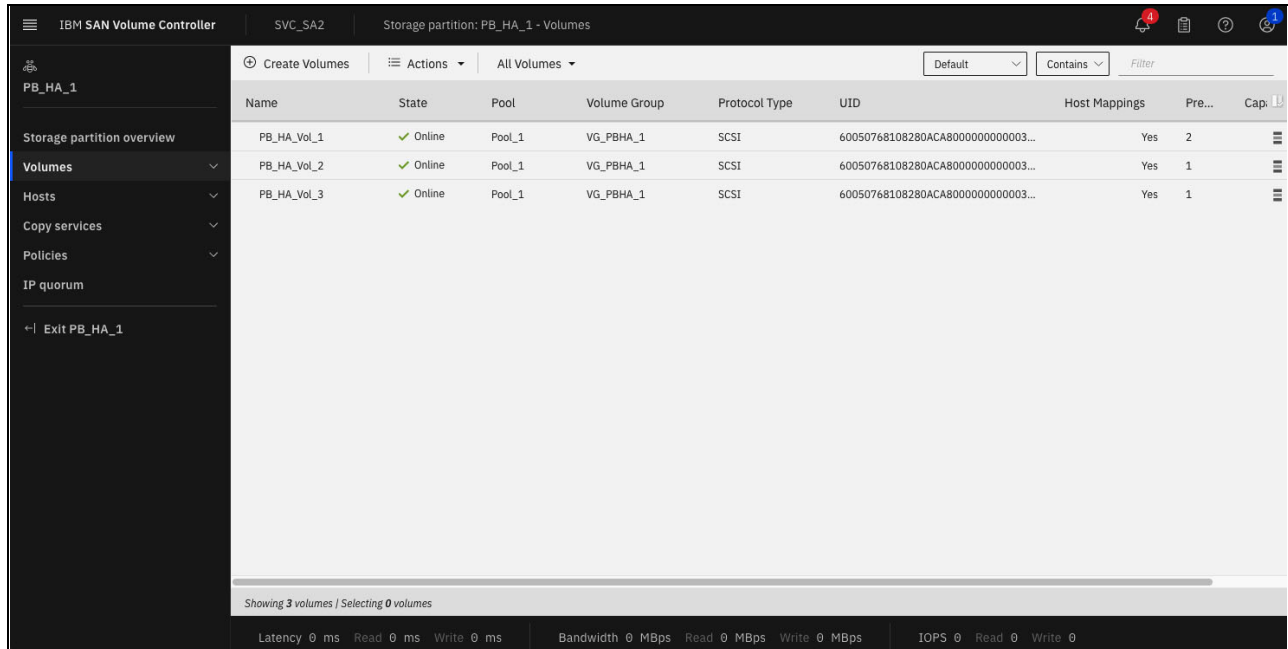


Figure 7-8 Check current policy-based HA volume status

In our example we demonstrate how to seamlessly add two existing volumes with data to the existing policy-based HA partition named PB_HA_1.

Here is a step-by-step breakdown:

- **Volume group preparation:** Create a new, common volume group to house the two existing volumes.
- **Temporary partition creation:** Establish a new partition and assign the newly created volume group during this process.

Refer to Figure 7-9 for a visual guide on creating a new storage partition.

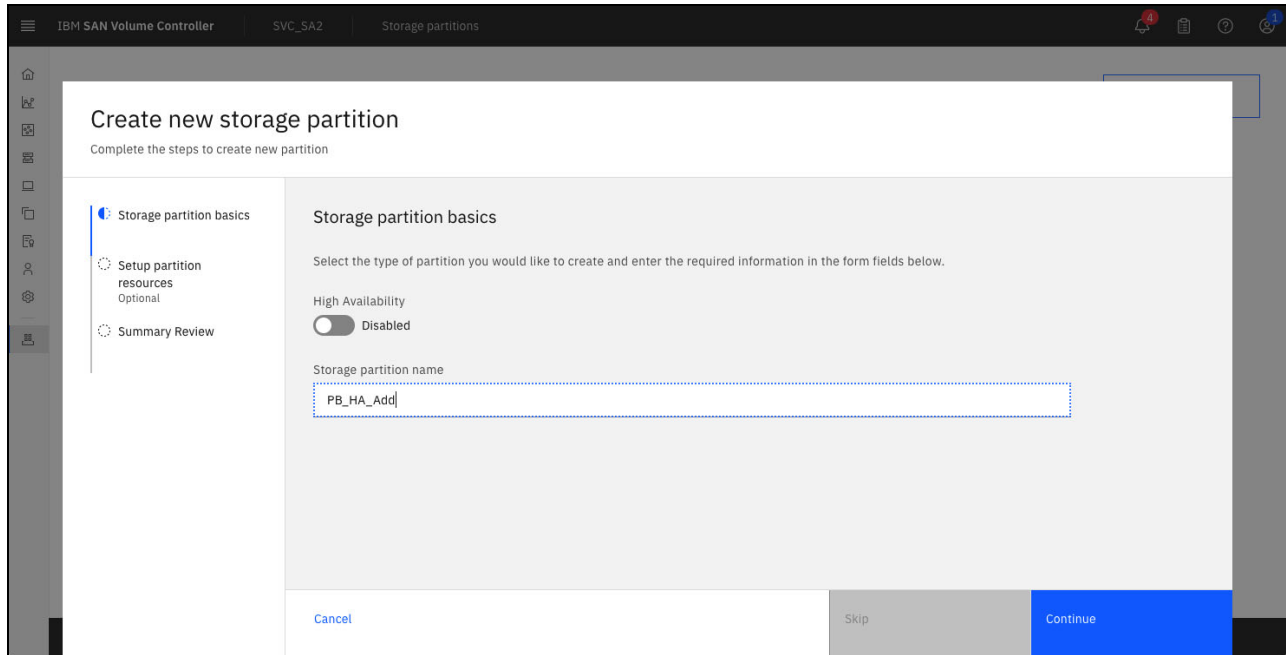


Figure 7-9 Create new storage partition

3. Next, we select **Select existing volume group** as shown in Figure 7-10 on page 128.

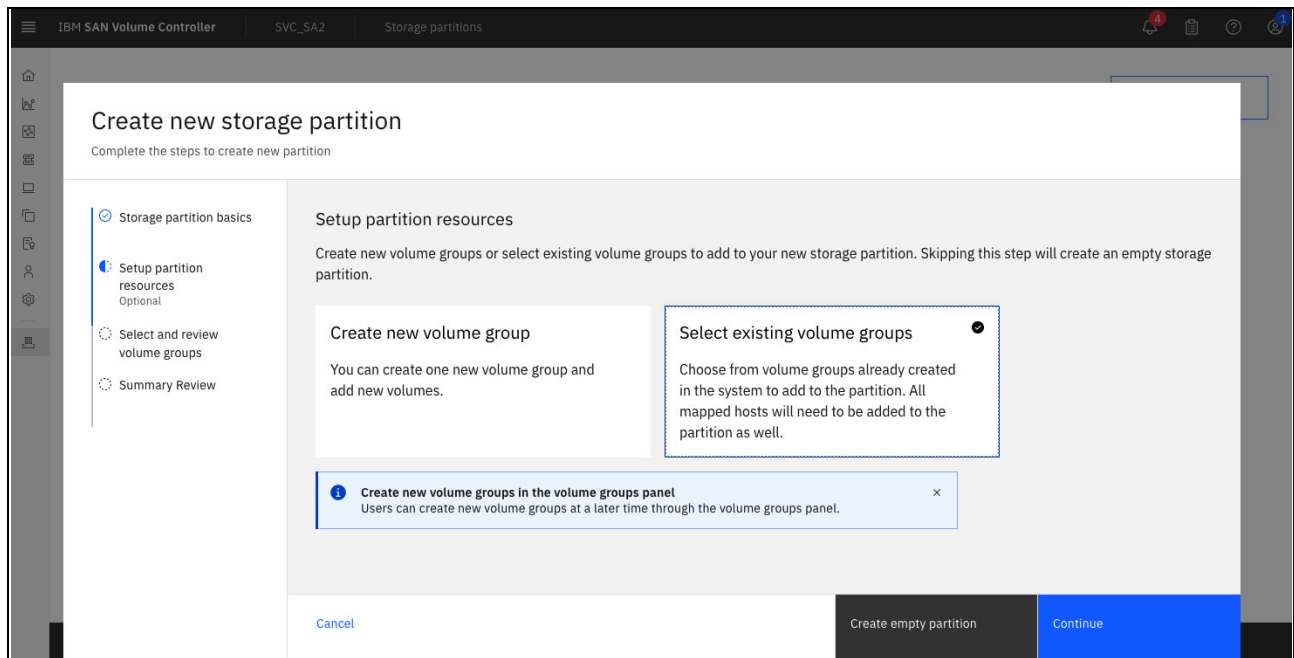


Figure 7-10 Create new storage partition and select volume group

4. Click **Select volume groups** as shown in Figure 7-11 on page 129.

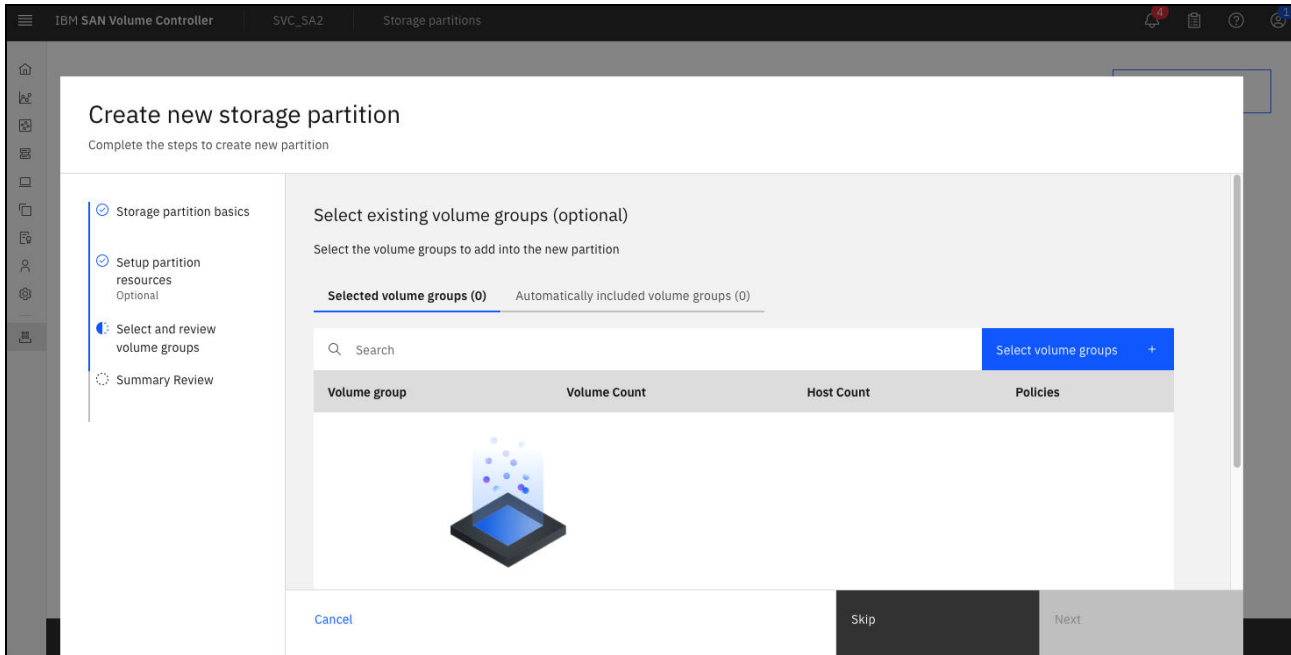


Figure 7-11 Create new storage partition and select volume group

5. Finally, select the appropriate volume group, follow the process as shown in Figure 7-12 on page 129. The system will guide you through the assignment process using a user-friendly GUI wizard.

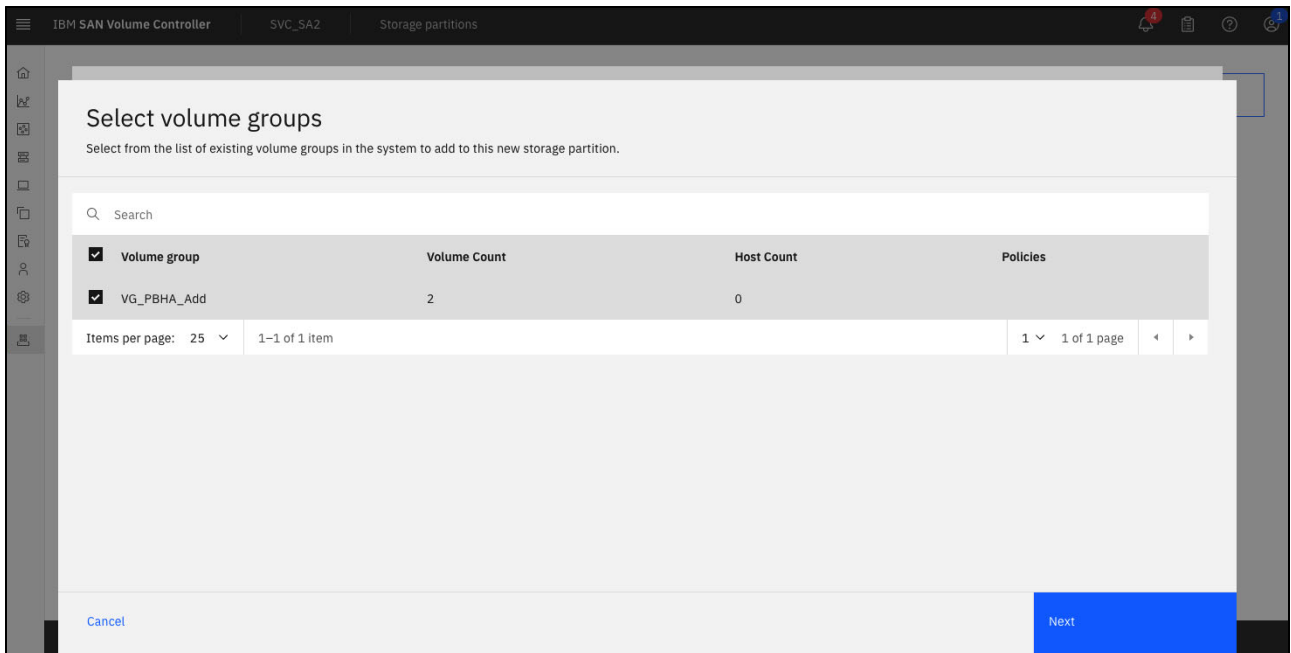


Figure 7-12 Create new storage partition and select volume group finish

6. We now get a new partition with two volumes and two hosts as shown in Figure 7-13 on page 130.

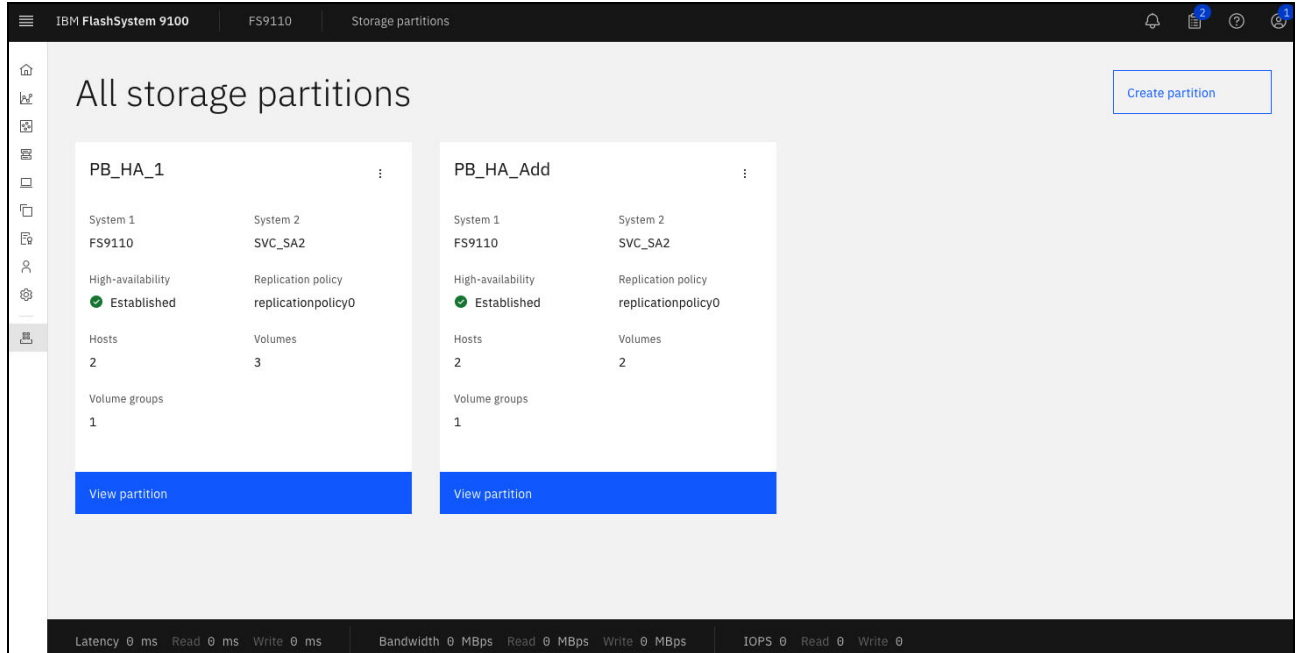


Figure 7-13 Check all storage partitions

We already created a new partition, the data volume are now managed by the partition. The data merge process requires exactly the same properties for both candidates. The original partition is running a 2-site high available configuration, whereas the newly created partition is running without the additional 2-site protection. Assigning the appropriate policy-based HA policy will eliminate this difference and automatically create the required volume copies on the recovery site.

7. Open the newly created partition and click **Select high availability replication** as shown in Figure 7-14.

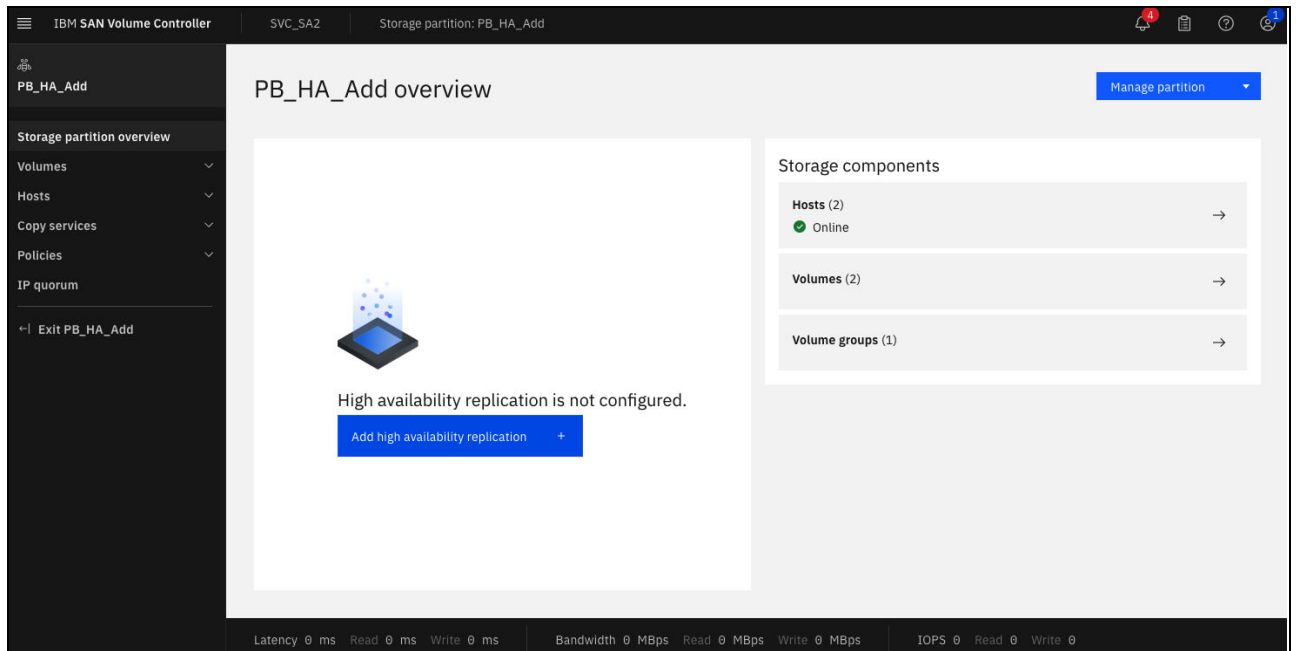


Figure 7-14 Partition: Check for details for the new partition

8. Follow the process and select the appropriate policy, set your preferred management system, and download the IP application (if not already done), link the pools. As shown in Figure 7-15, select the appropriate replication policy to activate the configured settings.

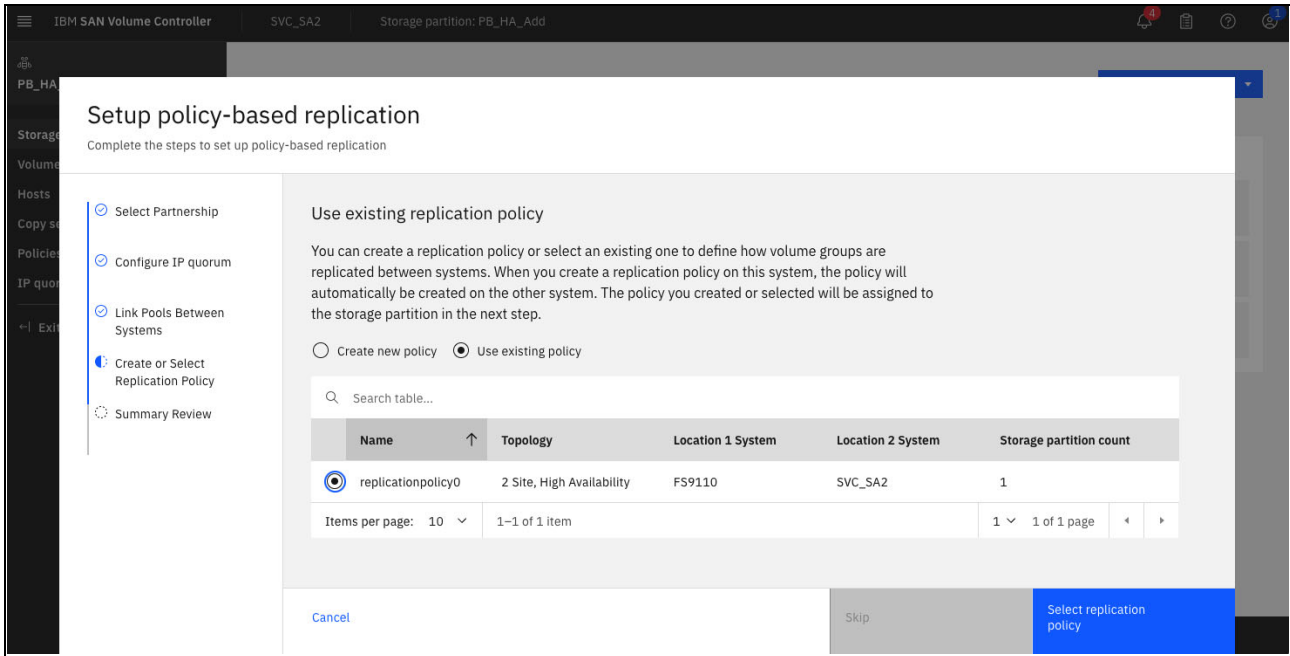


Figure 7-15 Partition: Select replication policy

9. Finalize the wizard and verify the new settings as shown in Figure 7-16. Note that the configuration should be identical for both partitions.

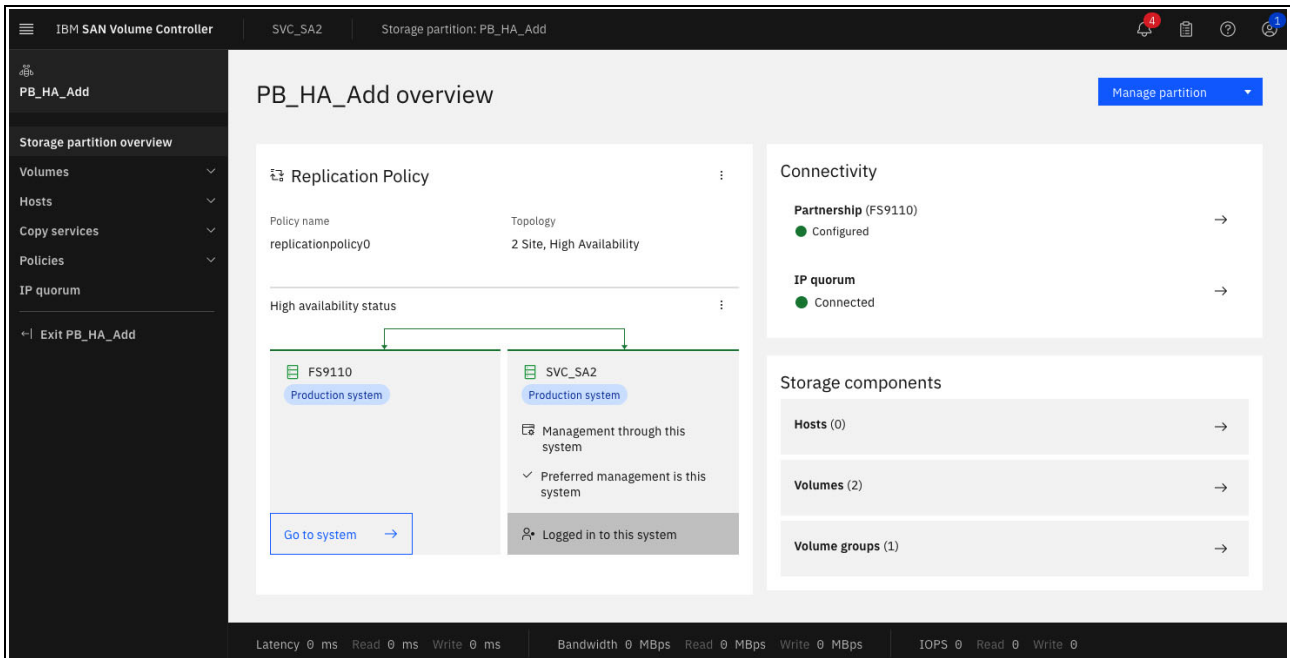


Figure 7-16 Partition: Review and finalize partition changes

10. As both partitions identical, we can initiate the partition merge. On the storage partition which should be merged (in our example “PB_HA_Add”, click the **Manage partition** button and select **Merge partition**, as shown in Figure 7-17.

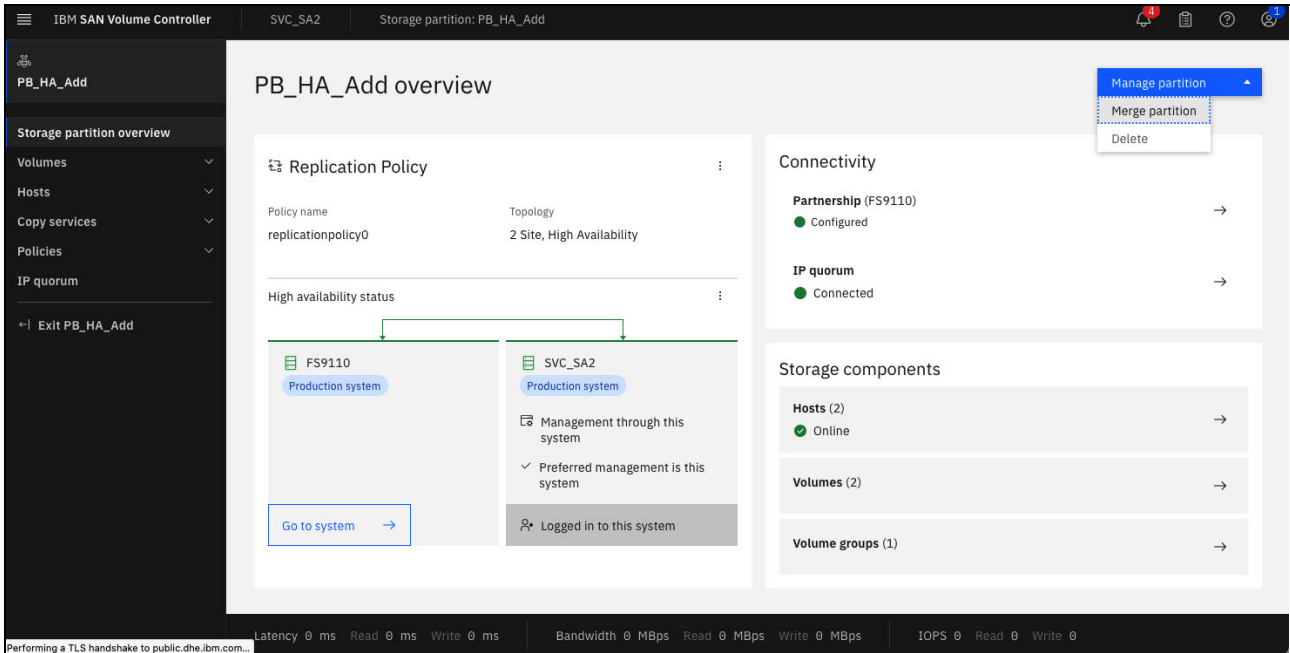


Figure 7-17 Start partition merge process

11. Follow the process and select the target partition as shown in Figure 7-18 on page 132.

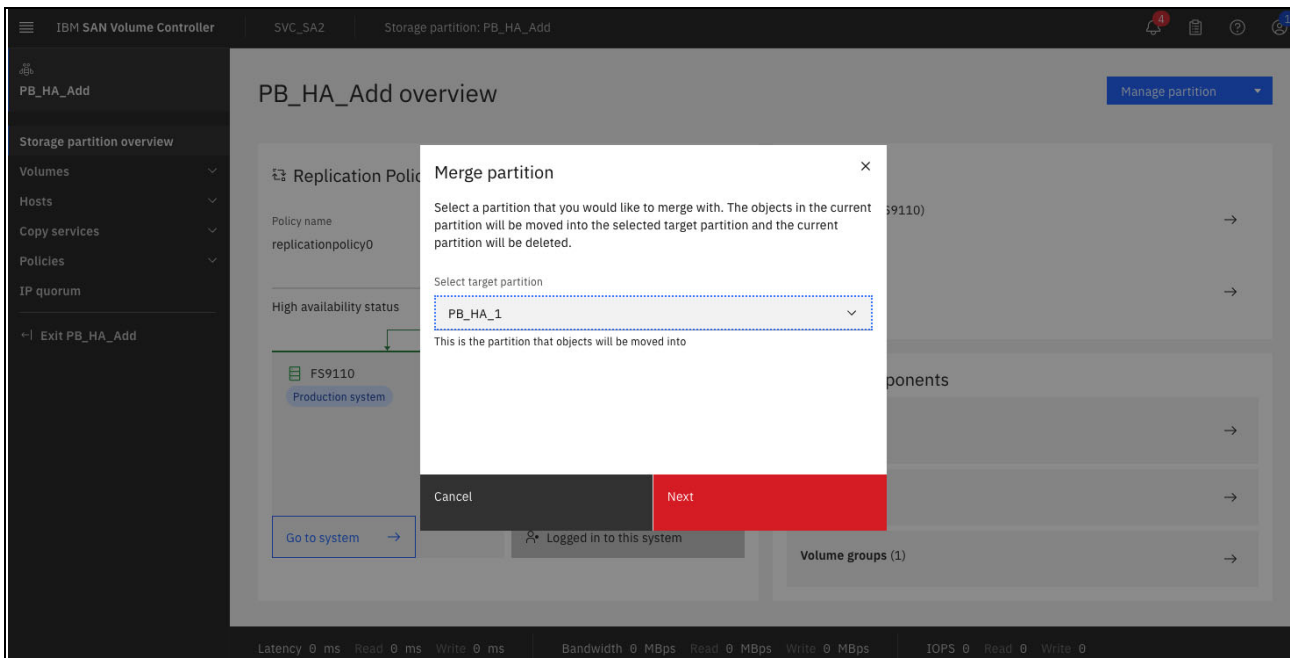


Figure 7-18 Set options for partition merge

12. Review the Merge configuration, as shown in Figure 7-19 on page 133, to make sure that the configuration meets your requirements.

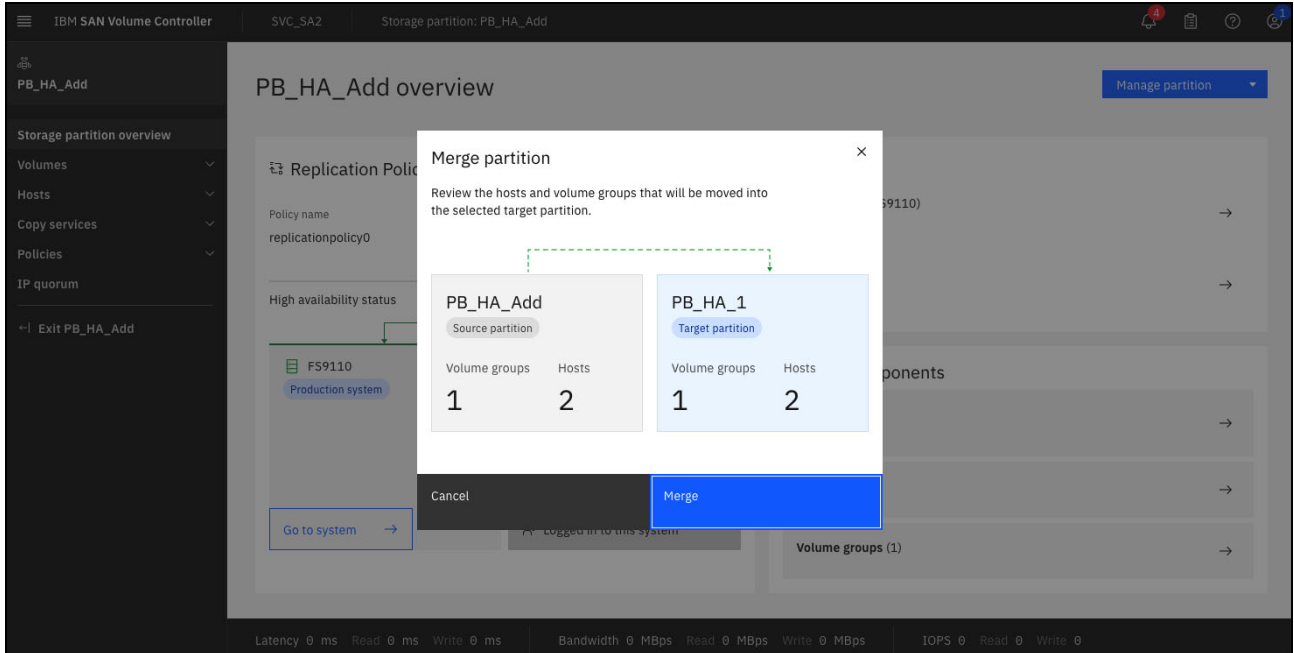


Figure 7-19 Review the partition merge settings and start the merge process

13. Finally, verify the new configuration (the partition now manages five volumes, four hosts). All volumes and hosts are available at both sites as shown in Figure 7-20.

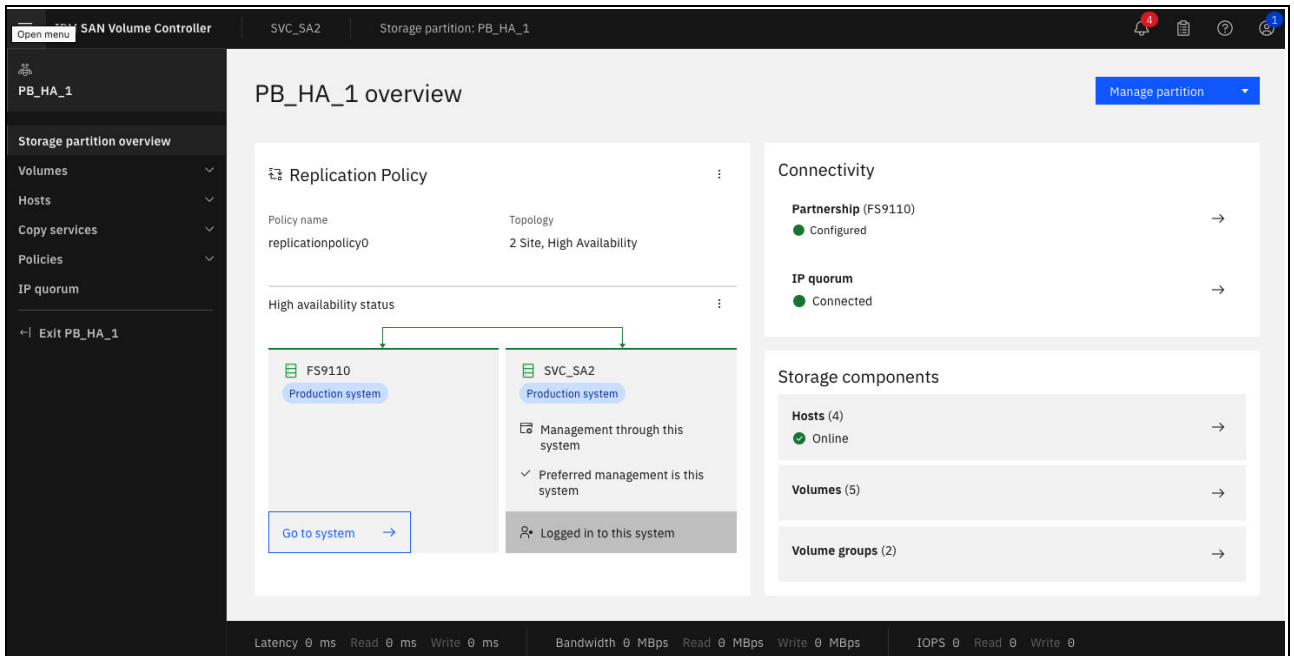


Figure 7-20 Verify the partition after merge

14. Verify the details like volume or host settings as shown in Figure 7-21 on page 134.

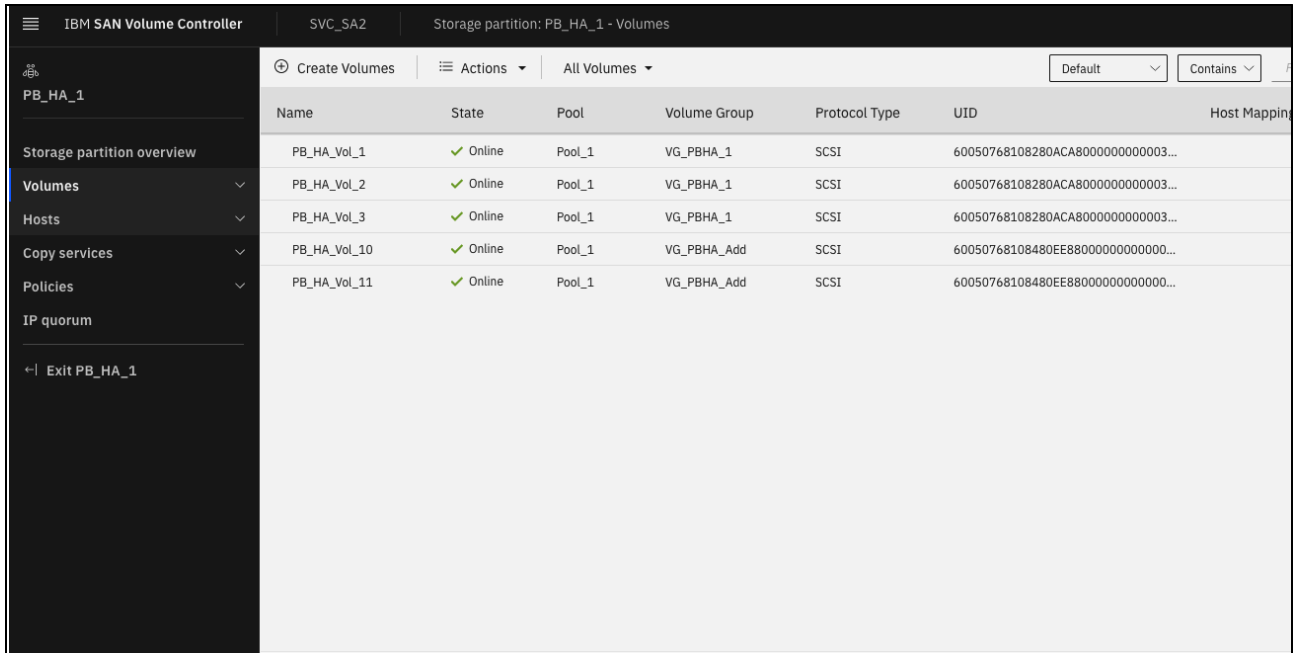


Figure 7-21 Check for policy-based HA volumes in detail after merge

We successfully assigned existing volumes to an existing policy-based HA partition using the merge process. The system automatically created the required host definitions and volumes at the target site. It merged all volumes, volume groups and hosts from two independent partitions in a single partition.

7.3 Host management

The appropriate host configuration may have an impact on the overall performance, as it defines the data path usage between both sites. Without setting up the host “location” parameter the host will always use the active management system for this partition for read and write access.

Setting the host “location” parameter to the local storage system optimizes the data flow and ensures always local read and write access at each site, which significantly reduces the long distance traffic between both sites.

The host "location" setting allows you to optimize data paths for geographically distant locations. Here is how to configure it:

- ▶ **Identify storage system names:** Locate the name of the storage system you're currently working on and the name of the remote storage system within your established partnership.
- ▶ **Set host location:** Based on the storage system names you identified, assign the appropriate location setting to your host.

In Example 7-11 showcases two storage systems. Their names, which also serve as location identifiers, are SVC_SA2 and FS9110.

Example 7-11 Identify storage locations - optional needed for host definition

```
IBM_2145:SVC_SA2:superuser>lssystem | grep name
```

```

name SVC_SA2
...
IBM_2145:SVC_SA2:superuser>lspartnership
id          name      location partnership      type cluster_ip
event_log_sequence link1 link2 link1_ip_id link2_ip_id
0000020421203BA2 SVC_SA2 local
0000020420A02B2A FS9110 remote  fully_configured fc

```

We are currently working on SVC_SA2; the remote system is FS9110. Those names can be used for setting up the location during the host creation or modification process.

7.3.1 Add hosts to partition

Create a new host within your storage partition by selecting **Hosts** → **Hosts**. Start the wizard by selecting the **Add Host** button. Once you've successfully created new hosts, you can proceed with assigning appropriate volumes to them. The following examples demonstrate the process of creating new volumes with location definitions and assigning them to your newly created hosts.

1. Create the new hosts as shown in Example 7-12.

Example 7-12 Create new hosts with location settings

```

svctask mkhost -fcwvpn 100000109B55XXXX -force -name SR650_111 -partition 0
-protocol fcscsi -location FS9110
svctask mkhost -fcwvpn 100000109B55YYYY -force -name SR650_112 -partition 0
-protocol fcscsi -location SVC_SA2

```

2. Assign the newly created volumes to the host or host cluster. See Example 7-13 on page 135.

Example 7-13 Assign volume to the host

```

svctask mkvdiskhostmap -force -host 3 -scsi 1 4
svctask mkvdiskhostmap -force -host 4 -scsi 1 4

```

7.3.2 Optimize policy-based HA internal data flow: Assign host location

Assigning the host location can make a significant performance difference in a policy-based HA configuration.

Note: Only hosts with a location setting will use the optimized data path management. You can modify existing host objects and assign the appropriate location to the host. In your partition select **Hosts** → **Hosts**, and select the host and modify the location.

The host location can be modified using the CLI. See Example 7-14.

Example 7-14 Change host location setting

```

svctask chost -location FS9110 3

```

7.3.3 Remove host from partition

A host removal is similar to legacy host management. Select **Hosts** → **Hosts**, right click to the host you want to delete and select **Remove host**. Follow the process and the system will remove the host including all volume mappings (if selected) from both policy-based HA systems.

The host can also be removed using the CLI. See Example 7-15.

Example 7-15 Host removal

```
svctask rmhost -force 3
```

7.4 Partition management

In this section we discuss partition management.

7.4.1 Change replication policy

Only a single replication policy can be active on each partition at any given time. Changing a replication policy may change the overall behavior of the replication, depending on the specific policy settings. Replacing an existing policy with a similar one that only has different timeout settings is possible. However, be aware that even minor changes can impact replication behavior. See Example 7-15.

Example 7-16 Assign a different replication policy to a partition

```
IBM_2145:SVC_SA2:superuser>chpartition -replicationpolicy 1 0
```

Replacing an existing policy with a new topology requires a policy removal first. To change a topology from policy-based replication to policy-based HA you must first remove the existing disaster recovery policy from the partition.

This step permanently removes:

- ▶ All volume and host settings associated with the policy.
- ▶ *Crucially, all replicated data on the remote site.*

After that you need to assign a new policy-based HA policy to the partition, which creates all volumes and hosts and initiates an initial data copy to the remote system.

Recommendation: Before proceeding, ensure you have a comprehensive backup plan in place, as the initial data copy to the remote site can take significant time depending on data volume.

7.4.2 Delete partition

Deleting a partition is permanent and removes all associated data. To proceed, you must first ensure there are no active replication policies linked to the partition.

Steps to delete a partition:

1. Within your partition select **Storage partition overview**. Click the three dots beside **Replication Policy**, press the **Remove Replication Policy** as shown in Figure 7-22 and follow the wizard. Note that all remote volume (including data) and host objects related to this partition will be removed.

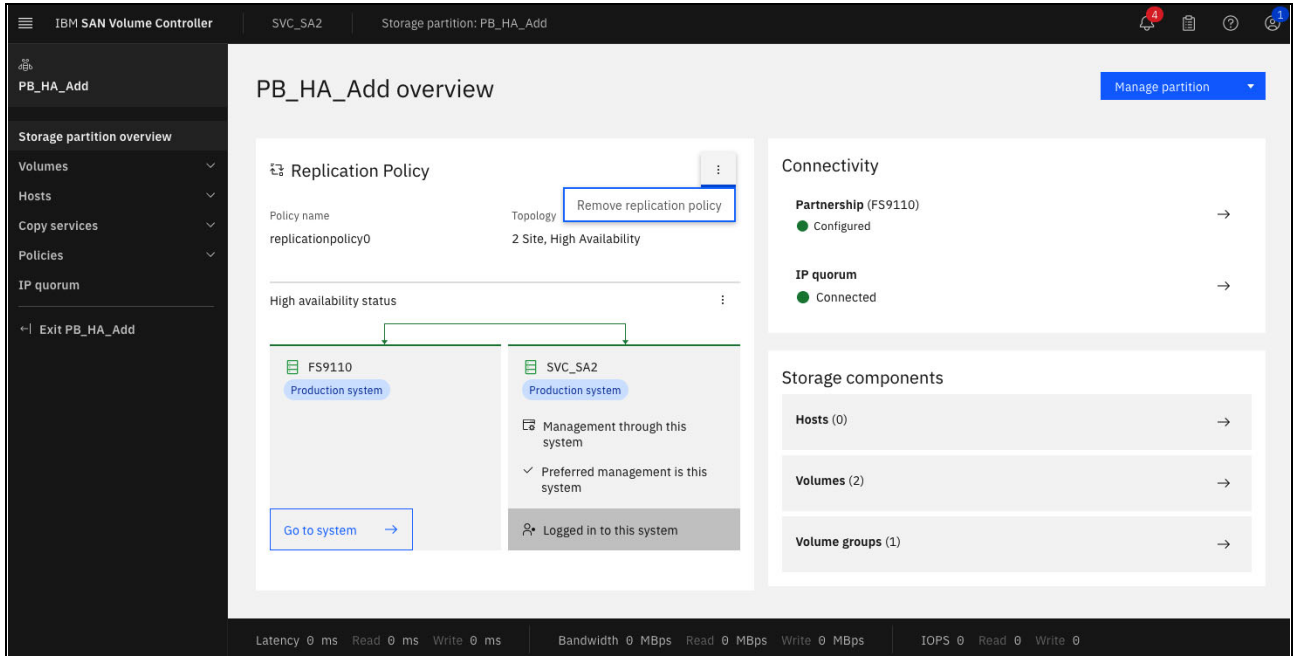


Figure 7-22 Remove replication policy

2. The partition can now be removed. Click **Manage partition** and **Delete** as shown in Figure 7-23. Follow the wizard, fill in the partition name you want to remove and click **Delete storage partition**.

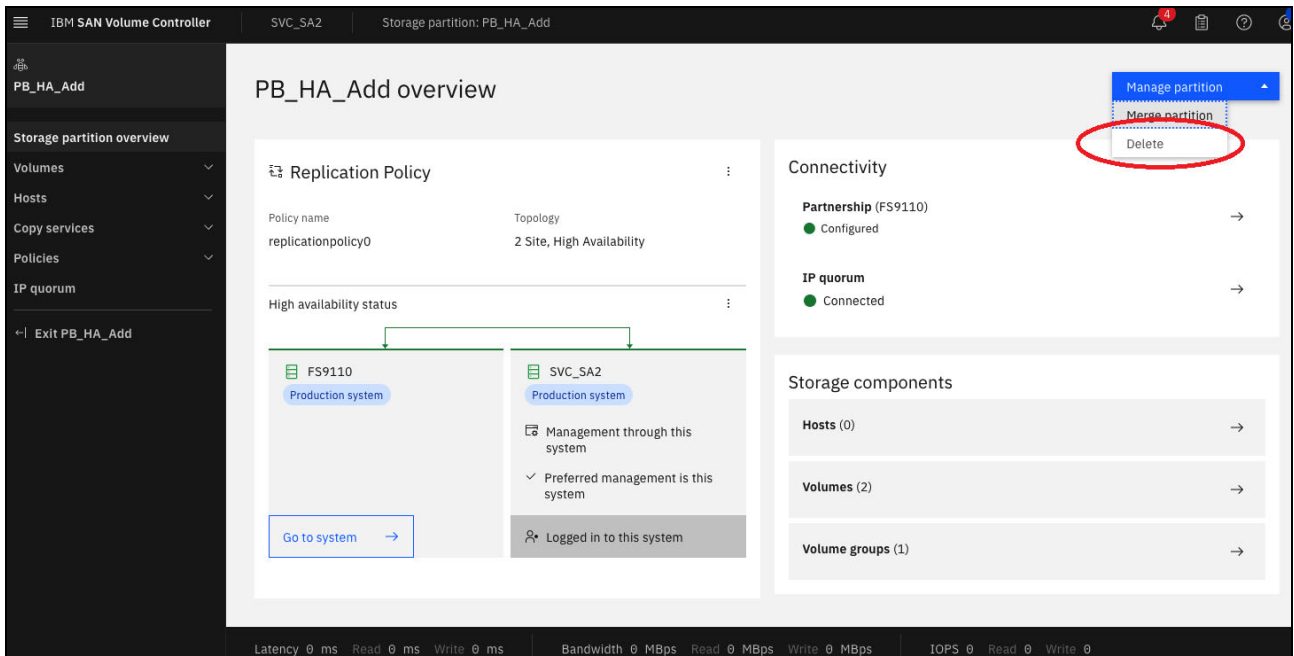


Figure 7-23 Delete the storage partition

Those actions can be performed from the CLI as well.

1. Remove the replication policy from the partition. See Example 7-17.

Example 7-17 Remove replication policy

```

IBM_2145:SVC_SA2:superuser>lspartition
id name      preferred_management_system_name active_management_system_name
replication_policy_id replication_policy_name location1_system_name
location1_status location2_system_name location2_status host_count
volume_group_count ha_status link_status desired_location_system_name
migration_status draft draft_volume_group_count draft_host_count
0 PB_HA_1 SVC_SA2 SVC_SA2 1
Test_PB_HA FS9110 healthy SVC_SA2
healthy 4 2 established synchronized
IBM_2145:SVC_SA2:superuser>chpartition -noreplicationpolicy PB_HA_1

```

2. Delete the partition. See Example 7-18.

Example 7-18 Delete the partition

```

IBM_2145:SVC_SA2:superuser>svctask rmpartition 0
IBM_2145:SVC_SA2:superuser>lspartition
IBM_2145:SVC_SA2:superuser>

```

If there is no replication required anymore on the system you may want to remove the storage system partnership to the remote storage system as well.

Note: Before deleting the partnership you must remove all active replication policies (using this particular partnership) from the system. Also, before deleting a partition, it is crucial to have a comprehensive backup plan in place, as any data stored within the partition will be permanently erased.

7.5 Migration options for policy-based HA partitions

Partitions can be migrated non-disruptively between storage systems. There is a partnership between source and target system required. Those migrations can be done for standard partitions without an active policy-based remote copy relationship like policy-based HA or policy-based replication.

Important: The required zoning changes between the storage systems and to the host systems are not part of the migration process and must be completed upfront.

If the partition is already in a policy-based remote copy relationship, the partition migration can not be done using the same method. There are multiple options available, which will be discussed next.

7.5.1 Migration of policy-based HA data by temporary HA protection removal

Perform the following steps for partition migration:

1. Define the appropriate copy direction by setting the migration target system as active management system.
2. Remove the existing policy-based HA policy from the storage partition. This step removes all volume and host definitions and deletes the data from the old policy-based HA partner site.
3. Assign a new policy-based HA policy to the storage partition for the target site. This will automatically create necessary volumes, hosts, and initiate data replication.
4. Verify the new configuration.

7.5.2 Migration of policy-based HA data to a third site by keeping the policy-based HA protection

Expanding your replication to a third site using policy-based replication is not currently supported. However, you can achieve migration using host-based mirroring to the storage system at the third, independent site. This approach replicates your data directly at the host level, keeping your current storage configuration unchanged and allowing you to implement the third-site protection without losing your existing 2-site HA functionality.

7.5.3 Migration of policy-based HA or policy-based replication data to a third site

The already existing policy-based HA solution or the policy-based replication solution should be migrated to a different system at a third site.

3-site configurations are currently not supported. IBM has already released a statements that it is planning policy-based HA plus policy-based replication to achieve replication to a third site in second half of 2024.

7.6 Snapshots and policy-based HA

All policy-based HA volumes are managed by the storage partition. You can assign snapshot policies to any volume group, regardless of whether they are standalone or managed by a partition. This assignment can be done independently on each storage system, allowing for different scheduling if needed.

7.6.1 Key differences from the previous snapshot solutions

The following are key differences from the previous snapshot solutions:

- ▶ **Cloning flexibility:** You can create clones and thin clones of these volumes to existing volumes or newly created volume groups outside the partition.
- ▶ **Policy removal impact:** Removing the replication policy from the partition will cause the associated volume and data to be deleted on the remote site. Additionally, the associated snapshots will also be removed when they expire.



Configuring FlashSystem and SVC partnerships over high-speed Ethernet

This chapter discusses how to configure FlashSystem and SVC partnerships over high-speed Ethernet to be used for policy-based replication and policy-based-HA. We focus on the deployment of short-distance partnerships using RDMA (Remote Direct Memory Access).

We cover the following key aspects:

- ▶ **High-speed replication portset setup:** Learn how to configure high-speed replication portsets for optimal performance.
- ▶ **IP address assignment:** Discover the process for assigning IP addresses to these portsets.
- ▶ **Partnership creation:** Explore methods for creating partnerships between storage systems, using both the Command Line Interface (CLI) and Graphical User Interface (GUI).
- ▶ **Deployment guidelines:** Gain valuable insights into the recommended practices for deploying short-distance partnerships that leverage RDMA for efficient data replication.

Note: This chapter is based on a [whitepaper](#) authored by Abhishek Jaiswal, Aakanksha Mathur, Akshada Thorat, Akash Shah and Santosh Yadav from IBM India.

This chapter has the following sections:

- ▶ “Introduction to replication over high-speed Ethernet” on page 142
- ▶ “Short-distance partnership using RDMA” on page 142
- ▶ “Setup considerations” on page 142

8.1 Introduction to replication over high-speed Ethernet

IBM Storage Virtualize V8.6.2 introduced support for setting up replication over high-speed Ethernet using the Remote Direct Memory Access (RDMA) protocol, enabling disaster recovery (DR) with high bandwidth over short distances. This capability enables customers with Ethernet infrastructure to implement DR with performance close to Fibre Channel (FC).

This chapter guides you through configuring a DR solution utilizing high-speed Ethernet partnerships between two systems. It details the prerequisites and setup considerations for establishing these partnerships, providing a comprehensive procedure that incorporates RDMA technology. Visual aids such as topology diagrams and step-by-step instructions through both GUI and CLI interfaces are included to facilitate the setup process.

To maintain real-time data copies at the remote DR site, this chapter explores the utilization of policy-based replication and remote copy technologies. In this setup, one system acts as the production system, where hosts access the data, while the other system serves as the DR system at a distant location. Visual aids such as topology diagrams and step-by-step instructions through both GUI and CLI interfaces are included to facilitate the setup process.

Policy-based replication offers asynchronous data replication with a variable recovery point greater than zero, aiming to achieve an optimal recovery point considering business needs. In contrast to remote copy, it minimizes overhead, providing higher throughput and lower latency between systems. Notably, it eliminates complex configuration requirements at the DR site, saving user time and ensuring streamlined failover procedures in DR scenarios.

8.2 Short-distance partnership using RDMA

IBM introduced replication over high-speed Ethernet, also known as short-distance partnership using the RDMA technology, for data transfer over replication links. RDMA offloads data transfer from CPUs and operating systems, resulting in low latency and high throughput. This is ideal for reliable, short-distance connections with low round-trip time (RTT), enabling superior performance.

8.3 Setup considerations

This section explains the setup considerations for configuring replication over high-speed Ethernet using RDMA.

8.3.1 Initial setup considerations

It is possible to configure short-distance partnership using RDMA between two systems. Both the production and recovery systems should either be in the replication or the storage layer.

You can use the `svcinfo lssystem` command to find the layer in which the system is. Refer to [IBM Documentation for lssystem](#).

You can use the `svctask chsystem` command to change the layer of the system. There can be up to two redundant fabrics established between the two systems. Refer to [IBM Documentation for chsystem](#).

Hardware considerations

Short-distance partnership using RDMA is supported on IBM FlashSystem 9XX0, FlashSystem 7X00, FlashSystem 5X00, and IBM System Storage SAN Volume Controller platforms.

Note: The system must contain RDMA-capable Ethernet adapters for establishing short distance partnership using RDMA. The adapter part number of the supported adapter is 01LJ587.

Software considerations

Make a note of the following software considerations to establish a short-distance partnership using RDMA:

- ▶ Short-distance partnerships using RDMA is supported in IBM FlashSystem 8.6.2.0 and later versions. Both the systems should have 8.6.2.0 or later.
- ▶ Systems where IBM HyperSwap solution is already deployed in an Ethernet environment cannot be a part of a short-distance partnership using RDMA.

Additionally, the candidate systems participating in the partnership must not be visible to each other over FC connections. You can check this using the following command: `svcinfolspartnershipcandidate`.

8.3.2 Network requirements

For high availability purpose, there are two replication links that are supported for short-distance replication using RDMA. These links could be of type, routed or direct attach.

Short-distance partnerships using RDMA is supported on both layer 2 and layer 3 networks, as shown in Figure 8-1 on page 144.

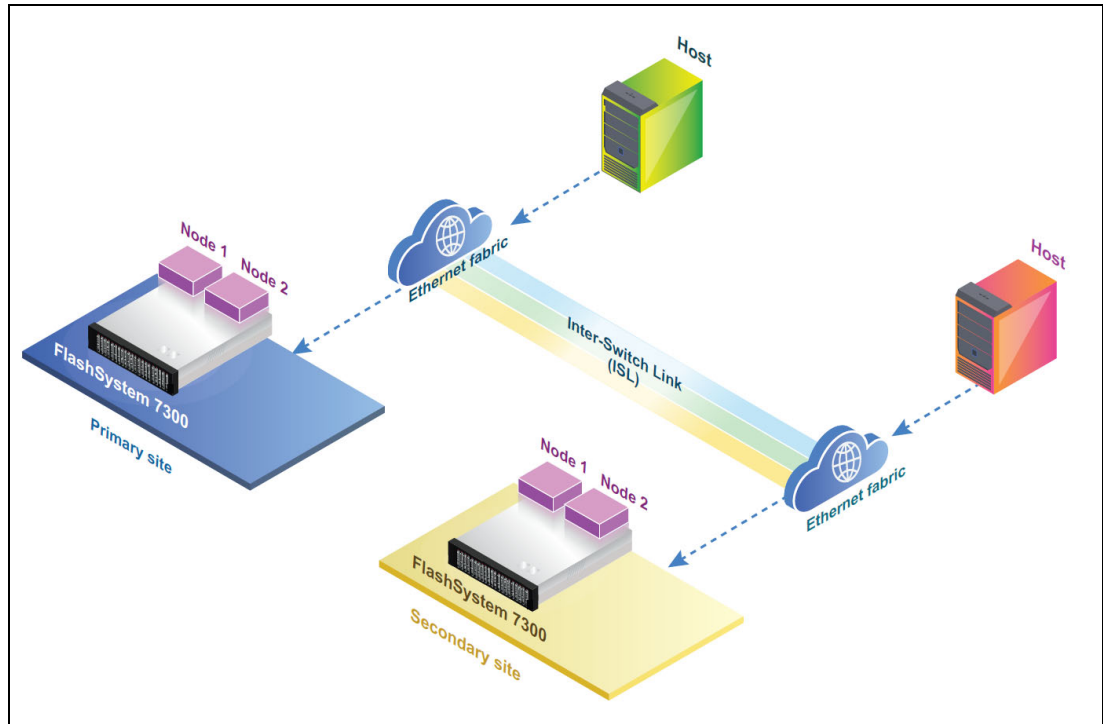


Figure 8-1 Configuration topology for short-distance partnership using RDMA

8.3.3 Deployment of short-distance partnership using RDMA

To establish a short-distance RDMA partnership between two systems, you must first configure them. Each partnership can have up to two links, each associated with a single highspeedreplication portset. Use the `-link1` or `-link2` attribute in the `mkippartnership` command to specify the desired highspeedreplication portset for each link. For partnerships with dual links, both `-link1` and `-link2` attributes must be used, along with their corresponding highspeedreplication portsets.

If there is a single ISL, either `-link1` or `-link2` replication links can be used. If there are two ISLs, both the links, `-link1` and `-link2` can be used.

All the adapters in the configuration are RDMA-capable Ethernet adapters. To avoid network congestion, the ISL between the two systems should be provisioned in such a way that it should be able to accommodate all the traffic passing through it. See Figure 8-2 on page 145

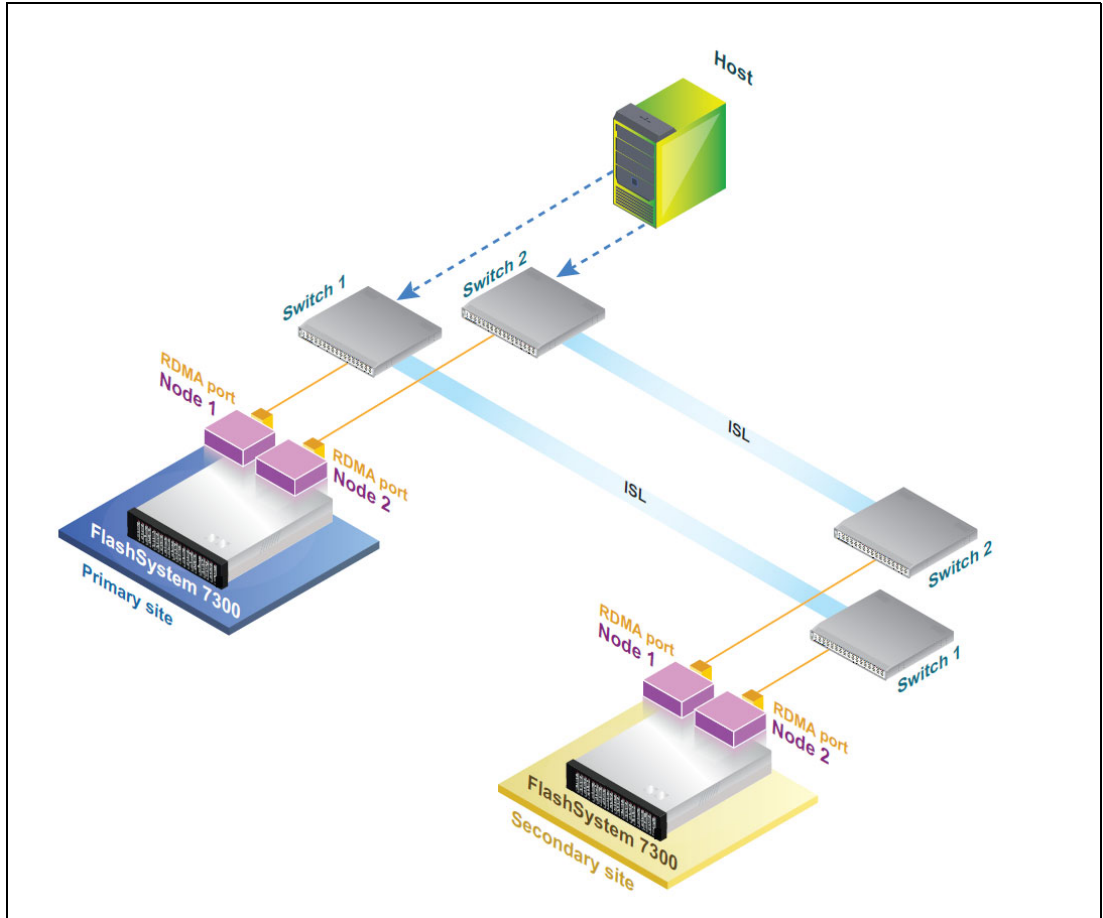


Figure 8-2 Configuration of a short-distance partnership using RDMA over ISL

You can configure the partnership by directly connecting the ports of the two systems, as shown in Figure 8-3 on page 146.

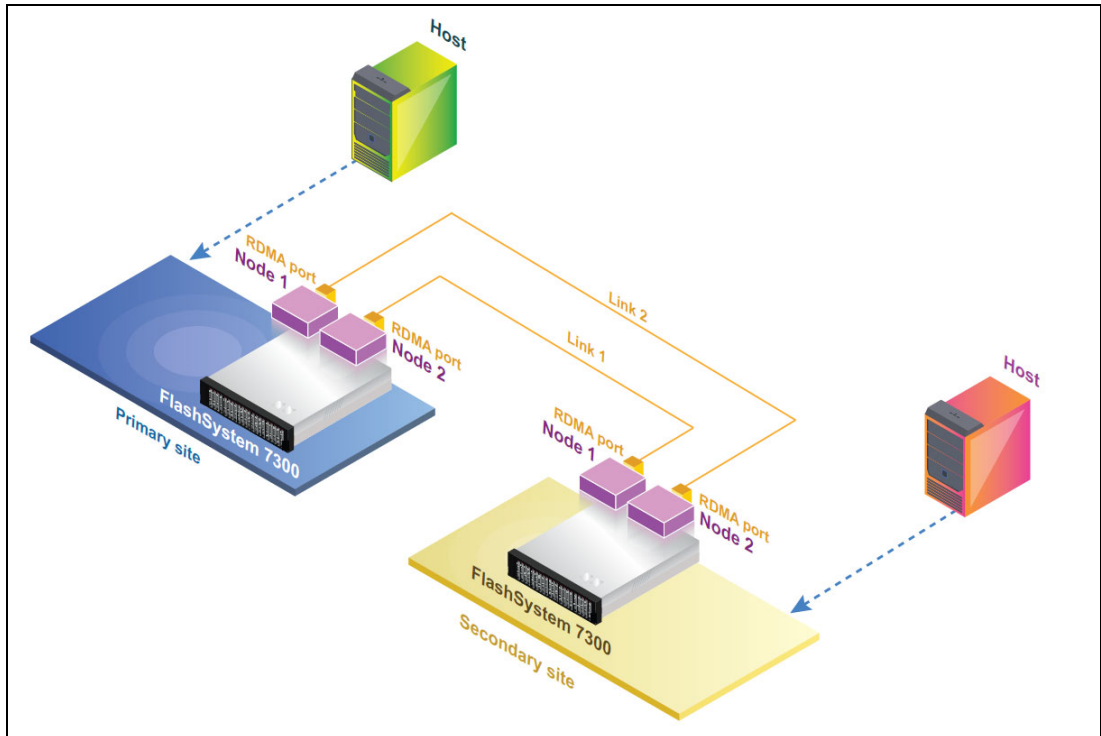


Figure 8-3 Configuration of a short-distance partnership using RDMA using direct-attach connections

8.3.4 Configuring a short-distance partnership using RDMA

The configuration task includes the following steps:

1. Identify the systems that participate in the partnership.
2. Identify the RDMA ports on each system, which will be a part of partnership links and has connectivity between them.
3. Create highspeedreplication portsets on each setup.
4. Configure these RDMA ports with IP addresses and map them to the respective highspeedreplication portset.
5. Establish the partnership between both the systems.

The following section takes you through detailed configuration steps using GUI and CLI.

Configuration using the GUI

The procedures and screen captures in this section exhibit a walkthrough of the IBM Storage Virtualize GUI and explain the steps to create and configure a high-speed replication portset.

Create high-speed replication portsets

Perform the following steps to create a high-speed replication portset:

1. In the IBM FlashSystem GUI, click **Settings** → **Network** → **Portsets**. Then click **Create Portset**, as shown in Figure 8-4 on page 147.

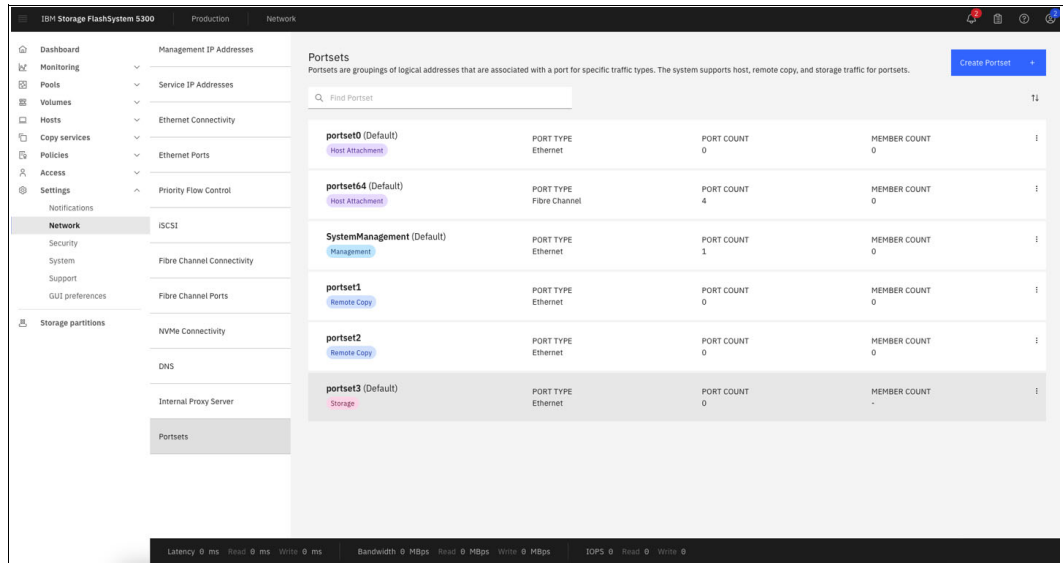


Figure 8-4 Creating a port-set

- In the **Create Portset** dialog, enter a name for the portset (for example, **portset4**, in this instance). Portset name is a user-defined variable and you can give any name to the portset. Select the portset type as **High speed replication**. In the **Port Type** section, select Ethernet and then click **Create**. See Figure 8-5 on page 147.

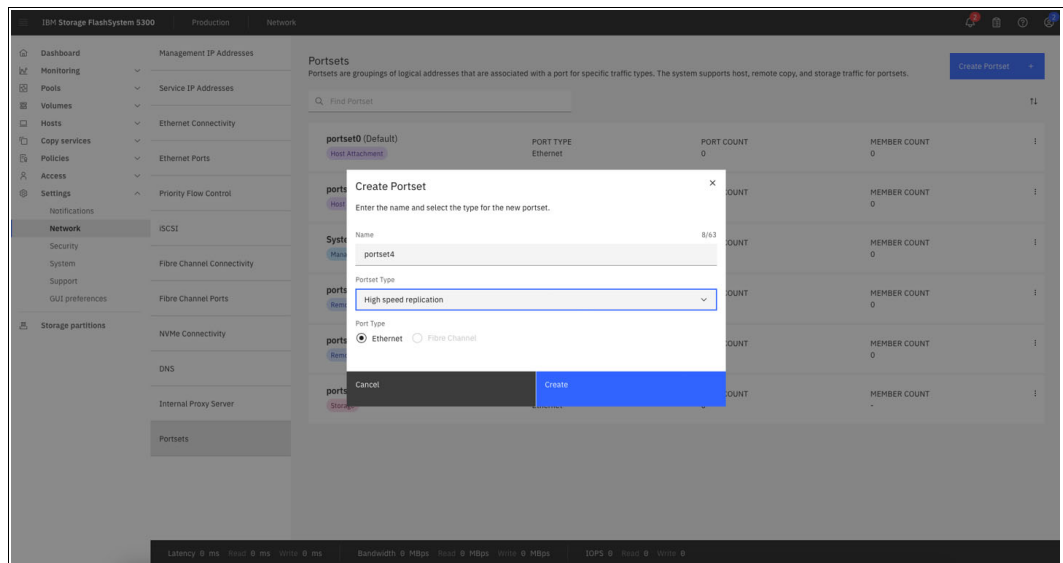


Figure 8-5 Specify port-set name and type

- To create another portset named "portset5," simply repeat the previous steps. Once created, both "portset4" and "portset5" will be listed on the Portsets page. See Figure 8-6 on page 148.

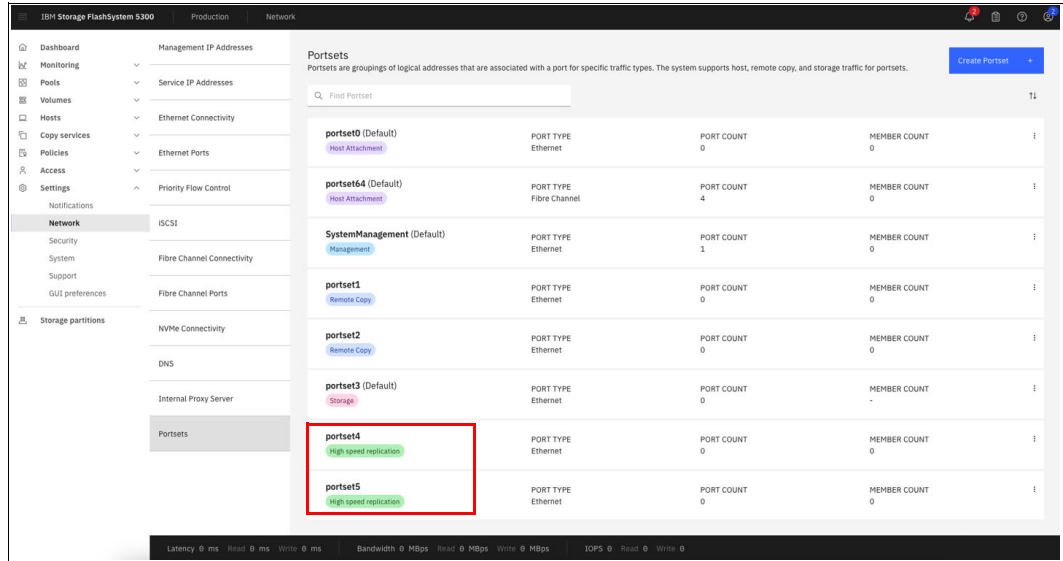


Figure 8-6 Listing of portsets

Assign IP addresses to the portsets

After the creation of high-speed replication portsets, you can assign IP addresses to them by performing the following steps:

1. Click **Settings** → **Network** → **Ethernet Ports**. On the Ethernet Ports page, right-click the port to which you need to assign an IP address and click **Manage IP Addresses**, as shown in Figure 8-7 on page 148.

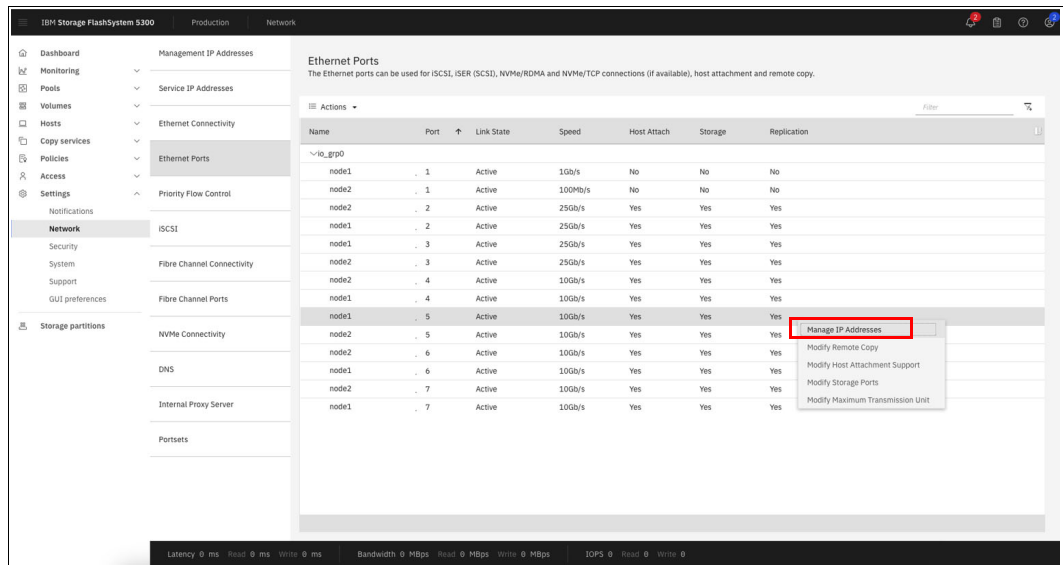


Figure 8-7 Right click on RDMA port

2. Click **Add IP Address**. See Figure 8-8 on page 149

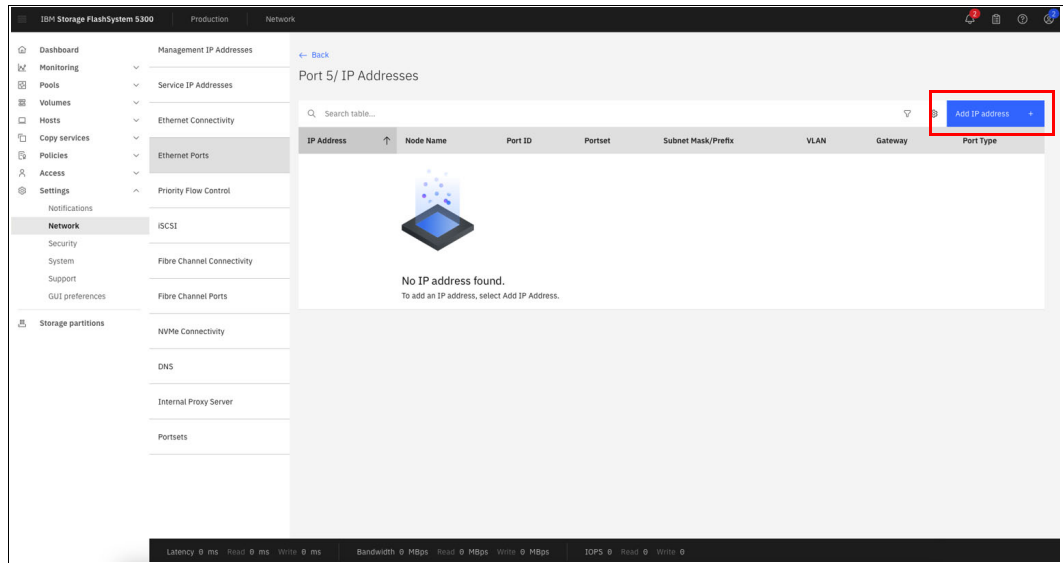


Figure 8-8 Add IP address

3. Enter information as shown in Figure 8-9 on page 149 for the IP address that you are adding to the selected port: IP address, subnet mask, VLAN, and gateway. Specify IPv4 as the type.

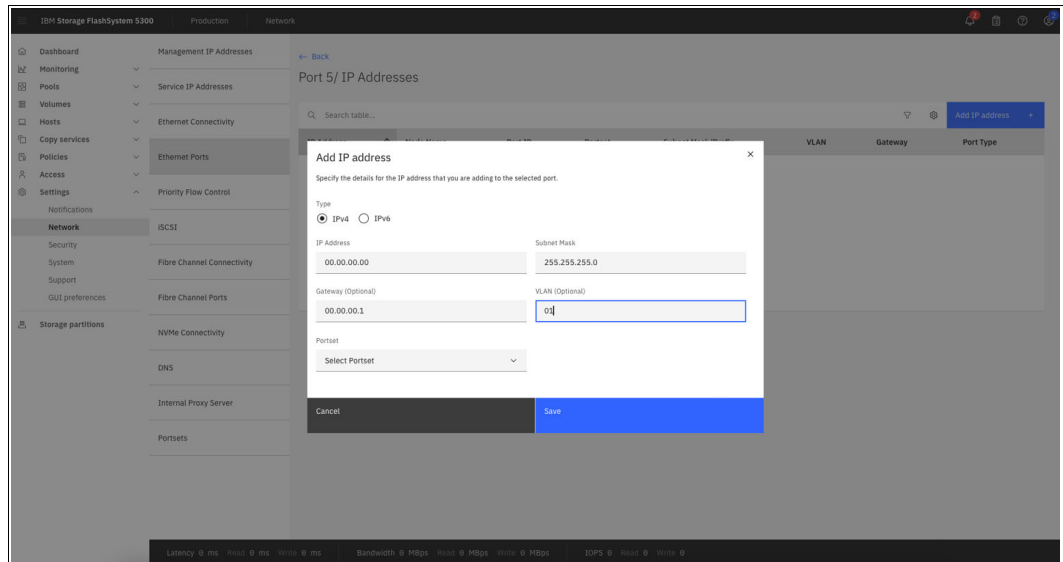


Figure 8-9 Enter IP address, subnet mask, VLAN, and gateway

4. Select the name of the portset and ensure that the portset type matches the specified traffic type. See Figure 8-10 on page 150.

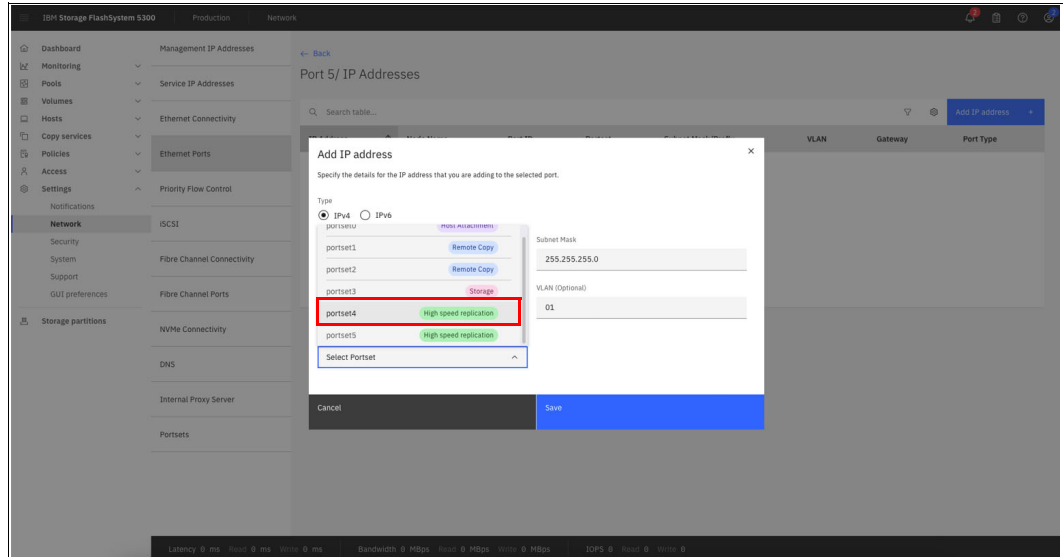


Figure 8-10 Select a portset

5. After selecting the portset (portset4), click **Save**. You can now see, in Figure 8-11 on page 150, the assigned IP to the respective portset.

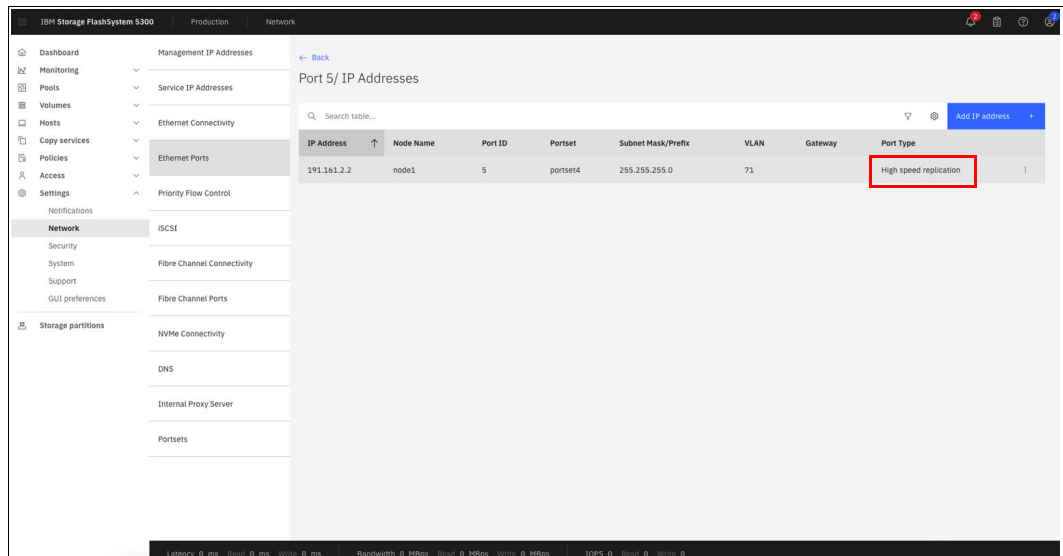


Figure 8-11 IP address of RDMA port

6. Repeat the same procedure for assigning IP to another portset (portset5).

Create a partnership

After assigning the IP addresses, perform the following steps to create a partnership.

1. Click **Copy Service** → **Partnership and remote copy** → **Create Partnership**.
2. Select **2-site partnership** then click **Continue**, as shown in Figure 8-12 on page 151.

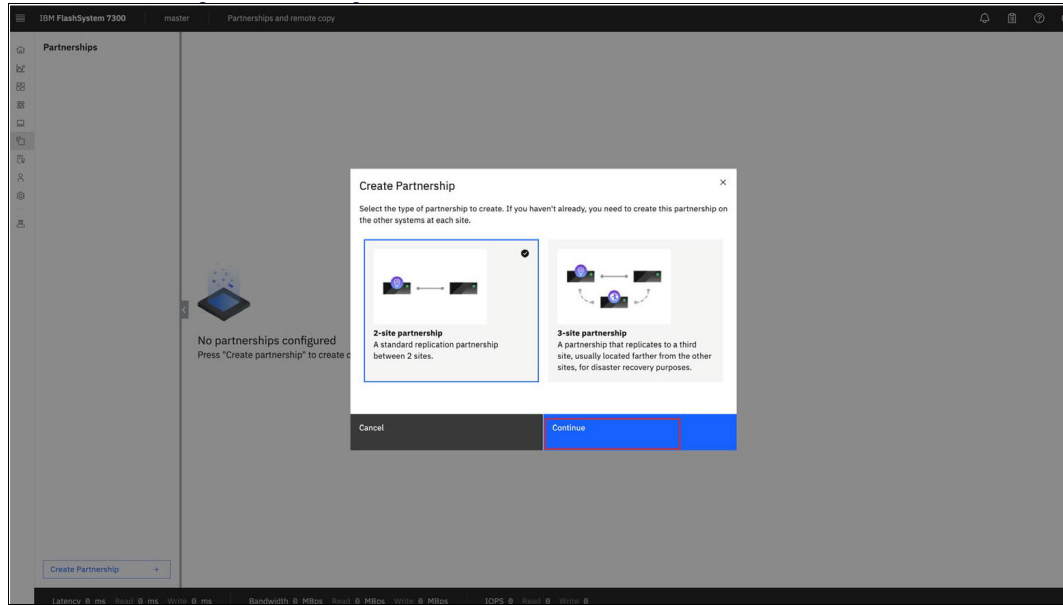


Figure 8-12 Creating a partnership

3. Select **IP (short distances using RDMA)** as the partnership type, enter the partner cluster IP address, and click **Test Connection**. See Figure 8-13 on page 151.

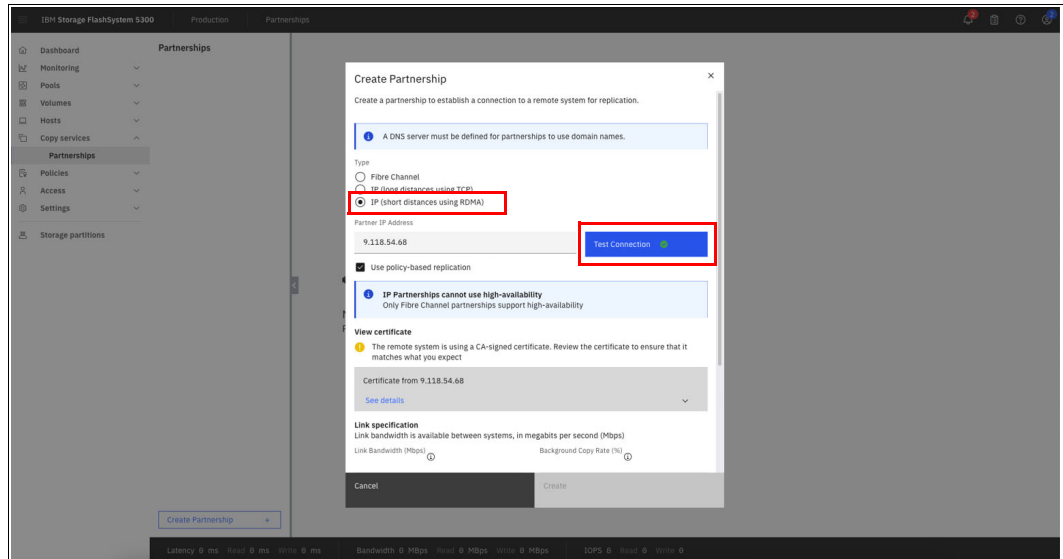


Figure 8-13 Select a short-distance partnership using RDMA

4. After testing the connection, select policy-based replication based on your requirement. See Figure 8-14 on page 152.

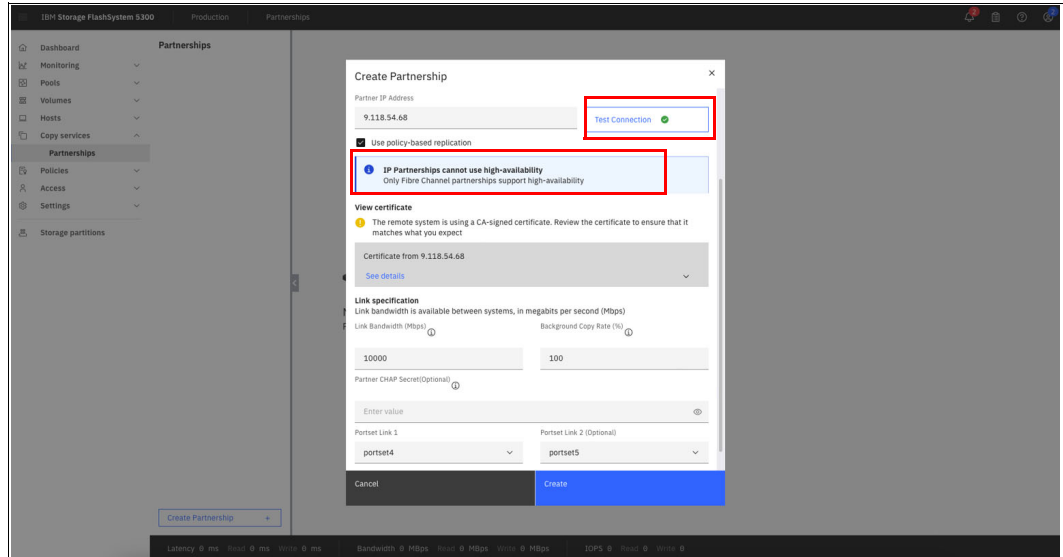


Figure 8-14 Test a connection for partnership

5. Enter link bandwidth and background copy rate. Then select the portsets of type highspeedreplication for Portset Link1 and Portset Link2. In this example, **portset4** and **portset5** are selected for Portset Link1 and Portset Link2 respectively. After that click **Create**, as shown in Figure 8-15 on page 152.

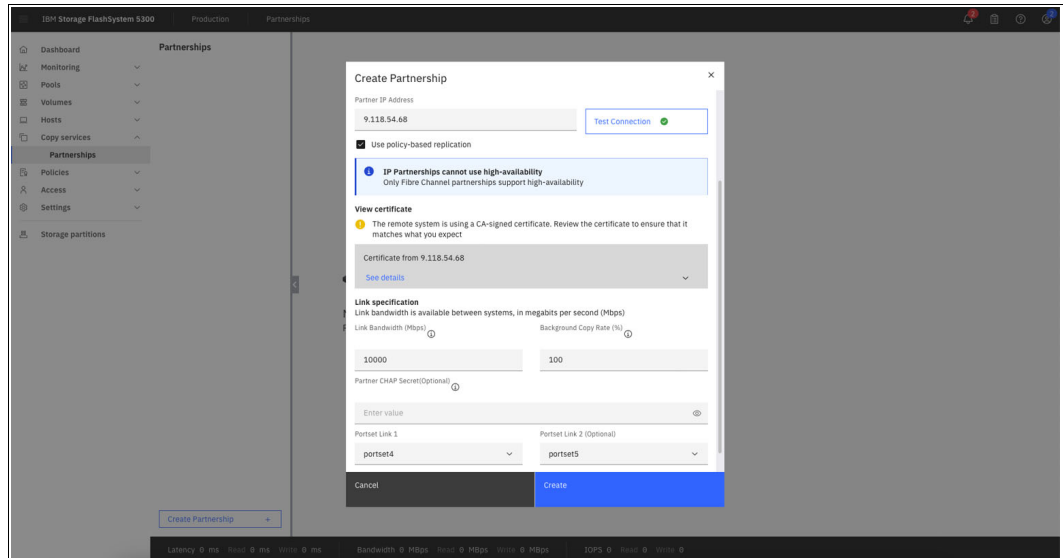


Figure 8-15 Select Portset Link1 and Portset Link2

6. Notice that the partially configured partnership is created, as shown in Figure 8-16 on page 153.

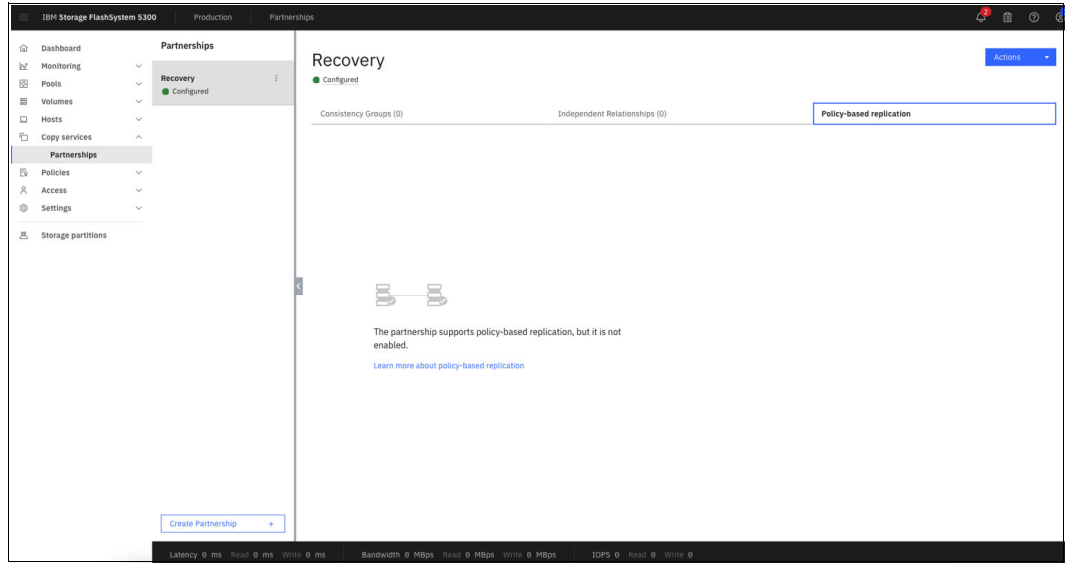


Figure 8-16 Partially configured partnership

You can follow the preceding steps on the remote cluster as well to change the partnership status to fully configured state.

7. After completing the steps for the remote cluster, notice that the partnership status shows **Configured**. See Figure 8-17 on page 153.

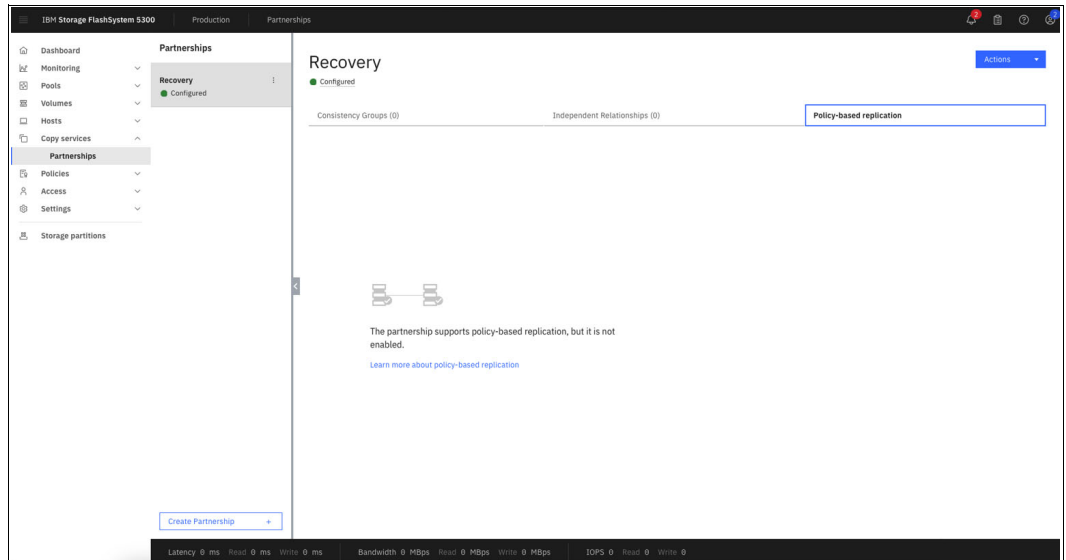


Figure 8-17 Configured partnership

8. After partnership is configured click the **Overflow** menu and then select **Partnership Properties**, as shown in Figure 8-18 on page 154.

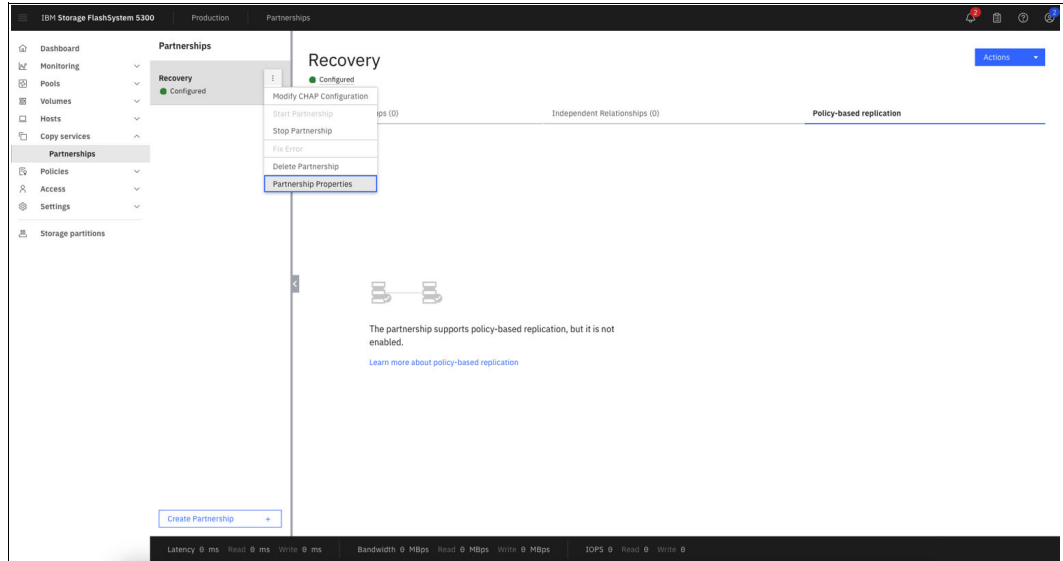


Figure 8-18 Select partnership properties

In the Properties dialog, notice that you can see a detailed view of partnership such as links, configuration status, type (as short distance using RDMA) and so on. See Figure 8-19 on page 154.

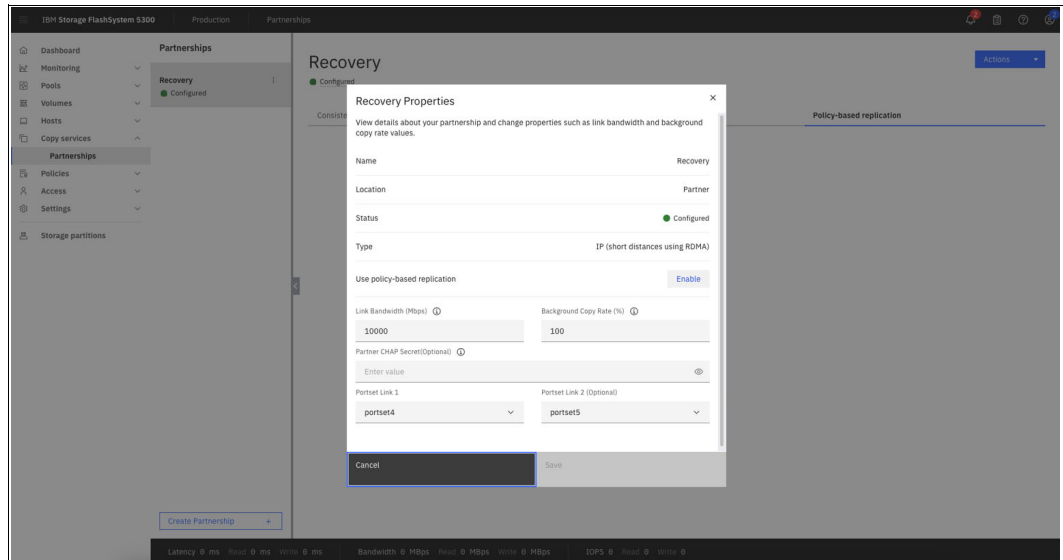


Figure 8-19 Partnership properties

Configuration using the CLI

This section takes you through the configuration of a short-distance partnership using RDMA using CLI.

Create a highspeedreplication portset

You can create the newly introduced highspeedreplication portset using the `mkportset` command with type as highspeedreplication. The portsets that are created using the `mkportset` are listed under the `lspportset` command, as shown in Figure 8-20.


```

IBM_FlashSystem:master:superuser>svctask mkportset -type highsppedreplication
Portset, id [4], successfully created
IBM_FlashSystem:master:superuser>svctask mkportset -type highsppedreplication
Portset, id [5], successfully created
IBM_FlashSystem:master:superuser>lsportset
id name          type                port_count host_count lossless owner_id owner_name port_type is_default
0 portset0       host                0          0          no      0          host      ethernet yes
1 portset1       replication         0          0          no      0          host      ethernet no
2 portset2       replication         0          0          no      0          host      ethernet no
3 portset3       storage             0          0          no      0          host      ethernet no
4 portset4       highsppedreplication 0          0          no      0          host      ethernet no
5 portset5       highsppedreplication 0          0          no      0          host      ethernet no
64 portset64     host               4          69         yes     0          host      fc        yes
72 SystemManagement management          1          0          no      0          host      ethernet yes

```

Figure 8-20 *mkportset and lsportset command output*

In Figure 8-20, there are two highsppedreplication portsets with default names as portset4 and portset5. Portset name is a user-defined attribute. The **mkportset** command has the following syntax:

```
#mkportset -type <portset_type>
```

Assign IP addresses

Configuration of IP addresses to the portsets created in previous step can be done using the **mkip** command. Users should select only the RDMA port for configuring IP addresses and mapping those to the highsppedreplication portsets. In the following screenshot, RDMA port 5 of node1 and node2 are selected for configuring IP addresses. See Figure 8-21.

```

IBM_FlashSystem:master:superuser>lsportethernet
port_id node_id node_name MAC duplex speed link_state dcbx_state rdma_type adapter_location adapter_port_id host storage replication eth clustering management
1 1 node1 0c:48:c6:7f:85:9a Full 10b/s active unsupported 0 1 yes yes yes no yes
2 1 node1 0c:48:c6:7f:85:9b inactive 0 2 yes yes yes no yes
3 1 node1 0c:48:c6:7f:85:9c inactive 0 3 yes yes yes no no
4 1 node1 0c:48:c6:7f:85:9d inactive 0 4 yes yes yes no no
5 1 node1 00:07:43:5a:9d:58 Full 25Gb/s active disabled iWARP 1 1 yes yes yes no no
6 1 node1 00:07:43:5a:9d:50 Full 25Gb/s active disabled iWARP 1 2 yes yes yes no no
7 1 node1 b8:59:9f:d9:cc:81 Full 25Gb/s active enabled RoCE 2 1 yes yes yes no no
8 1 node1 b8:59:9f:d9:cc:80 Full 25Gb/s active enabled RoCE 2 2 yes yes yes no no
1 2 node2 0c:48:c6:7f:86:76 Full 10b/s active unsupported 0 1 yes yes yes no yes
2 2 node2 0c:48:c6:7f:86:77 inactive 0 2 yes yes yes no yes
3 2 node2 0c:48:c6:7f:86:78 inactive 0 3 yes yes yes no no
4 2 node2 0c:48:c6:7f:86:79 inactive 0 4 yes yes yes no no
5 2 node2 00:07:43:48:4c:d8 Full 25Gb/s active disabled iWARP 1 1 yes yes yes no no
6 2 node2 00:07:43:48:4c:d0 Full 25Gb/s active disabled iWARP 1 2 yes yes yes no no
7 2 node2 b8:59:9f:fc:61:4d Full 25Gb/s active enabled RoCE 2 1 yes yes yes no no
8 2 node2 b8:59:9f:fc:61:4c Full 25Gb/s active enabled RoCE 2 2 yes yes yes no no

```

Figure 8-21 *lsportethernet command output*

While assigning the IP address, you can either provide the highsppedreplication portset name or the portset ID. The following example shows how to assign IP addresses and map it to an highsppedreplication portset. See Figure 8-22 on page 155.

```

IBM_FlashSystem:master:superuser>mkip -node node1 -port 5 -portset portset4 -ip xx.xx.xx.xx -prefix xx -vlan xx
IP Address, id [1], successfully created
IBM_FlashSystem:master:superuser>mkip -node node2 -port 5 -portset portset5 -ip xx.xx.xx.xx -prefix xx -vlan xx
IP Address, id [2], successfully created
IBM_FlashSystem:master:superuser>lsip
id node_id node_name port_id portset_id portset_name IP_address prefix vlan gateway owner_id owner_name
0 0 0 1 72 SystemManagement xx.xx.xx.xx xx xx.xx.xx.xx
1 1 node1 5 4 portset4 xx.xx.xx.xx xx xx
2 2 node2 5 5 portset5 xx.xx.xx.xx xx xx

```

Figure 8-22 *mkip and lsip command output*

The assigned IP addresses are listed using the **lsip** command, which has the following syntax:

```
#mkip -node <node_name> -port <port_id> -ip <x.x.x.x> -prefix <subnet_prefix>
-portset <portset_id | portset_name> -vlan <vlan_id >
```

Note: The RDMA port should be used for short distance partnership using the RDMA type. The same port which has been used for partnership may not be used for host attachment, storage attachment, replication, or Ethernet clustering. Also, the same IP address assignment for two different RDMA port is not allowed.

Create a partnership

Establishing a short-distance partnership between production and recovery systems can be done using the **mkippartnership** command. The command takes two options **-link1** and **-link2**. Users should provide an individual portset to each of these two options. Users can provide a maximum of two links per partnership.

It is advisable to provide two portsets corresponding to each of the link options. A partnership could also be created with a single link option, but users can use both the replication links for redundancy purposes.

The example in Figure 8-23 shows a short-distance partnership creation with **highspeedreplication** portsets (as created in earlier steps). Users can provide either a **highspeedreplication** portset ID or name while creating a partnership.

In the example the partnership status is **partially_configured_local**.

```
IBM_FlashSystem:master:superuser>svctask mkippartnership -clusterip xx.xx.xx.xx -linkbandwidthbits 25000 -backgroundcopyrate 100 -link1 portset4 -link2 portset5
IBM_FlashSystem:master:superuser>lspartnership
id      name      location partnership      type cluster_ip  event_log_sequence link1  link2  link1_ip_id link2_ip_id
000002043BE162B0 master    local                    partially_configured_local ipv4 xx.xx.xx.xx  portset4 portset5
0000020437415F70 aux      remote
```

Figure 8-23 **mkippartnership** and **lspartnership** command output

The created partnership is listed using the **lspartnership** command which has the following syntax:

```
#svctask mkippartnership -type <partnership_type> -clusterip <remote_cluster_ip>
-linkbandwidthbits <link_bandwidth> -backgroundcopyrate <copy_rate> -link1 <link1
portset_id| portset_name> -link2 <link2 portset_id| portset_name>
```

To create a fully configured partnership, repeat the preceding command on the remote system. You can verify the partnership status using the **lspartnership** command on each system, as shown in Figure 8-24.

```
IBM_FlashSystem:master:superuser>lspartnership
id      name      location partnership      type cluster_ip  event_log_sequence link1  link2  link1_ip_id link2_ip_id
000002043BE162B0 master    local                    fully_configured      ipv4 xx.xx.xx.xx  portset4 portset5
0000020437415F70 aux      remote
```

Figure 8-24 **lspartnership** command output

For short distance partnerships, run the **sainfo lsnodeipconnectivity** command to observe RDMA connectivity. Figure 8-25 on page 156 shows that the status is connected for both the links of the fully configured partnership created in “Create a partnership” on page 150.

```
IBM_FlashSystem:master:superuser>sainfo lsnodeipconnectivity 01-1
status      local_port_id local_vlan local_rdma_type local_ip_addr remote_port_id remote_vlan remote_rdma_type remote_ip_addr remote_uwnn  remote_panel_name cluster_id  error_data
Connected:  IMARP 5      xx      IWARP      xx.xx.xx.xx  5      xx      IWARP      xx.xx.xx.xx  5005076810000184 01-1  000002043BE162B0

IBM_FlashSystem:master:superuser>sainfo lsnodeipconnectivity 01-2
status      local_port_id local_vlan local_rdma_type local_ip_addr remote_port_id remote_vlan remote_rdma_type remote_ip_addr remote_uwnn  remote_panel_name cluster_id  error_data
Connected:  IMARP 5      xx      IWARP      xx.xx.xx.xx  5      xx      IWARP      xx.xx.xx.xx  5005076810000158 01-2  000002043BE162B0
```

Figure 8-25 **sainfo lsnodeipconnectivity** command output

Note: The native IP replication can be done using a replication type portset while short-distance partnership using RDMA is possible only with highspeedreplication type portsets. Compression and secured IP partnership are not supported with short-distance partnership using RDMA portsets.

Change attribute using the `chpartnership` command

The `chpartnership` command can be used for changing the attributes of an already created partnership. For more information of the `chpartnership` command, refer to [IBM Documentation for `chpartnership`](#).

8.3.5 Bandwidth utilization

For 100% write workload, IBM FlashSystem achieves maximum possible throughput with policy-based replication using high-speed Ethernet. The built-in performance monitor will show fully saturated links and maximum bandwidth utilization when using IBM FlashSystem with high-speed replication over Ethernet. The graph in Figure 8-26 illustrates this, showing fully saturated links and maximum bandwidth utilization with IBM FlashSystem using high-speed replication over Ethernet.

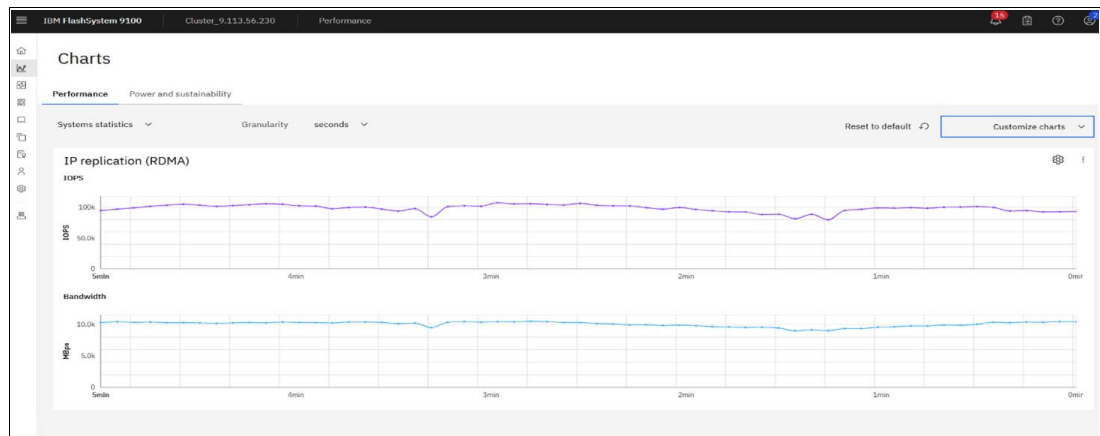


Figure 8-26 Throughput versus time

When the connection between the two near DR sites utilizes reliable links, replication performance reaches its optimal level. Since iWARP leverages TCP, performance remains relatively stable even during temporary link issues.

The graph data is for informational purposes only and does not reflect benchmark results.

8.3.6 Policy-based replication configuration checklist

Perform the following steps to configure policy-based replication:

1. Set up mutual Transport Layer Security (mTLS) between the systems.
2. Configure the partnership at both the sites.
3. Link the pools (using the `mdiskgrp` command) between the systems.
4. Create a replication policy of topology, 2-site-async-dr.
5. Create a volume group and assign the newly created replication policy to the group.
6. Create new volumes or add existing volumes to the volume group.

Refer to Chapter 4, “Implementing policy-based replication” on page 47 for more details.

8.3.7 General guidelines

Guidelines for creating a highspeedreplication portset:

- ▶ Only RDMA port should be used for short-distance partnerships using RDMA and the port cannot be used for any other traffic such as host attachment, storage attachment, replication, or Ethernet clustering.
- ▶ Up to two RDMA ports can be assigned per highspeedreplication portset per node.
- ▶ A maximum of six highspeedreplication portset creation is supported per system.
- ▶ Users can assign IPV4 and IPV6 addresses to a highspeedreplication portset.

Guidelines for creating a short-distance partnership:

- ▶ Ethernet-based RDMA clustering and short-distance partnership using RDMA cannot coexist.
- ▶ Partnerships can be created using a single replication link or a dual replication link.
- ▶ You can view the partnership status using the `lspartnership` command. Ideally, it should be fully configured. If partnership status is not present, then use the instructions in the following Troubleshooting section to fix it.

8.3.8 Troubleshooting

This section lists a few troubleshooting tips to validate the configuration.

1. Validate the partnership status:

- a. In the GUI, click **Copy Service** → **Partnership and remote copy** → **Partnerships**. Notice that the status is displayed as **Configured**, as shown in Figure 8-27.

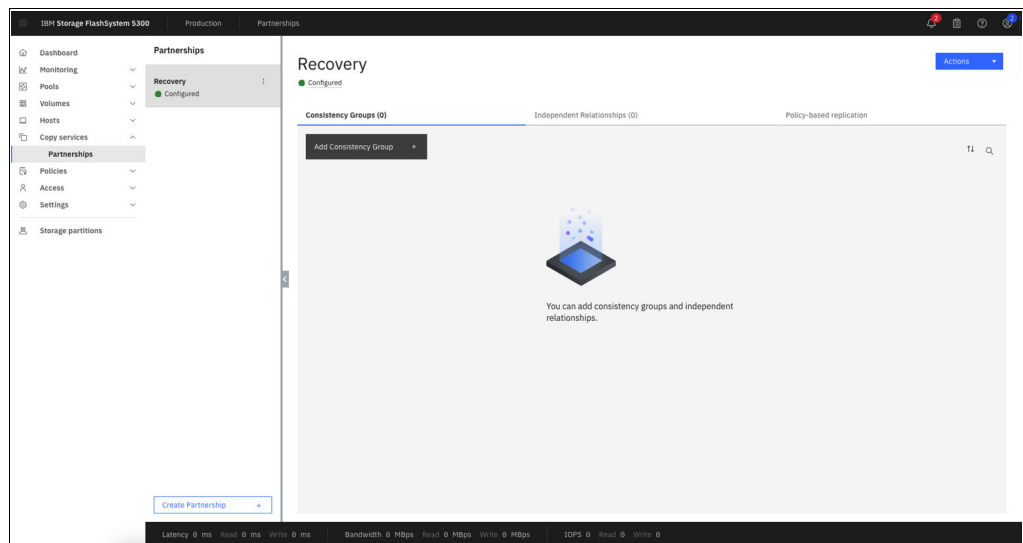


Figure 8-27 View the configuration status in GUI

- b. In the CLI, run the `lspartnership` command and check if the output shows the partnership status as `fully_configured` along with the other partnership attributes, as shown in Figure 8-28.

```

IBM_FlashSystem:master:superuser>lspartnership
id          name      location partnership      type cluster_ip  event_log_sequence link1  link2  link1_ip_id link2_ip_id
0000020438E162B0 master    local
0000020437415F70 aux      remote    fully_configured  ipv4 xx.xx.xx.xx          portset4 portset5
    
```

Figure 8-28 View the configuration status in CLI

2. Ensure connectivity between links: You need to ensure that all IP addresses associated with the link1 and link2 portsets on the production IBM FlashSystem storage system are connected with all IP addresses associated with the link1 and link2 portsets on the recovery IBM FlashSystem storage system. The same can be validated by using the following methods:

- a. In the GUI, click **Settings** → **Network** → **Ethernet Connectivity**, as shown in Figure 8-29 on page 159.

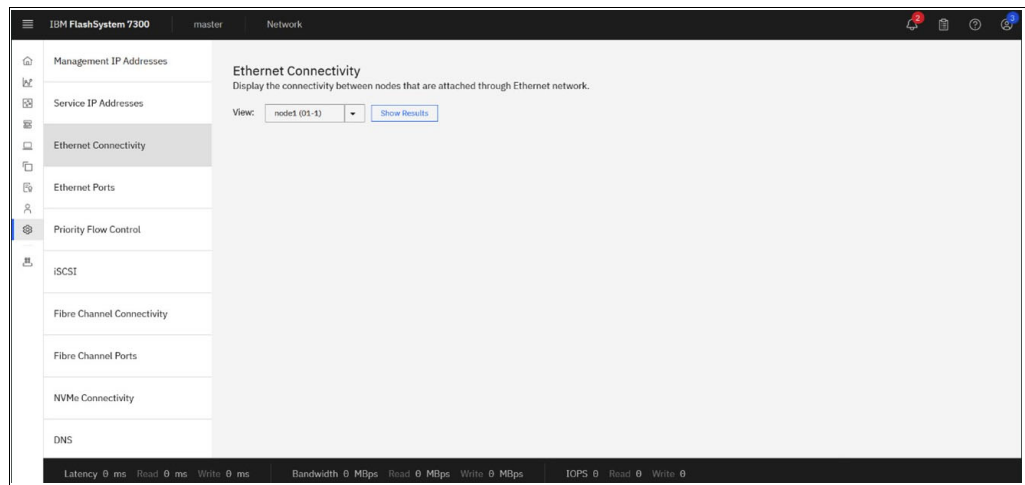


Figure 8-29 Ethernet Connectivity page

- b. Select the node and click **Show Results**, as in Figure 8-30.

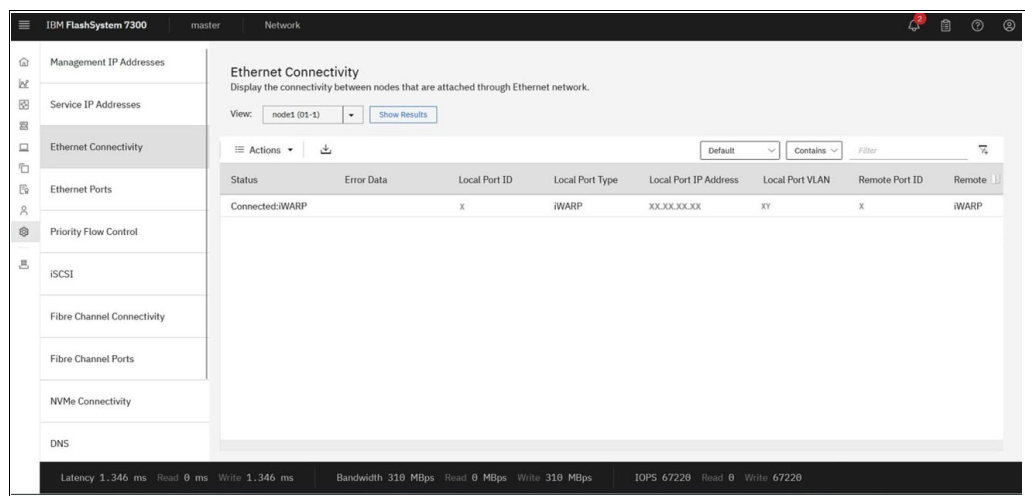


Figure 8-30 Displaying connectivity between nodes attached through Ethernet network

- c. In the CLI: Run the `sainfo lsnnodeipconnectivity` command on all the nodes of the system, as shown in Figure 8-31.

```

IBM_FlashSystem:master:superuser>sainfo lsnodeipconnectivity 01-1
status      local_port_id local_vlan local_rdma_type local_ip_addr remote_port_id remote_vlan remote_rdma_type remote_ip_addr remote_wann  remote_panel_name cluster_id  error_data
Connected:  INARP 5          xx          jNARP          xx.xx.xx.xx  5          xx          jNARP          xx.xx.xx.xx  5005076810000184 01-1          0000020438E162B0

IBM_FlashSystem:master:superuser>sainfo lsnodeipconnectivity 01-2
status      local_port_id local_vlan local_rdma_type local_ip_addr remote_port_id remote_vlan remote_rdma_type remote_ip_addr remote_wann  remote_panel_name cluster_id  error_data
Connected:  INARP 5          xx          jNARP          xx.xx.xx.xx  5          xx          jNARP          xx.xx.xx.xx  5005076810000158 01-2          0000020438E162B0

```

Figure 8-31 IP addresses configured on the ports

3. For a more resilient configuration to have maximum redundancy, ensure that the IP addresses configured on the ports for both the links are from different nodes.
4. In case the partnership status, is something other than `fully_configured`, further troubleshooting is required to understand why there is a change in the partnership and is not reflecting the required ideal state, which is `fully_configured`.
 - a. Partnership is in the `not_present` state: An IP partnership can transition to the `not_present` state due to multiple reasons and it means that the replication services have come to a halt. Check for alerts, warnings, or errors associated with partnership. Run the `!sevent log` command in the CLI or click **Monitoring** and then view the list in the Events tab in the GUI to find the events pertaining to the changes occurred in the system.
 - b. In the CLI, run the `sainfo lsnodeipconnectivity` command on all the nodes of the system to understand if there are any issues with the sessions established. The session status should ideally be `Connected` but apart from this state, there are also other states, such as `Protocol mismatch`, `Degraded`, and `Unreachable`.
5. While reference of event logs and directed maintenance procedure (DMP) from the GUI are the recommended ways to resolve any issue pertaining to the IBM FlashSystem, you can also perform system sanity by checking reachability to the remote system.
 - a. Check the connectivity to the remote cluster using the `svctask ping` command: For IPv4 / IPv6 use these commands:


```
# svctask ping -srcip4 <source_ip> <destination_ip>
```

```
# svctask ping6 -srcip6 <source_ip> <destination_ip>
```
6. In case you see the error codes 2021 or 2023 in the event logs, refer to the below links to determine the cause and action to be taken to resolve the issue:
 - Error code 2021: [IBM Documentation for error code 2021](#)
 - Error code 2023: [IBM Documentation for error code 2023](#)

In the GUI, follow the Directed Maintenance Procedure (DMP) from the menu **Monitoring** → **Events** to troubleshoot the issue. If you have followed the DMP and the issue is still not resolved, then another option is to raise a support ticket with IBM for further assistance.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- ▶ *Policy-Based Replication with IBM Storage FlashSystem, IBM SAN Volume Controller and IBM Storage Virtualize*, REDP-5704
- ▶ *Unleash the Power of Flash: Getting Started with IBM Storage Virtualize Version 8.7 on IBM Storage FlashSystem and IBM SAN Volume Controller*, SG24-8561

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

Online resources

These websites are also relevant as further information sources:

- ▶ *Configure policy-based replication over high-speed Ethernet transport on IBM FlashSystem* whitepaper:
<https://www.ibm.com/downloads/cas/NP4RWMKX>

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 526. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the .5" spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize(->Hide)>Set** . Move the changed Conditional text settings to all files in your book by opening the book file with the spine:fm still open and **File>Import>Formats** the Conditional Text Settings (ONLY!) to the book files.

Draft Document for Review July 12, 2024 9:54 am

8569spine.fm 163



Redbooks

Ensuring Business Continuity with Policy-Based

SG24-8569-00

ISBN



(1.5" spine)

1.5" <-> 1.998"

789 <-> 1051 pages



Redbooks

Ensuring Business Continuity with Policy-Based

SG24-8569-00

ISBN



(1.0" spine)

0.875" <-> 1.498"

460 <-> 788 pages

Redbooks

Ensuring Business Continuity with Policy-Based Replication and

SG24-8569-00

ISBN



(0.5" spine)

0.475" <-> 0.873"

250 <-> 459 pages

Redbooks

Ensuring Business Continuity with Policy-Based Replication and

(0.2" spine)

0.17" <-> 0.473"

90 <-> 249 pages

(0.1" spine)

0.1" <-> 0.169"

53 <-> 89 pages

To determine the spine width of a book, you divide the paper PPI into the number of pages in the book. An example is a 250 page book using Plainfield opaque 50# smooth which has a PPI of 526. Divided 250 by 526 which equals a spine width of .4752". In this case, you would use the ".5" spine. Now select the Spine width for the book and hide the others: **Special>Conditional Text>Show/Hide>SpineSize(->Hide)>Set** . Move the changed Conditional text settings to all files in your book by opening the book file with the spine:fm still open and **File>Import>Formats** the Conditional Text Settings (ONLY!) to the book files.

Draft Document for Review July 12, 2024 9:54 am

8569spine.fm 164



Ensuring Business Continuity with

SG24-8569-00

ISBN

(2.5" spine)
2.5" <-> mmm.n"
1315 <-> mmm pages



Ensuring Business Continuity with Policy-Based Replication and Policy-Based

SG24-8569-00

ISBN

(2.0" spine)
2.0" <-> 2.498"
1052 <-> 1314 pages





SG24-8569-00

ISBN

Printed in U.S.A.

Get connected

