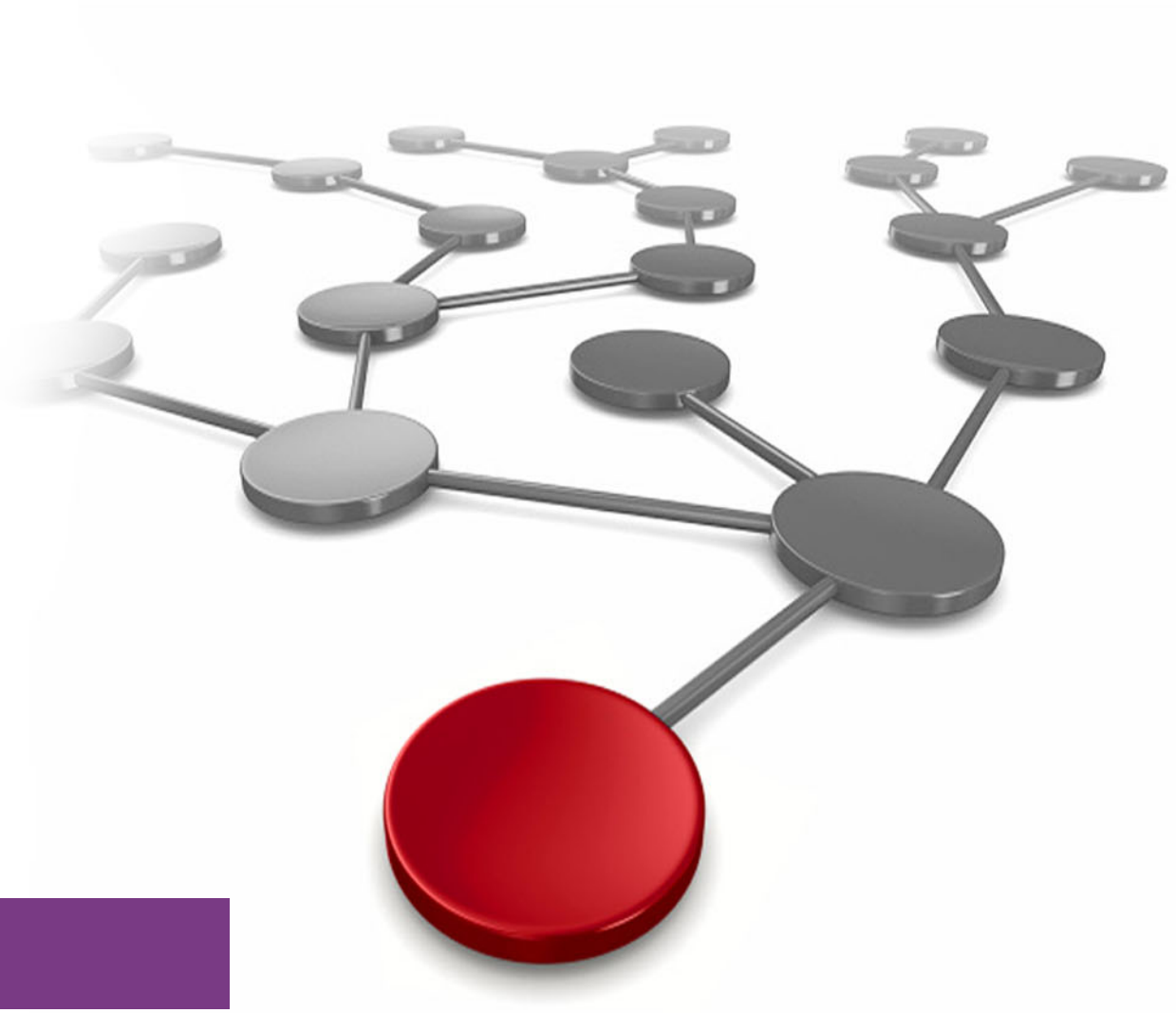
Redbooks
ibm.com/redbooks

IBM FlashSystem and VMware Implementation and Best Practices Guide

Vasfi Gucer
Duane Bolland
Nezih Boyacioglu
Jordan Fincher
David Green
Warren Hawkins
Ibrahim Alade Rufai
Leandro Torolho



Storage

IBM

Redbooks

IBM Redbooks

IBM FlashSystem and VMware Implementation and Best Practices Guide

October 2022

SG24-8505-01

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

Second Edition (October 2022)

This edition applies to IBM Spectrum Virtualize 8.5 and VMware vSphere ESXi Version 7.0.

Contents

Figures

Tables

Examples

Notices

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

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

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

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

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

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

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

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

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

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

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

COPYRIGHT LICENSE:

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

Trademarks

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

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

AIX®
Easy Tier®
FlashCopy®
HyperSwap®
IBM®
IBM Cloud®
IBM FlashCore®
IBM FlashSystem®
IBM Garage™
IBM Spectrum®
Redbooks®
Redbooks (logo)  ®
Storwize®
Tivoli®

The following terms are trademarks of other companies:

ITIL is a Registered Trade Mark of AXELOS Limited.

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

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

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

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

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

Preface

This IBM® Redbooks® publication details the configuration and best practices for using the IBM FlashSystem® family of storage products within a VMware environment. The first version of this book was published in 2021 and specifically addressed IBM Spectrum® Virtualize Version 8.4 with VMware vSphere 7.0. This second version of this book includes all the enhancements that are available with IBM Spectrum Virtualize 8.5.

Topics illustrate planning, configuring, operations, and preferred practices that include integration of IBM FlashSystem storage systems with the VMware vCloud suite of applications:

- VMware vSphere Web Client (vWC)
- vSphere Storage APIs - Storage Awareness (VASA)
- vSphere Storage APIs – Array Integration (VAAI)
- VMware Site Recovery Manager (SRM)
- VMware vSphere Metro Storage Cluster (vMSC)
- Embedded VASA Provider for VMware vSphere Virtual Volumes (vVols)

This book is intended for presales consulting engineers, sales engineers, and IBM clients who want to deploy IBM FlashSystem storage systems in virtualized data centers that are based on VMware vSphere.

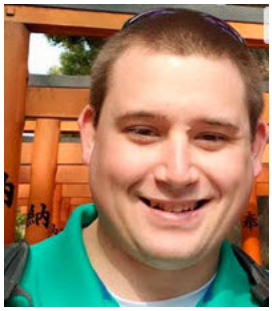
Authors

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

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

Duane Bolland is an IBM Storage Technical Advisor and works from Portland, Oregon. His first experience with IBM was as an IBM AIX® administrator in the 1990s. He has been with IBM® for 20 years. His past assignments include roles in the Open Systems Lab, SAN Volume Controller (SVC) product field engineer (PFE), and xSeries Lab Services. He is a technical focal for IBM Spectrum Virtualize and worked with the products since 2006. This book is his third IBM Redbooks publication.

Nezih Boyacioglu has 20 years of experience as a storage area network (SAN) storage specialist. He works for a premiere IBM Business Partner in Turkiye. His IBM storage journey started with IBM Tivoli® Storage Manager (IBM Spectrum Protect) and tape systems, and his main focus for the last 10 years is on the IBM Spectrum Virtualize family (SVC, IBM Storwize®, and IBM FlashSystem) and SANs. He is an IBM Certified Specialist for Enterprise Storage Technical Support, Flash Technical Solutions, Virtualized Storage, and IBM Spectrum Storage software.



Jordan Fincher is an SVC and IBM FlashSystem Level 3 Support Engineer. He has supported IBM Spectrum Virtualize products for IBM since 2015 and has contributed to several IBM Redbooks publications.

David Green works with the IBM Storage Area Network Central team troubleshooting performance and other problems on storage networks. He has authored, or contributed to, a number of IBM Redbooks publications. He is a regular speaker at IBM Technical University. You can find his blog at Inside IBM Storage Networking where he writes about all things that are related to Storage Networking and IBM Storage Insights.

Warren Hawkins has a background in Infrastructure Support in predominantly Microsoft Windows and VMware environments, and has gained 13 years experience working in 2nd and 3rd line support in both public and private sector organizations. Since joining IBM in 2013, Warren has played a crucial part in customer engagements. Using his field experience, he established himself as the Test Lead for the IBM Spectrum Virtualize product family, focusing on clustered host environments.

Ibrahim Alade Rufai has expertise on all aspects of designing, building, and implementing Enterprise Cloud and AI projects, Storage, and Software-defined infrastructure systems for cognitive era. He assists clients across Middle East and Africa design for cognitive business, build with collaborative innovation, and deliver through a cloud platform (Private, Public, Hybrid, and Multicloud).

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

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

Pierre Sabloniere
IBM France

Matthew Smith
IBM UK

Markus Oscheka
IBM Germany

Erica Wazevski
IBM US

Rivka Pollack
IBM Israel

Wade Wallace and Wendi Gervis Watson
IBM Redbooks, Raleigh Center

The authors team would also like to thank the following people from VMware for reviewing the VMware-related technical content in the book:

- Weiguo Hel, Peter Toro, Paul Turner, Ken Werneburg

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 by 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

- Look for us on LinkedIn:

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

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

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

- Stay current on recent Redbooks publications with RSS Feeds:

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

Summary of changes

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

This book is the second edition of the *IBM FlashSystem and*, SG24-8505 that was originally published on July 15, 2021. The new information that is included in this revision is described next.

October 2022, Second Edition

This revision includes the following new and changed information.

New information

- Chapter 6, “Embedded VASA Provider for Virtual Volumes” on page 129 is a new chapter.
- “Abbreviations and acronyms” on page 185 added.
- VMware key terminology.
- Claim rules to manage storage devices.
- Various performance best practices updates.
- Non-Volatile Memory Express (NVMe) volumes protocol types.
- Small Computer System Interface (SCSI) and NVMe zones.
- SCSI UNMAP feature.
- Virtual machine (VM) level performance screens.

Changed information

- Revised section on IBM Storage Insights integration

1

Introduction

This IBM Redbooks publication describes the configuration and best practices for using IBM Spectrum Virtualize based storage systems within a VMware environment. The first version of this book was published in 2021 and addressed IBM Spectrum Virtualize 8.4 with VMware Elastic Sky X integrated (ESXi) 7.0. This book is the second version, and it includes all enhancements that are available with IBM Spectrum Virtualize 8.5.

This publication is intended for storage and VMware administrators. The reader is expected to have a working knowledge of IBM Spectrum Virtualize and VMware. Initial storage and server setup is not covered.

This chapter includes the following sections:

1.1 IBM and VMware

IBM and VMware have a long record of collaboration. VMware was founded in 1998, and the beginning of IBM Spectrum Virtualize dates back to the early 2000s. Almost since inception, IBM and VMware have been technology partners.

IBM is a VMware Technology Alliance Partner. IBM storage is deployed in the VMware Reference Architectures Lab. Therefore, VMware products run well on IBM storage.

1.2 Overview of IBM Spectrum Virtualize and IBM FlashSystem

IBM Spectrum Virtualize refers to the storage software. *IBM FlashSystem* is a family brand name for the hardware.

For this book, we mostly refer to the storage as IBM FlashSystem. However, other IBM Spectrum Virtualize storage products work in similar fashion.

For more information, see the following IBM Redbooks publications:

- *Implementation Guide for IBM Spectrum Virtualize Version 8.5*, SG24-8520
- *Performance and Best Practices Guide for IBM Spectrum Virtualize 8.5*, SG24-8521

1.2.1 IBM Spectrum Virtualize

IBM Spectrum Virtualize refers to the software that runs on various IBM storage hardware. Its models include the IBM FlashSystem 5000, IBM FlashSystem 7300, IBM FlashSystem 9500, and IBM SAN Volume Controller (SVC). The former IBM Storwize family also uses IBM Spectrum Virtualize. Except for differences in hardware and licensing, all IBM Spectrum Virtualize products behave identically.

The primary function of IBM Spectrum Virtualize is *block-level storage virtualization*. IBM defines *storage virtualization* as a technology that makes one set of resources resemble another set of resources, preferably with more desirable characteristics.

The storage that is presented to the host is virtual and does not correspond to a specific back-end storage resource so that IBM Spectrum Virtualize can perform many enterprise-class features without impacting the hosts.

IBM Spectrum Virtualize first came to market in 2003 in the form of the IBM SVC. In 2003, the SVC was a cluster of commodity servers attached to a storage area network (SAN). The SVC did not contain its own storage. Instead, SVC used back-end storage that was provided from other storage systems. At the time of writing, IBM Spectrum Virtualize supports up to 500+ different storage controllers.

IBM Spectrum Virtualize offered many advantages, including:

- Data compression and deduplication
- Software and hardware encryption
- SafeGuarded Copy
- IBM Easy Tier® for workload balancing
- Multi-tenancy
- Thin-provisioning

- 2 - 3 site replication and point in time copy
- Software-defined storage (SDS) capability to cloud service providers (CSPs)

These features combine to make a compelling suite of storage management tools.

IBM Spectrum Virtualize is software, so new features and functions are added regularly. Over the years, IBM Spectrum Virtualize acquired a number of technologies that integrate it well with VMware.

Additionally, all IBM storage includes the following benefits:

- IBM Storage Insights, which provides a cloud-based dashboard to monitor all storage across the enterprise. It is excellent for tracking both performance and capacity usage.
- Remote support and remote upgrades.
- Backed by various premium support offerings, including Technical Advisors.

1.2.2 IBM FlashSystem

Since 2011, IBM Spectrum Virtualize is also available in various canister and enclosure hardware architectures. Originally branded under the Storwize name, these products are now called IBM FlashSystem. The lab environment that was used for this IBM Redbooks publication was built around an IBM FlashSystem 9200. These models include internal storage that is virtualized the same as with SVC. IBM FlashSystem products can also virtualize external storage, but the market has shifted focus toward using internal storage.

IBM FlashSystem 9200 supports a range of storage media with an emphasis on high-performance Non-Volatile Memory Express (NVMe) drives. For most storage administrators, capacity and performance optimization includes IBM FlashCore® Module (FCM) modules. FCMs are the next generation of IBM FlashSystem Micro Latency Modules; they offer high throughput and built-in compression without a performance penalty.

On the high end is Storage Class Memory (SCM), which is built around Intel's 3D-Xpoint and Samsung Z-NAND technologies. SCM offers unprecedented throughput and low latency, with limited capacities. The IBM FlashSystem 9500 control enclosure supports NVMe SCM/FCM and solid-state drives (SSDs), and serial-attached SCSI (SAS) SSDs are supported only by using expansion enclosures.

Figure 1-1 shows the front of an IBM FlashSystem 9200 enclosure. It is configured with 24 dual-ported NVMe drive slots.

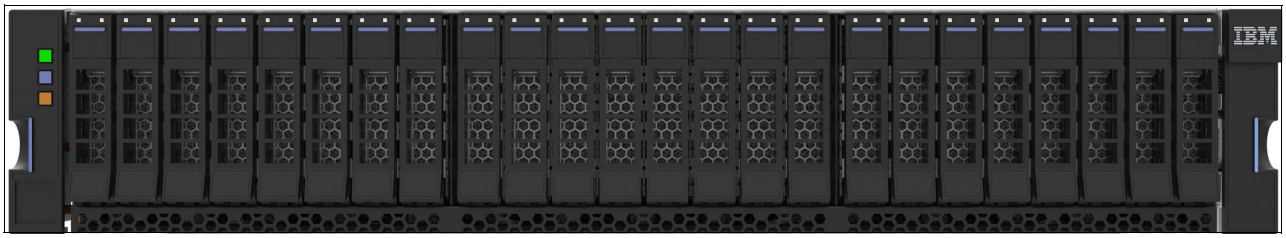


Figure 1-1 The front of an IBM FlashSystem 9200

Each system can be scaled up to include addition control enclosures or scaled out to include more expansion enclosures.

Figure 1-2 shows the back of an IBM FlashSystem 9200 enclosure. In the center are two canisters that run the IBM Spectrum Virtualize software and perform I/O. The canisters are identical, except that the upper canister is flipped upside down. The far left and right sides contain redundant power supplies.

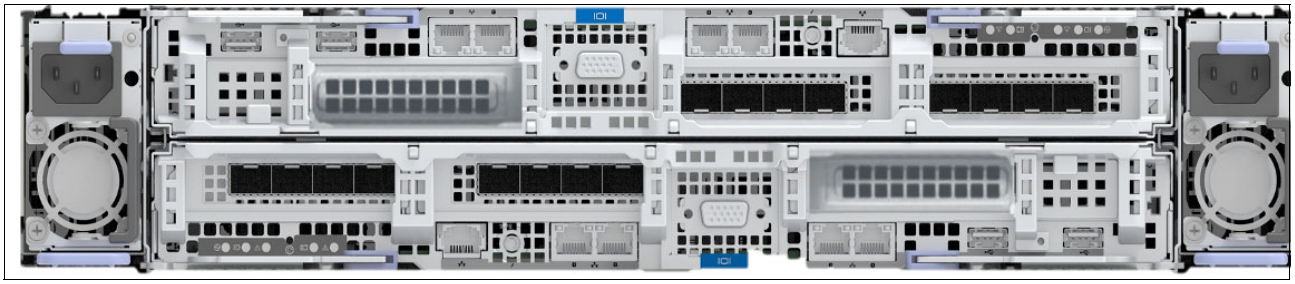


Figure 1-2 The back of an IBM FlashSystem 9200

IBM FlashSystem 9200 can be configured with various I/O ports that support various protocols. The most common configuration is 8 or 12 Fibre Channel (FC) ports. These ports can support the Fibre Channel protocol (FCP) or Non-Volatile Memory Express over Fibre Channel (FC-NVMe).

1.2.3 Key IBM Spectrum Virtualize terminology

Table 1-1 lists the key IBM Spectrum Virtualize terminology.

Table 1-1 Key IBM Spectrum Virtualize terminology

Term	Definition
Canister	The IBM FlashSystem hardware that runs the IBM Spectrum Virtualize software.
Node	The software representation of the canister in system.
I/O group	A pair of nodes or canisters that work together to service I/O.
Drive	FlashCore Module, Non-Volatile Memory Express (NVMe) SSD, or hard disk drive (HDD) storage hardware.
Array	A RAID array of drives.
Managed disk (MDisk)	Either an array of internal storage or a logical unit number (LUN) provided by an external storage controller.
Pool	A collection of MDisks that provide a pool or storage for allocation to volumes.
Volume or VDisk	The virtual LUN that is presented to the host.
SafeGuarded Copy	Immutable copies of primary volumes.
Host	The server that uses the storage. An ESXi server in this case.
IBM HyperSwap®	A business continuity solution that copies data across two sites.

1.2.4 Key VMware terminology

Table 1-2 lists the key VMware terminology.

Table 1-2 Key VMware terminology

Term	Definition
Host	An VMware ESXi server running on a physical server.
Cluster	A group of ESXi servers.
Data center	A group of ESXi host clusters.
Data store	SAN-based shared storage resources for ESXi clusters. Local disk-based data stores cannot be shared by multiple servers.
Native multipathing (NMP)	NMP plug-in. Most of the FC SAN-based storage systems are controlled by NMP.
High-performance plug-in (HPP)	NVMe-based storage systems are controlled by HPP.
Claim rules	Claim rules determine which multipathing module owns the paths to a storage device. They also define the type of multipathing support that the host provides to the device.
VAAI	VMware vSphere Storage APIs – Array Integration (VAAI), also referred to as hardware acceleration or hardware offload application programming interfaces (APIs), are a set of APIs to enable communication between VMware vSphere ESXi hosts and storage devices.
VMware vSphere Virtual Volumes (vVols)	vVols are virtual machine disk (VMDK) granular storage entities that are exported by storage arrays. vVols are exported to the ESXi host through a small set of Protocol Endpoints (PEs). PEs are part of the physical storage fabric, and they establish a data path from virtual machines (VMs) to their respective vVols on demand. Storage systems enable data services on vVols. The results of these data services are newer vVols. Data services configuration and management of virtual volume systems are exclusively done out-of-band regarding the data path.

1.3 Overview of IBM FlashSystem with VMware

Figure 1-3 summarizes the various of VMware and IBM software components that are discussed in this publication.

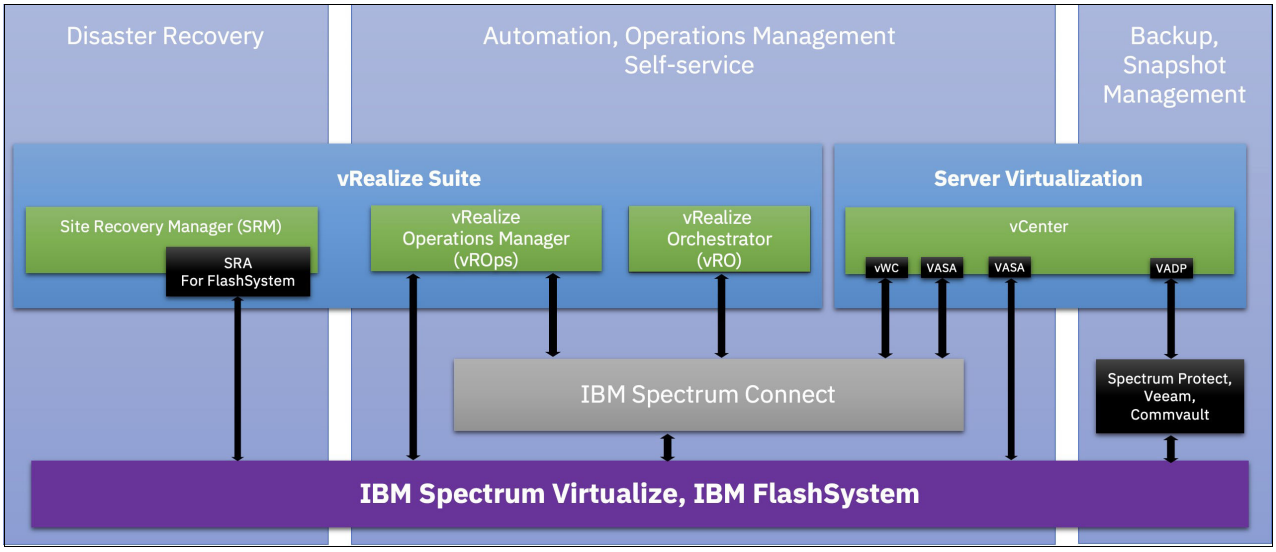


Figure 1-3 Integrating IBM Storage with VMware

Two key components of integrating IBM Spectrum Virtualize with VMware are as follows:

- IBM Spectrum Connect, which is a no additional charge software solution that is available with all IBM storage. For more information about IBM Spectrum Connect, see Chapter 4, “Integrating with VMware by using IBM Spectrum Connect” on page 41.
- Integration with VMware Site Recovery Manager, which is not part of IBM Spectrum Connect. For more information about Integration with VMware Site Recovery Manager, see Chapter 5, “VMware and IBM Spectrum Virtualize multi-site guidelines” on page 105.

2

Host and storage connectivity

This chapter describes the test environment setup that is used in this book. It also discusses options for host to storage connectivity, which includes host-cluster configuration, protocols such as Fibre Channel and Internet Small Computer System Interface (iSCSI), iSCSI Extensions for RDMA (iSER), and multipath configuration options.

This chapter includes the following sections:

2.1 Test environment implementation

Figure 2-1 depicts the configuration and connectivity for the test environment for host connectivity that is used in this book. An IBM FlashSystem 9100 and an IBM FlashSystem 9200 are cabled to two IBM SAN24B-5 Fibre Channel switches. Each canister on each IBM FlashSystem has two ports that are connected to each switch. Four VMware hosts are used for this book.

Each host is connected to each switch:

- The hosts with the blue-colored connections are zoned to the FS9100 with blue-colored connections.
- The hosts with the green-colored connections are zoned to the FS9200 with green-colored connections.

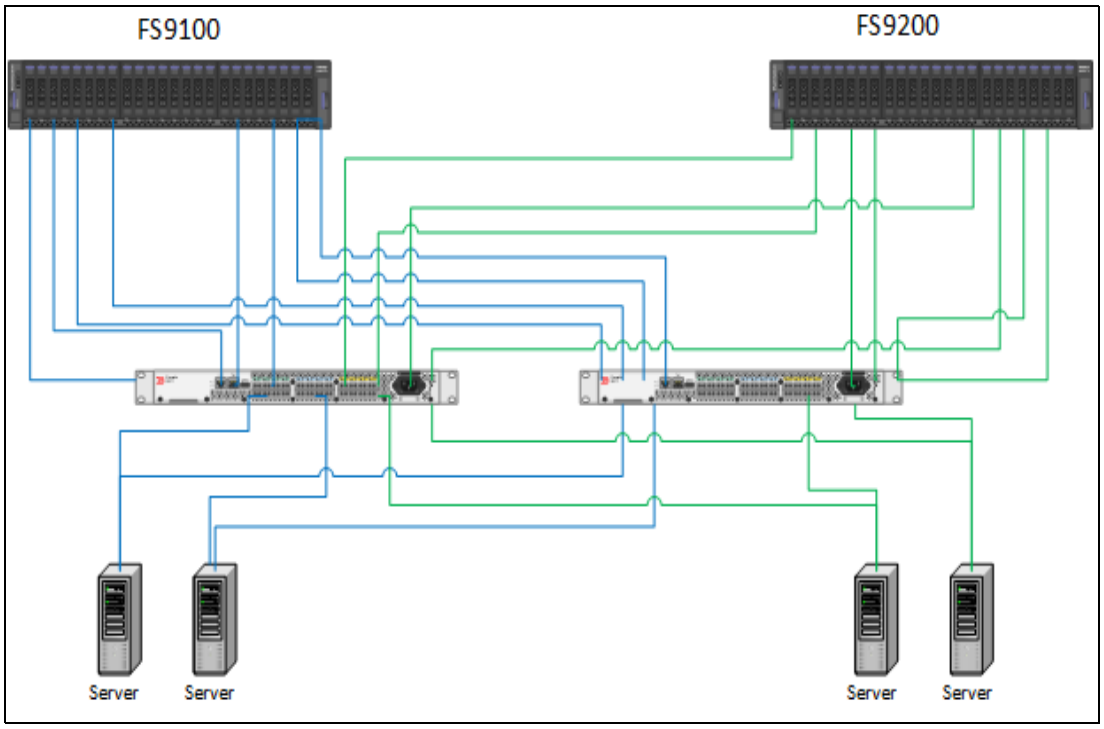


Figure 2-1 Configuration and connectivity for the test environment that is used in this book

2.1.1 IBM FlashSystem host clusters

IBM FlashSystem products support host clusters starting with IBM Spectrum Virtualize 7.7.1. With a host cluster, a user can create a group of hosts to form a cluster. A cluster is treated as a single entity, which allows multiple hosts to have access to the same set of volumes.

Volumes that are mapped to a host cluster are assigned to all members of the host cluster that use the same Small Computer System Interface (SCSI) ID. Before this feature was implemented in IBM Spectrum Virtualize, as an example, an Elastic Sky X (ESX) cluster would be created as a single host object, containing all the worldwide node names (WWNNs) for the hosts in the cluster, up to 32.

With host clusters, a storage administrator can define individual host objects for each Elastic Sky X integrated (ESXi) host and add them to a host cluster object that represents each vSphere cluster. If hosts are later added to the host cluster, they automatically inherit shared host cluster volume mappings. Within the host cluster object, you can have up to 128 hosts in a single host cluster object. Host clusters are easier to manage than single host objects because the 32 worldwide port name (WWPN) limitation is removed.

The minimum size of a cluster is two nodes for vSphere high availability (HA) to protect workloads if one host stops functioning. However, in most use cases, a 3-node cluster is more appropriate because you have the option of running maintenance tasks on an ESXi server without having to disable HA.

Configuring large clusters has benefits too. You typically have a higher consolidation ratio, but there might be a downside if you do not have enterprise-class or correctly sized storage in the infrastructure. If a data store is presented to a 32-node or a 64-node cluster and the virtual machines (VMs) on that data store are spread across the cluster, there is a chance that you will run into SCSI-locking contention issues. Using a VMware vSphere Storage APIs – Array Integration (VAAI) aware array helps reduce this problem with Atomic Test and Set (ATS). However, if possible, consider starting small and gradually growing the cluster size to verify that your storage behavior is not impacted.

Figure 2-2 shows one of the VMware host clusters that was used in the test configuration for this book. There are two hosts that are defined in the VMware host cluster.

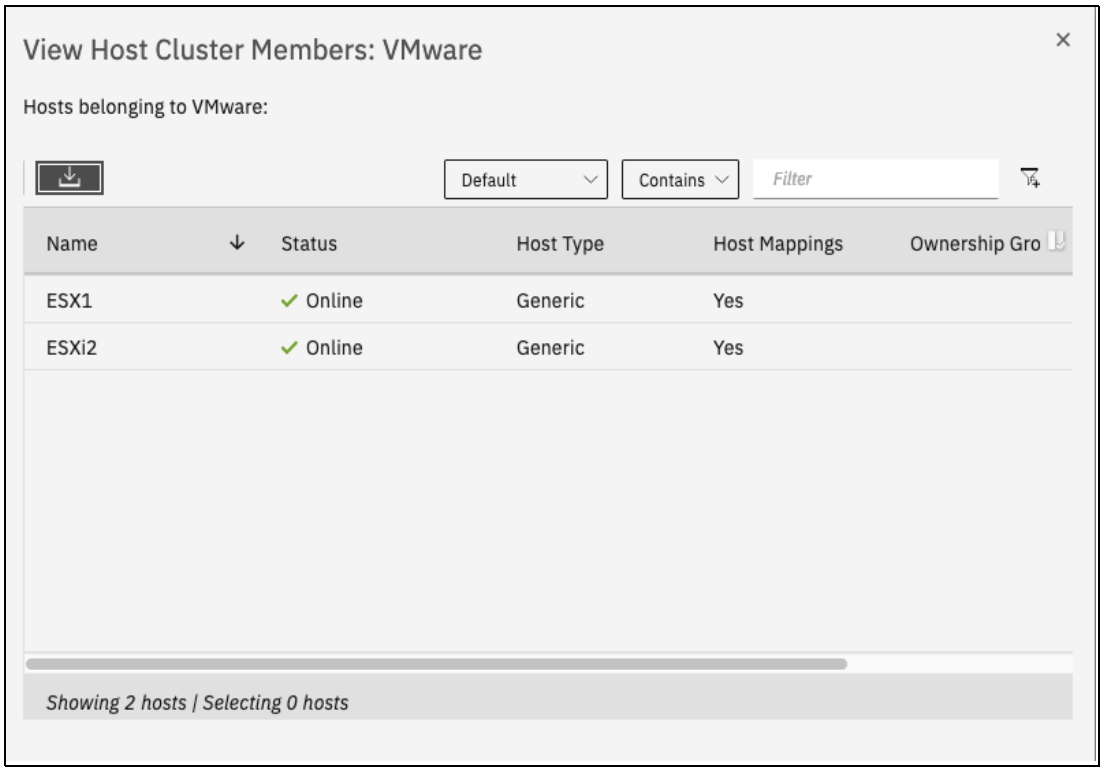


Figure 2-2 VMware host clusters that were used in the test configuration

Note: Do not add Non-Volatile Memory Express (NVMe) hosts and SCSI hosts to the same host cluster.

Figure 2-3 shows the volumes that are assigned to the host cluster.

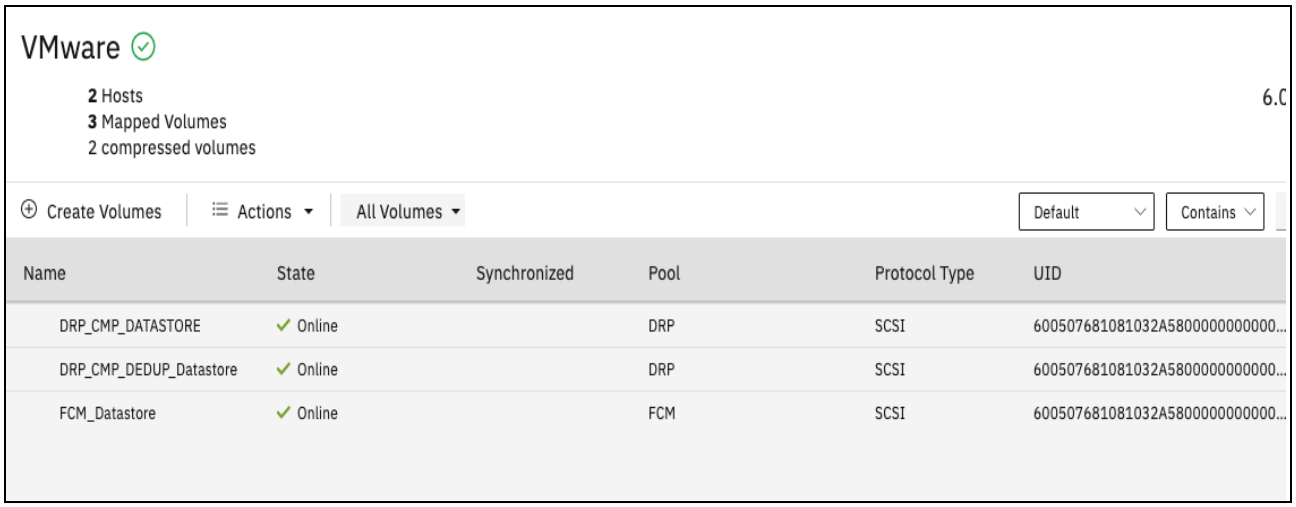


Figure 2-3 Volumes that are assigned to the host cluster

Figure 2-4 shows one of the hosts in the host cluster with the volumes that are connected to it. The volumes were assigned to the host cluster, and not directly to the host. Any hosts that are added to a host cluster have all of the volumes mapped to the host cluster automatically assigned to the hosts.

Note: Private mappings can still be provisioned to individual hosts within a host cluster, for example for storage area network (SAN) Boot configurations.

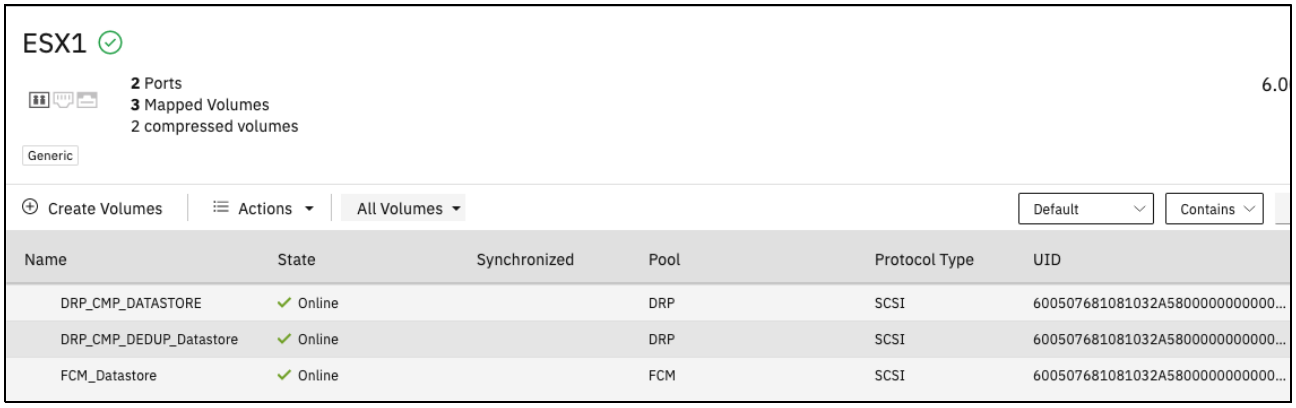


Figure 2-4 One of the hosts in the host cluster with the volumes that are connected to it

2.1.2 Use cases for implementing throttles

With IBM FlashSystem storage you can configure throttles for the following items:

- Hosts
- Host clusters
- Volumes
- SCSI offload
- Storage pools

A common use case for throttles on hosts, host clusters, and volumes can be applied when test and production workloads are mixed on the same IBM FlashSystem. Test-related workloads should not affect production, so you can throttle test hosts and volumes to give priority to production workloads.

IBM Spectrum Virtualize supports commands that are used for SCSI offload and VMware VAAI.

- SCSI offload enables the host to offload some data operations to the storage system.
- VAAI enables VMware hosts to also offload some operations to supported storage systems.

Both technologies reduce traffic on the storage network, and load on the host. Hosts use these offload commands to perform tasks such as formatting new file systems or performing data copy operations without a host needing to read and write data. Examples are the **WRITE SAME** and **XCOPY** commands. IBM Spectrum Virtualize 8.1.0.0 introduced support for **WRITE SAME** when **UNMAP** is enabled. **WRITE SAME** is a SCSI command that tells the storage system to write the same pattern to a volume or an area of a volume.

When SCSI **UNMAP** is enabled on IBM FlashSystem storage, it advertises this situation to hosts. At versions 8.1.0.0 and later, some hosts respond to the **UNMAP** command by issuing a **WRITE SAME** command, which can generate large amounts of I/O. If the back-end storage system cannot handle the amount of I/O, volume performance can be impacted. IBM Spectrum Virtualize offload throttling can limit the concurrent I/O that is generated by the **WRITE SAME** or **XCOPY** commands.

To enable offload throttling:

1. Run the **svcupgradetest** utility to obtain the recommended bandwidth throttle value:
   a. For systems managing any enterprise or nearline storage, the recommended value is 100 MBps.
   b. For systems managing *only* tier1 flash or tier0 flash, the recommended value is 1000 MBps.
2. Enable the offload throttle by using the following command-line interface (CLI) command:

```
throttle -type offload -bandwidth bandwidth_limit_in_MB
```

2.1.3 Data reduction pools

Data reduction can increase storage efficiency and reduce storage costs, especially for IBM FlashSystem storage systems. Data reduction reduces the amount of data that is stored on both the internal drives and the virtualized external storage systems by reclaiming previously used storage resources that are no longer needed by host systems.

IBM FlashSystem storage systems implement data reduction by using data reduction pools (DRPs). A DRP can contain thin-provisioned or compressed volumes. DRPs also provide more capacity to volumes in the pool by supporting data deduplication.

With a log-structured pool implementation, DRPs help to deliver more consistent performance from compressed volumes. DRPs also support compression of all volumes in a system, potentially extending the benefits of compression to all data in a system. Traditional storage pools have a fixed allocation unit of an extent, and that does not change with DRPs. However, features like *Thin Provisioning* and *IBM Real-time Compression (RtC)* use smaller allocation units and manage this allocation with their own metadata structures. These features are described as *Binary Trees* or *Log Structured Arrays (LSAs)*.

For thin-provisioned volumes to stay thin, you need to be able to reclaim capacity that is no longer used, or for LSAs (where all writes go to new capacity), garbage-collect the old overwritten data blocks. This action also needs to be done at the smaller allocation unit size in a DRP volume.

Figure 2-5 shows the types of volumes that can be created in a DRP.

- DRP fully allocated volumes provide the best performance for the IBM FlashSystem products, but storage efficiency and space savings are not realized.
- Thin-compressed volumes provide storage-space efficiency with the best performance of the four options for space-efficient volumes.

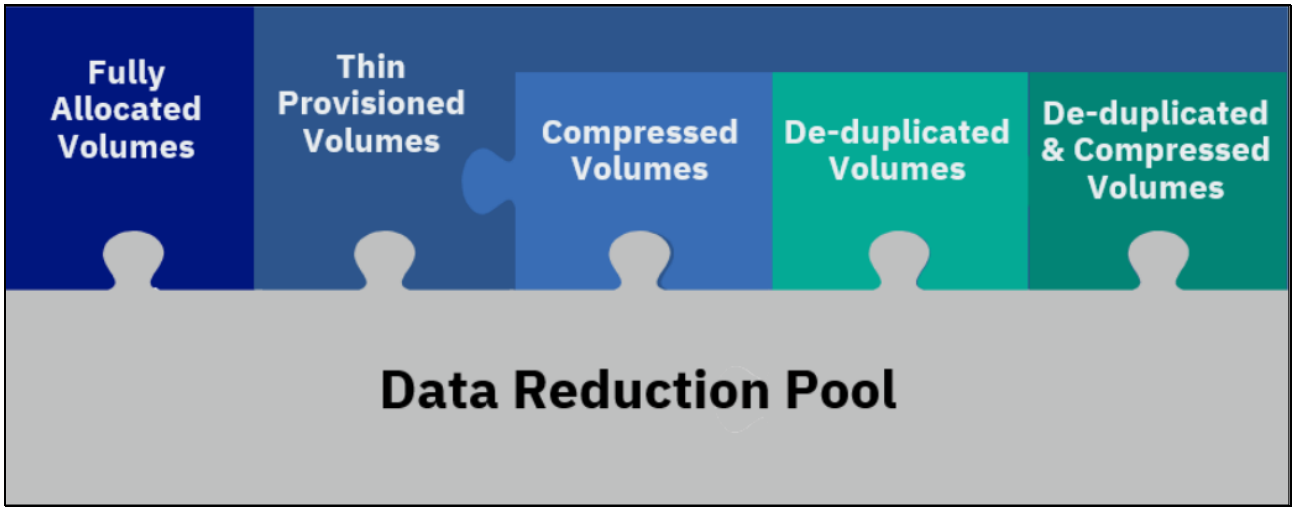


Figure 2-5 Types of volumes that can be created in a data reduction pool

For more information about data reduction pools, see *Implementation Guide for IBM Spectrum Virtualize Version 8.5*, SG24-8520.

2.2 Host connectivity protocols

IBM FlashSystem support both Ethernet-based and Fibre-Channel-based host-attachment protocols:

- Ethernet-based protocols include iSCSI and iSER. Both of these protocols can be implemented over existing Ethernet networks and do not require a dedicated storage network.
- Fibre-Channel-based protocols include Non-Volatile Memory Express over Fibre Channel (FC-NVMe) and traditional SCSI Fibre Channel, which is most often referred to as Fibre Channel.

2.2.1 iSCSI

iSCSI connectivity is a software feature that is provided by the SAN Volume Controller (SVC) code. The iSCSI protocol is a block-level protocol that encapsulates SCSI commands into Transmission Control Protocol/Internet Protocol (TCP/IP) packets. Therefore, iSCSI uses an IP network rather than requiring the Fibre Channel (FC) infrastructure. The iSCSI standard is defined by Request for Comment (RFC) 3720.

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

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

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

Ensure that the iSCSI initiators and targets that you plan to use are supported. Use the following sites for reference:

- IBM FlashSystem 8.5 Support Matrix
- IBM FlashSystem 9x00 8.5
- IBM System Storage Interoperation Center (SSIC)

iSCSI qualified name

An IBM FlashSystem cluster can provide up to eight iSCSI targets, one per canister. Each canister has its own iSCSI Qualified Name (IQN), which, by default, is in the following format:

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

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

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

2.2.2 iSER

IBM FlashSystem that run IBM Spectrum Virtualize v8.2.1 or greater support iSER for host attachment, which is implemented by using RDMA over Converged Ethernet (RoCE) or Internet Wide-Area RDMA Protocol (iWARP). This feature supports a fully Ethernet-based infrastructure (and not Fibre Channel) in your data center:

- IBM FlashSystem internode communication with 2 or more IBM FlashSystem in a cluster.
- HyperSwap.

Using iSER requires that an Ethernet adapter is installed in each node, and that dedicated Remote Direct Memory Access (RDMA) ports are used for internode communication. RDMA enables the Ethernet adapter to transfer data directly between nodes. The direct transfer of data bypasses the central processing unit (CPU) and cache and makes transfers faster.

Requirements for RDMA connections:

- 25 Gbps Ethernet adapter is installed on each node.
- Ethernet cables between each node are connected correctly.
- Protocols on the source and destination adapters are the same.
- Local and remote IP addresses can be reached.
- Each IP address is unique.
- The local and remote port virtual LAN identifiers are the same.

- A minimum of two dedicated ports are required for node-to-node RDMA communications to ensure the best performance and reliability. These ports cannot be used for host attachment, external storage, or IP replication traffic.
- A maximum of four ports per node are allowed for node-to-node RDMA connections.

2.2.3 FC-NVMe

The NVMe transport protocol provides enhanced performance on high-demand IBM FlashSystem drives. NVMe is a logical device interface specification for accessing non-volatile storage media. Host hardware and software use NVMe to fully leverage the levels of parallelism possible in modern solid-state drives (SSDs).

Compared to the SCSI protocol, NVMe improves I/O and brings performance improvements such as multiple, long command queues, and reduced latency. SCSI has one queue for commands, unless multi-queue support such as *blk_mq* is enabled on the operating system, and you are limited to the number of cores in the CPUs on the host.

NVMe is designed to have up to 64 thousand queues. In turn, each of those queues can have up to 64 thousand commands that are processed simultaneously. This queue depth is much larger than SCSI typically has. NVMe also streamlines the list of commands to only the basic commands that Flash technologies need.

IBM FlashSystem implements NVMe by using the FC-NVMe protocol. FC-NVMe uses the Fibre Channel protocol as the transport so that data can be transferred from host memory to the target, which is similar to RDMA. For more information about NVMe, see *IBM Storage and the NVM Express Revolution*, REDP-5437.

Every physical FC port on IBM FlashSystem storage supports four virtual ports: one for SCSI host connectivity, one for FC-NVMe host connectivity, one for SCSI host failover, and one for FC-NVMe host failover. Every NVMe virtual port supports the functions of NVMe discovery controllers and NVMe I/O controllers. Hosts create associations (NVMe logins) to the discovery controllers to discover volumes or to I/O controllers to complete I/O operations on NVMe volumes. Up to 128 discovery associations are allowed per node, and up to 128 I/O associations are allowed per node. An extra 128 discovery associations and 128 I/O associations per node are allowed during N_Port ID virtualization (NPIV) failover.

At the time of this writing, IBM Spectrum Virtualize 8.5 supports a maximum of 64 NVMe hosts. For more information, see IBM Support.

If FC-NVMe is enabled on the IBM FlashSystem, each physical WWPN reports up to four virtual WWPNs. Table 2-1 lists the NPIV ports and port usage when FC-NVMe is enabled.

Table 2-1 NPIV ports and port usage when FC-NVMe is enabled

NPIV port	Port description
Primary Port	The WWPN that communicates with back-end storage if the IBM FlashSystem is virtualizing any external storage.
SCSI Host Attach Port	The virtual WWPN that is used for SCSI attachment to hosts. This WWPN is a target port only.
Failover SCSI Host Port	The standby WWPN that is brought online only if the partner node in an I/O group goes offline. This WWPN is the same WWPN as the primary host WWPN of the partner node.

NPIV port	Port description
NVMe Host Attach Port	The WWPN that communicates with hosts for FC-NVMe. This WWPN is a target port only.
Failover NVMe Host Attach Port	The standby WWPN that is brought online only if the partner node in an I/O group goes offline. This WWPN is the same WWPN as the primary host WWPN of the partner node.

For more information about FC-NVMe and configuring hosts to connect to IBM FlashSystem storage systems by using FC-NVMe, see VMware ESXi installation and configuration for NVMe over Fibre Channel hosts

Figure 2-6 shows how to add a FC-NVMe host to an IBM FlashSystem from the Add Host window.

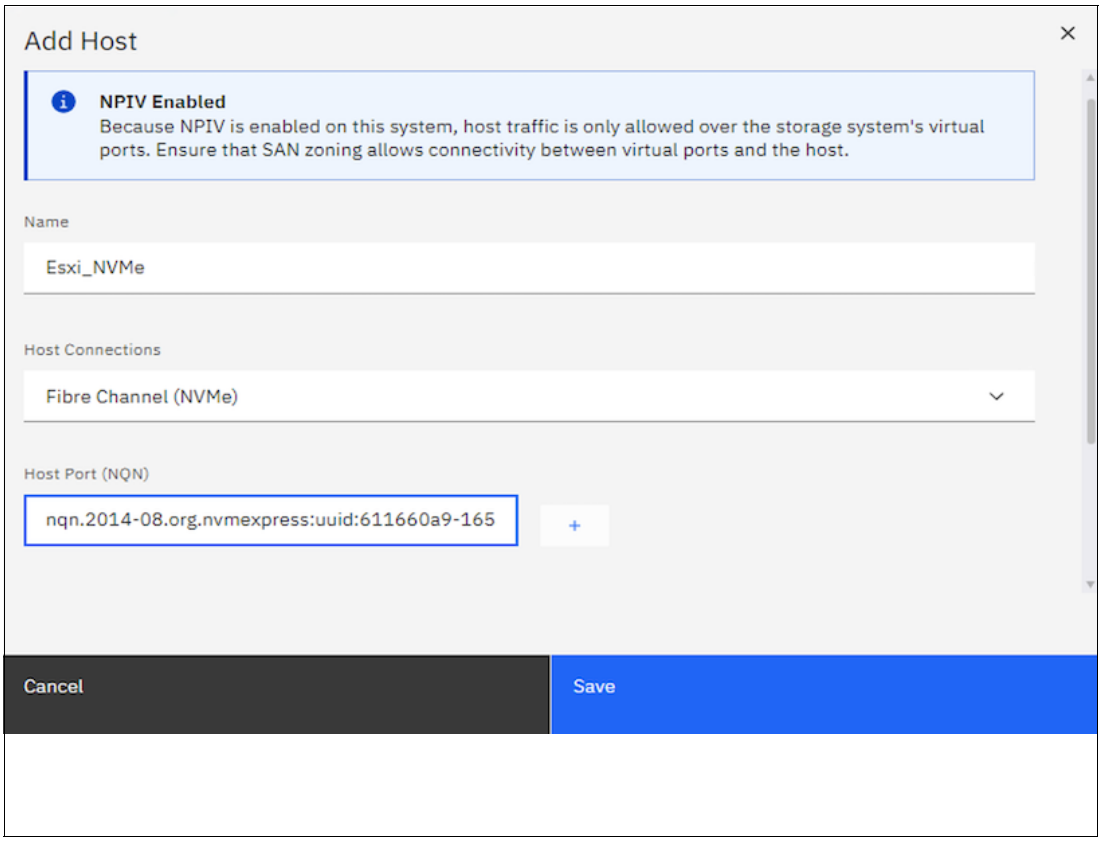


Figure 2-6 Adding an FC-NVMe host to IBM FlashSystem

After adding a VMware FC-NVMe host to an IBM FlashSystem, you must discover the storage subsystem on the VMware ESXi host. You must discover both nodes for each adapter on the IBM FlashSystem by using the NVMe WWPNs that are zoned with the ESXi host. To do so, click **DISCOVER CONTROLLERS** in the window that is shown in Figure 2-7.

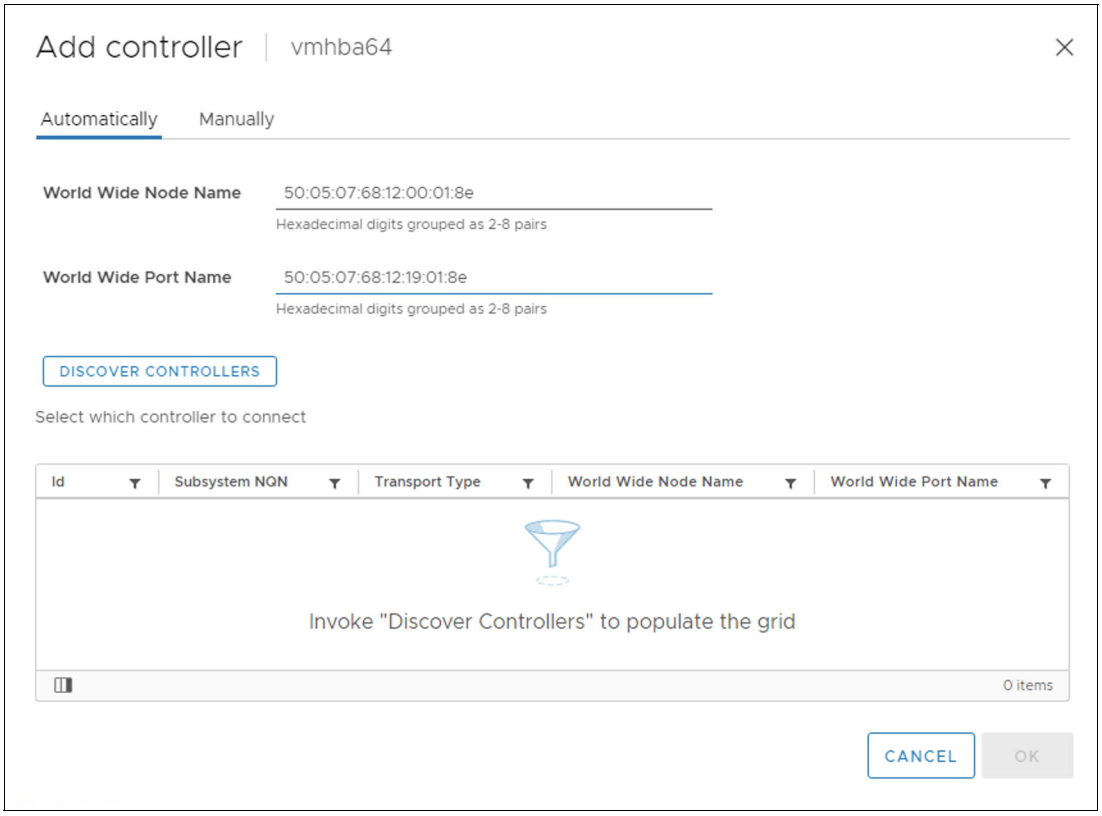


Figure 2-7 Discovering the IBM FlashSystem NVMe controller on VMware

The window that is shown in Figure 2-8 opens and shows the available SCSI and NVMe devices and data stores.

Name	State	Pool	Protocol Type	UID	Host Mappings	Capacity
ESXVOL_01	Online	VmwarePool	SCSI	600507681280000C70000000000004...	Yes	500.00 GiB
ESXVOL_02	Online	VmwarePool	SCSI	600507681280000C70000000000004...	Yes	500.00 GiB
NVMEVOL_01	Online	VmwarePool	NVMe	70000000000004B40050760812800...	Yes	501.00 GiB
NVMEVOL_02	Online	VmwarePool	NVMe	70000000000004B50050760812800...	Yes	501.00 GiB
FILES	Online	VmwarePool	SCSI	600507681280000C70000000000004...	Yes	100.00 GiB

Figure 2-8 SCSI and NVMe devices and data stores

NVMe devices are managed by the VMware high-performance plug-in (HPP). To see the NVMe devices, run the **`esxcli storage hpp device list`** command on `esxcli`.

2.2.4 SCSI Fibre Channel

FC is a storage networking transport that transports SCSI commands and data from hosts to storage. The Fibre Channel protocol (FCP) is the transport layer that transmits the SCSI commands. Hosts and storage systems are connected to switches to create a SAN.

Figure 2-1 on page 8 is an example of two single-switch SANs that are used to create this book.

The IBM Spectrum Virtualize software that runs on IBM FlashSystem storage uses the SCSI protocol to communicate with its clients, and presents storage space in form of SCSI logical units (LUs) identified by SCSI logical unit numbers (LUNs).

Note: In formal practice, LUs and LUNs are different entities. In practice, the term LUN is often used to refer to a logical disk or LU.

Since most applications do not directly access storage, but work with files or records, the OS of a host must convert these abstractions to the language of storage, which are vectors of storage blocks that are identified by logical block addresses within an LU. In IBM Spectrum Virtualize, each of the externally visible LUs is internally represented by a volume, which is an amount of storage that is taken out of a storage pool. Hosts use the SCSI protocol to send I/O commands to IBM FlashSystem storage to read and write data to these LUNs.

As with FC-NVMe host attachment, if NPIV is enabled on the IBM FlashSystem storage system, hosts attach to a virtual WWPN. Table 2-1 on page 14 lists the SCSI and Failover Host Attach Ports.

2.2.5 NVMe over Remote Direct Memory Access

IBM Spectrum Virtualize 8.5.0 can be attached to an NVMe host through NVMe over Remote Direct Memory Access (NVMe over RDMA). NVMe over RDMA uses RoCE v2 as the transport protocol. RoCE v2 is based on UDP. RDMA is a host-offload and host-bypass technology that allows an application (including storage) to make data transfers directly to and from another application's memory space. The RDMA-capable Ethernet network interface cards (RNICs), and not the host, manage reliable data transfers between source and destination.

VMware V7.0u2 and later supports RoCE v2 as host connectivity for IBM Spectrum Virtualize 8.5 storage systems.

RNICs can use RDMA over Ethernet through RoCE encapsulation. RoCE wraps standard InfiniBand payloads with Ethernet or IP over Ethernet frames, which is sometimes called InfiniBand over Ethernet. There are two main RoCE encapsulation types:

- RoCE v1

  Uses dedicated Ethernet Protocol Encapsulation to send Ethernet packets between source and destination MAC addresses by using Ethertype 0x8915.

- RoCE v2

  Uses dedicated UDP over Ethernet Protocol Encapsulation to send IP UDP packets by using port 4791 between source and destination IP addresses. UDP packets are sent over Ethernet by using source and destination MAC addresses.

  RoCE v2 is not compatible with other Ethernet options, such as RoCE v1.

Note: Unlike RoCE v1, RoCE v2 is routable.

For more information about configuring the VMware ESXi for NVMe over RDMA on IBM FlashSystem storage systems, see Configuring the VMware ESXi operating system for NVMe over RDMA host.

2.3 Multi-path considerations

This section describes multi-path considerations, such as path selection policies and zoning considerations.

2.3.1 Native multipathing path-selection policies

There are three general VMware native multipathing (NMP) plug-in path-selection policies or path-selection plug-ins (PSPs). A PSP is a VMware ESXi host setting that defines a path policy to an LUN. The three PSPs are *Most Recently Used (MRU)*, *Fixed*, and *Round-Robin (RR)*.

Most Recently Used

The policy selects the first working path, discovered at system start time. If this path becomes unavailable, the ESXi or ESX host switches to an alternative path and continues to use the new path while it is available. This policy is the default for LUNs that are presented from an Active/Passive array. ESXi and ESX host switches do not return to the previous path if it returns, and the host switch remains on the working path until the working path fails.

Tip: The VMware **`preferred`** flag can be set on a path. This flag is not applicable if the path selection policy is set to **`Most Recently Used`**.

Fixed

The *Fixed* policy uses the designated preferred path flag if it is configured. Otherwise, it uses the first working path that is discovered at system start time. If the ESXi host cannot use the preferred path or it becomes unavailable, the ESXi host selects an alternative available path. The host automatically returns to the previously defined preferred path when it becomes available. This policy is the default for LUNs that are presented from an active/active storage array.

Round-Robin

The *Round-Robin* policy is the recommended policy for IBM FlashSystem products. This path selection policy uses a round-robin algorithm to load balance paths across all LUNs when connecting to a storage array. This policy is the default for VMware starting with ESXi 5.5. You must explicitly set Round-Robin for versions earlier than ESXi 5.5.

Data can travel through only one path at a time:

- For active/passive storage arrays, only the paths to the preferred storage array are used.
- For an active/active storage array, all paths are used for transferring data, assuming that paths to the preferred yes are available.

With Asymmetric Logical Unit Access (ALUA) in an active/active storage array, such as the IBM FlashSystem 9200 and 9500 systems, only the optimized paths to the preferred control enclosure node are used for transferring data. Round-Robin cycles through only those optimized paths. You should configure pathing so that half the LUNs are preferred by one control enclosure node, and the other half are preferred by the other control enclosure node.

Latency Round-Robin is activated by default when Round-Robin is selected as the path selection policy. Latency Round-Robin considers I/O bandwidth and path latency when selecting an optimal path for I/O. When this latency mechanism is used, Round-Robin dynamically selects the best path for better load-balancing. For more information about Latency Round-Robin, see Change Default Parameters for Latency Round-Robin.

Round-Robin path selection limit

Round-Robin Path switching supports two limits:

- Input/output operations per second (IOPS) limit: A new path is used after a specified number of IOPS are completed on the current path.
- Bytes limit: A new path is used after a specified number of bytes are transferred on the current path.

The default path selection limit is IOPS, and the default value is 1000 IOPS before the path changes. In some cases, a host can experience latency to storage with no latency seen on the SAN. In these cases, the load of 1000 IOPS saturates the bandwidth of the path. Lowering this value can increase storage performance and help prevent this cause of latency. The recommended path-selection limit setting for IBM FlashSystem is to use IOPS and set the value to 1. For more information about the IOPS limit, see Adjusting Round Robin IOPS limit from default 1000 to 1 (2069356).

Path selection with claim rules

Claim rules help to set the path selection limit and path selection policy settings in new LUNs that are assigned to ESXi host. Example 2-1 shows an example of creating a claim rule.

Example 2-1 Creating a claim rule for an IBM Spectrum Virtualize system to set the path selection limit to 1

```
esxcli storage nmp satp rule add -s VMW_SATP_ALUA -V IBM -M "2145" -c tpgs_on --psp="VMW_PSP_RR"
-e "IBM arrays with ALUA support" -O "iops=1"
```

To configure the claim rule, use vSphere Host Profile window, as shown in Figure 2-9.

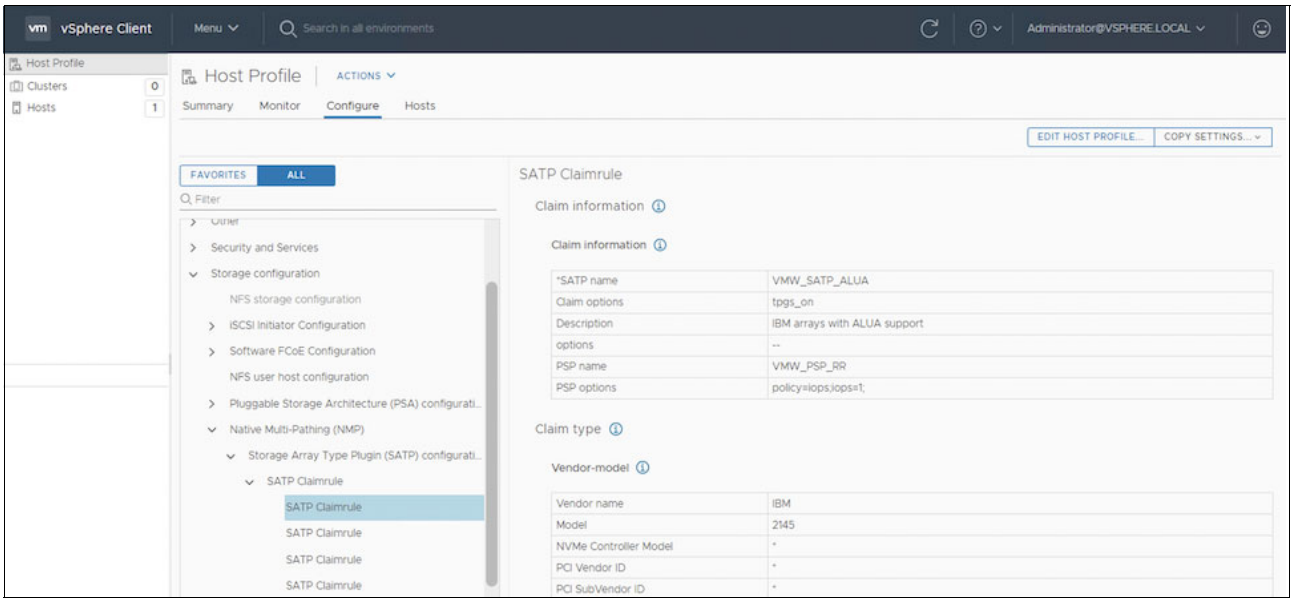


Figure 2-9 Configuring a claim rule by using the Host Profile window

Note: Existing and previously presented devices must be manually set to Round-Robin with an IOPS limit of 1. Optionally, the ESXi host can be restarted so that it can inherit the multipathing configuration that is set by the new rule.

2.3.2 High-performance plug-in and path selection policies

Since VMware 6.7, there is a new multipath plug-in that is called HPP. The HPP replaces the NMP for high-speed devices, such as NVMe. The HPP is the default plug-in that claims NVMe over Fabrics (NVMe-oF) targets. Within ESXi, the NVMe-oF targets are emulated and presented to users as SCSI targets. The HPP supports only active/active and implicit ALUA targets.

In vSphere 7.0 Update 1 and earlier, NMP remains the default plug-in for local NVMe devices, but you can replace it with HPP. Starting with vSphere 7.0 Update 2, HPP becomes the default plug-in for local NVMe and SCSI devices, but you can replace it with NMP.

Consider the following configuration recommendations for HPP:

- Use the vSphere version that supports HPP.
- Use HPP for local NVMe and SCSI devices, and NVMe-oF devices.
- If you use NVMe-oF, do not mix transport types to access the same namespace.
- Configure your VMs to use VMware Paravirtual controllers.
- Set the latency sensitive threshold to bypass the I/O scheduler.
- If a single VM drives a significant share of the device's I/O workload, consider spreading the I/O across multiple virtual disks. Attach the disks to separate virtual controllers in the VM.

By default, ESXi passes every I/O through the I/O scheduler. However, using the scheduler might create internal queuing, which is not efficient with the high-speed storage devices.

You can configure the latency sensitive threshold and enable the direct submission mechanism that helps I/O to bypass the scheduler. With this mechanism enabled, the I/O passes directly from Pluggable Storage Architecture (PSA) through the HPP to the device driver.

For the direct submission to work properly, the observed average I/O latency must be lower than the latency threshold that you specify. If the I/O latency exceeds the latency threshold, the system stops the direct submission, and temporarily reverts to using the I/O scheduler. The direct submission is resumed when the average I/O latency drops below the latency threshold again.

Note: HPP does not benefit when the systems perform lower than 200,000 IOPS.

Example 2-2 shows how to list the devices that are controlled by the HPP.

Example 2-2 Output of using the esxcli storage hpp device list command

```
[root@localhost:~] esxcli storage hpp device list
eui.70000000000004b5005076081280000c
   Device Display Name: NVMe Fibre Channel Disk (eui.70000000000004b5005076081280000c)
   Path Selection Scheme: LB-RR
   Path Selection Scheme Config: {iops=1,bytes=10485760;}
   Current Path: vmhba64:C0:T0:L3
   Working Path Set: vmhba64:C0:T0:L3, vmhba65:C0:T0:L3
   Is SSD: true
   Is Local: false
   Paths: vmhba64:C0:T0:L3, vmhba65:C0:T1:L3, vmhba65:C0:T0:L3, vmhba64:C0:T1:L3
   Use ANO: false
```

To support multipathing, HPP uses the Path Selection Schemes (PSSs) when selecting physical paths for I/O requests.

ESXi supports the following path selection mechanisms for HPP.

Load Balance - Round-Robin

Load Balance - Round-Robin (LB-RR) is the default scheme for the devices that are claimed by HPP. After transferring a specified number of bytes or I/Os on a current path, the scheme selects the path by using the Round-Robin algorithm.

Load Balance - Latency

To achieve better load-balancing results, Load Balance - Latency (LB-Latency) dynamically selects an optimal path by considering the following path characteristics:

- The `latency evaluation time` parameter indicates at what time interval, in milliseconds, that the latency of paths must be evaluated.
- The `sampling I/Os per path` parameter controls how many sample I/Os must be issued on each path to calculate the latency of the path.

Load Balance - IOPS

When using Load Balance - IOPS (LB-IOPS), after transferring a specified number of I/Os on a current path (the default is 1000), the system selects an optimal path that has the least number of outstanding I/Os.

Load Balance - Bytes

When using Load Balance - Bytes (LB-BYTES), after transferring a specified number of bytes on a current path (the default is 10 MB), the system selects an optimal path that has the least number of outstanding bytes.

Fixed

With this scheme, a designated preferred path is used for I/O requests. If the preferred path is not assigned, the host selects the first working path that is discovered at start time. If the preferred path becomes unavailable, the host selects an alternative available path. The host returns to the previously defined preferred path when it becomes available again.

When you configure `FIXED` as a path selection mechanism, select the preferred path.

2.4 Zoning considerations

Modern SAN switches have three types of zoning available:

- Port zoning
- Worldwide node name (WWNN) zoning
- WWPN zoning

The preferred method is to use only WWPN zoning. You should not mix zoning types. WWPN-based zoning is more flexible than Switchport-based and is required if the IBM Spectrum Virtualize NPIV feature is enabled. Switch-port based zoning can cause failover of the NPIV ports to not work correctly, and in certain configurations can cause a host to be connected to the IBM FlashSystem on both the physical and virtual WWPNs.

For more information about the NPIV feature and switch-port zoning, see Using Switch Port-Based Zoning with the IBM Spectrum Virtualize NPIV Feature.

A common misconception is that WWPN zoning provides poorer security than port zoning. However, modern SAN switches enforce the zoning configuration directly in the switch hardware. Also, you can use port-binding functions on SAN switches to enforce a WWPN to be connected to a particular SAN switch port, or to prevent unauthorized devices from logging in to your fabric if they are connected to switch ports. Lastly, the default zone on each of your virtual fabrics should have a zone policy of deny, which means that any device in the default zone cannot communicate with any other device on the fabric. All unzoned devices that are not in at least one named zone) are in the default zone.

Naming convention

When you create and maintain a Storage Network zoning configuration, you must have a defined naming convention and zoning scheme. If you do not define a naming convention and zoning scheme, your zoning configuration can be difficult to understand and maintain. Environments have different requirements, which means that the level of detailing in the zoning scheme varies among environments of various sizes. Therefore, ensure that you have an understandable scheme with an appropriate level of detailing for your environment. Then, use it consistently whenever you change the environment.

Aliases

Use zoning aliases when you create your IBM FlashSystem zones. Aliases make your zoning easier to configure and understand and minimize errors. Define aliases for the IBM FlashSystem physical WWPNs, the SCSI-FC WWPNs, and the FC-NVMe WWPNs if you have that feature enabled.

You should have the following zones:

- A zone containing all of the IBM FlashSystem aliases for the IBM FlashSystem physical WWPNs that are dedicated for internode use.
- A zone containing all of the IBM FlashSystem aliases for both the local and remote IBM FlashSystem physical WWPNs that are dedicated for partner use if replication is enabled.
- One zone for each host containing the aliases for the host and either the IBM FlashSystem SCSI-FC virtual WWPNs, or the FC-NVMe virtual WWPNS, depending on which type of host attachment the host is using. For an alternative to this approach, see “Multi-inititiator zoning” on page 23.
- One zone per storage system containing the aliases for the storage system and the IBM FlashSystem physical WWPNs if the IBM FlashSystem is virtualizing storage.

Tip: If you have enough IBM FlashSystem ports available and you have many hosts that you are connecting to an IBM FlashSystem, you should use a scheme to balance the hosts across the ports on the IBM FlashSystem. You can use a simple round-robin scheme, or you can use another scheme, such as numbering the hosts with the even-numbered hosts zoned to the even-numbered ports and the odd-numbered hosts zoned to the odd-numbered ports. Whichever load-balancing scheme that you choose to use, you should ensure that the maximum number of paths from each host to each volume is four paths. The maximum supported number is eight paths. The recommended number is four paths per volume.

Note: Do not add NVMe and SCSI ports to the same zone.

For SAN zoning best practices, see the IBM San Zoning Best Practices at Support page.

Multi-inititiator zoning

For host clusters such as VMware, it is desirable to have all hosts in the cluster in the same zone because it makes administration and troubleshooting easier. This setup can cause issues where a malfunctioning host affects all other hosts in the zone. Traditional best-practice zoning is to have only one initiator (host) per zone.

In recent years, Brocade released the Peer Zoning feature. Cisco released a similar feature that is called Smart Zoning. Both features allow multiple initiators to be in the same zone, but prevent them from connecting to each other. They can connect only to target ports in the zone, which allows multiple hosts to be in the same zone, but prevents the issue of a malfunctioning host port from affecting the other ports.

For VMware clusters, the preferred zoning configuration is to have the ports for all of the hosts in the cluster in a zone with the IBM FlashSystem virtual WWPN.

- Brocade Peer zoning must be enabled for this zone on Brocade fabrics. Brocade Peer Zoning was introduced in FOS v7.4.x.

  For more information about Brocade Peer zoning, see *Brocade Fabric OS Administration Guide 9.0.x*.

- Cisco Smart Zoning must be enabled for this zone on Cisco fabrics. Cisco Smart Zoning was introduced in NX-OS v5.2.x.

  For more information about Cisco Smart Zoning, see “Configuring and Managing Zones” in *Cisco MDS Family 9000 NX-OS Fabric Configuration Guide*.

2.5 Recommendations for tuning ESXi hosts

This section describes tuning ESXi hosts for better storage performance.

Marking the hosts as flash

In some cases, VMware sees volumes as a hard disk drives (HDD) even if they are not serial-attached SCSI (SAS) or near-line SAS (NL-SAS) volumes. Check the current drive type on ESXi and change it to **Flash** before creating the data store.

Setting the Round-Robin IOPS limit to 1

The recommended option for configuring Round-Robin and the correct IOPS limit is to create a rule that sets any new device that is added to that host automatically as a Round-Robin PSP with an I/O Operation Limit value of 1.

VMware Paravirtual SCSI

Paravirtual SCSI (PVSCSI) adapters are high-performance storage adapters that can result in greater throughput and lower CPU utilization. PVSCSI adapters are best for SAN environments, where hardware or applications drive a high amount of I/O throughput. The VMware PVSCSI adapter driver is also compatible with the Windows Storport storage driver. PVSCSI adapters are not suitable for direct-attached storage environments.

Large-scale workloads with intensive I/O patterns require adapter queue depths greater than the PVSCSI default values. At the time of writing, the PVSCSI queue depth default values are 64 (for device) and 254 (for adapter). You can increase PVSCSI queue depths to 254 (for device) and 1024 (for adapter) inside a Windows or Linux VM.

Eager-zeroed thick virtual disks

An eager-zeroed thick disk has all space allocated and zeroed out at the time of creation, which increases the time that it takes to create the disk, but results in the best performance, even on the first write to each block.

Latency sensitive threshold on NVMe volumes

When you use the HPP for your storage devices, set the latency sensitive threshold for the device so that I/O can avoid the I/O scheduler. By default, ESXi passes every I/O through the I/O scheduler. However, using the scheduler might create internal queuing, which is not efficient with the high-speed storage devices. You can configure the latency sensitive threshold and enable the direct submission mechanism that helps I/O to bypass the scheduler. With this mechanism enabled, the I/O passes directly from PSA through the HPP to the device driver. If the I/O latency exceeds the latency threshold, the system stops the direct submission, and temporarily reverts to using the I/O scheduler. The direct submission is resumed when the average I/O latency drops below the latency threshold.

3

Storage consumption

This chapter describes storage-related configurations in VMware vSphere for IBM FlashSystem storage systems.

This chapter includes the following sections:

3.1 Data store types

Data stores are logical containers that provide a uniform model for storing virtual machine (VM) files, ISO images, and VM templates. The following sections describe the types of data stores that can be deployed with IBM FlashSystem.

3.1.1 vSphere Virtual Machine File System

One type of data store that you can deploy on IBM FlashSystem storage systems uses the native vSphere Virtual Machine File System (VMFS) format. It is a special high-performance file system format that is optimized for storing VMs and allows shared access to multiple Elastic Sky X integrated (ESXi) hosts to concurrently read and write to the same storage. A VMFS data store can span multiple volumes (logical unit numbers (LUNs)), but this setup is not advisable. Instead, we recommend implementation in a one-to-one type of relationship. Several versions of the VMFS file system have been released since its introduction. Currently, ESXi supports VMFS5 and VMFS6.

When you work with VMFS data stores, consider the following items:

- Data store extents: Do not span more than one extent in a data store. The recommendation is to have a 1:1 ratio between the data store and the volume.
- Block size: The block size on a VMFS data store defines the maximum file size and the amount of space a file occupies. VMFS5 and VMFS6 data stores support the block size of 1 MB.
- Storage vMotion: Storage vMotion supports migration across VMFS, virtual storage area network (VSAN), and VMware vSphere Virtual Volume (vVol) data stores. A vCenter Server performs compatibility checks to validate Storage vMotion across different types of data stores.
- Storage Distributed Resource Scheduler (SDRS): VMFS5 and VMFS6 can coexist in the same data store cluster. However, all data stores in the cluster must use homogeneous storage devices. Do not mix devices of different formats within the same data store cluster.
- Device Partition Formats: A new VMFS5 or VMFS6 data store uses a globally unique identifier (GUID) partition table (GPT) to format the storage device. The GPT format enables you to create data stores larger than 2 TB. If your VMFS5 data store was previously upgraded from VMFS3, it continues to use the master boot record (MBR) partition format, which is characteristic for VMFS3. Conversion to GPT happens only after you expand the data store to a size larger than 2 TB.

Storage I/O Control

Storage I/O Control (SIOC) is a feature that provides I/O prioritization for virtual machine disks (VMDKs) that are on a shared data store. When a latency threshold is crossed for a shared data store, Storage I/O Control engages and starts prioritizing access to that data store. By default, all VMs have the same number of shares and a fair access to the data store. Therefore, SIOC prevents the "*noisy neighbor*" issue from occurring and makes sure that no one VM monopolizes access to that data store.

Starting from vSphere 6 VMware introduced IO reservations with SIOC. When reservations are used, the same I/O injector that is used for checking latency also samples the input/output operations per second (IOPS) capabilities of a data store. When the configured IOPS reservation that is set on the VMs exceeds the capabilities of the observed IOPS capabilities of that data store, IOPS is distributed to the VMs proportionally to their percentage of the number of set reservations.

The lowest value that you can set is 5 milliseconds (default is 30 ms). Typically, you cannot reach this value with IBM FlashSystem because it runs IOPS in microseconds. However, if the specified latency is reached, SIOC acts to reduce latency to acceptable levels.

For critical systems, the usual recommendation is to not employ limits or throttling on the VMs resources. Even though SIOC falls into the throttling category, it also provides a great fail-safe for unavoidable and unpredictable contention. This function might be helpful when there are multiple VMDKs that share a data store for manageability reasons.

For more information about SIOC on data stores used by production databases, see Storage I/O control for critical apps is a great idea.

Note: The goal of Distributed Resource Scheduler (DRS) I/O load-balancing is to fix long-term prolonged I/O imbalances, VMware vSphere SIOC addresses short-term burst and loads.

SIOC limitations and requirements

The following requirements and limitations apply to SIOC:

- Data stores that are SIOC-enabled must be managed by a single vCenter Server system.
- SIOC is supported on VMFS data stores only.
- SIOC does not support data stores with multiple extents.

Setting the Storage I/O Control threshold value

The congestion threshold value for a data store is the upper limit of latency that is allowed for a data store before SIOC assigns importance to the VM workloads according to their shares.

You do not need to adjust the threshold setting in most environments. If you change the congestion threshold setting, set the value based on the following considerations:

- A higher value typically results in higher aggregate throughput and weaker isolation. Throttling does not occur unless the overall average latency is higher than the threshold.
- If throughput is more critical than latency, do not set the value too low. For example, for Fibre Channel disks, a value below 20 ms might lower peak disk throughput. A high value (above 50 ms) might allow high latency without significant gain in overall throughput.
- A lower value results in lower device latency and stronger VM I/O performance isolation. Stronger isolation means that the shares controls are enforced more often. Lower device latency translates into lower I/O latency for the VMs with the highest shares, at the cost of higher I/O latency experienced by the VMs with fewer shares.
- A low value (lower than 20 ms) results in lower device latency and isolation among I/Os at the potential cost of a decrease in aggregat4e data store throughput.
- Setting the value high or low results in poor isolation.

Data store clusters (Easy Tier versus SDRS)

A data store cluster is a collection of data stores with shared resources and a shared management interface. You can use vSphere SDRS to manage storage resources when you create a data store cluster.

SDRS consists of features that can be used to balance storage space and load between data stores by using Storage vMotion to migrate VMs. Depending on your environment, you can automate these tasks or decide to be notified and implement actions yourself.

SDRS I/O balancing

SDRS is load-balancing based on I/O latency and I/O metrics. I/O metrics build upon SIOC, which continuously monitors the I/O latency (which is the time it takes for I/O to do a round trip). SDRS captures this performance data over a period. The initial period is 16 hours and after that, it is based on the advanced setting of checking for imbalance every defined period (default is 8 hours).

If the latency for a data store exceeds the threshold (default: 15 ms) over a percentage of time, SDRS migrates VMs to other data stores within the data store cluster until the latency is below the threshold limit. SDRS might migrate a single VM or multiple VMs to reduce the latency for each data store below the threshold limit. If SDRS is unsuccessful in reducing latency for a data store, it (at a minimum) tries to balance the latency among all data stores within a data store cluster.

When I/O metrics are enabled, SIOC is enabled on all data stores in the data store cluster.

Note: The I/O latency threshold for SDRS should be lower than or equal to the SIOC congestion threshold.

In an IBM FlashSystem Easy Tier environment, disable I/O-related functions of SDRS so that Easy Tier can work properly.

SDRS space balancing

SDRS continuously monitors the space usage of all data stores in a data store cluster. If the amount of used space in a data store falls below the defined threshold (default 80%), SDRS migrates one or more VMs to other data stores within the data store cluster to reduce the amount of used space below the defined threshold. However, the advanced setting of utilization difference between the source and destination data store (default 5%) can be used. If the difference between the source and destination is less than the threshold, SDRS does not migrate the VMs. SDRS does not migrate the VM if there is not enough space in the destination data store.

Powered-on VMs with snapshots are not considered for space balancing.

Tip: As a general recommendation, consider using a data store cluster or SDRS whenever possible. However, make sure to disable latency-based rules in Easy Tier environment. SDRS simplifies VM placement when creating, deploying, or cloning VMs. SRDS provides recommendations for balancing on space and I/O. In manual mode, recommendations can be applied on a case-by-case basis.

For answers to frequently asked questions about SDRS, see SDRS FAQ (2149938).

Capacity sizing and number of volumes

Table 3-1 summarizes the storage maximums for ESXi host at the time of writing.

Table 3-1 ESXi host maximums: storage

Category	Description	Limit
Virtual Disks	Virtual Disks per Host	2048
Fibre Channel	LUNs per host	1024
Fibre Channel	LUN size	64 TB
Fibre Channel	LUN ID	0 - 16383

Category	Description	Limit
Fibre Channel	Number of paths to a LUN	32
Fibre Channel	Number of total paths on a server	4096
Fibre Channel	Number of host bus adapters (HBAs) of any type	8
Fibre Channel	HBA ports	16
Fibre Channel	Targets per HBA	256
iSCSI Physical	LUNs per server	1024
iSCSI Physical	10 Gb Internet Small Computer System Interface (iSCSI) HBA initiator ports per server	4
iSCSI Physical	Network interface cards (NICs) that can be associated or port-bound with the software iSCSI stack per server	8
iSCSI Physical	Number of total paths on a server	4096
iSCSI Physical	Number of paths to a LUN (software iSCSI and hardware iSCSI)	32
iSCSI Physical	10 Gb iSCSI HBA targets per adapter port	128
iSCSI Physical	Software iSCSI targets	256
Common VMFS	Volume size	64 TB
Common VMFS	Volumes per host	1024
Common VMFS	Hosts per volume	64
Common VMFS	Powered on VMs per VMFS volume	2048
Common VMFS	Concurrent vMotion operations per VMFS volume	128
VMFS5 / VMFS-6	Raw Device Mapping size (virtual compatibility)	62 TB
VMFS5 / VMFS-6	Raw Device Mapping size (physical compatibility)	64 TB
VMFS5 / VMFS-6	Block size	1 MB
VMFS5 / VMFS-6	File size	62 TB
VMFS5 / VMFS-6	Files per volume	~130690
RDMA NVMe	Namespaces per server	32
RDMA NVMe	Number of paths to a namespace	4
RDMA NVMe	Number of total paths on a server	128
RDMA NVMe	Initiator ports per server	2

3.1.2 Raw Device Mappings

Raw Device Mapping (RDM) is used with VMware ESXi hosts to provide a VM with access to an entire LUN. RDM can be seen as a mapping file in a separate Virtual Machine File System (VMFS) volume, which acts as a proxy to a raw physical storage device (which is a LUN). The RDM contains metadata for managing and redirecting disk access to the physical device.

With RDM, a VM can access and use a storage LUN directly and it allows the use of VMFS manageability.

Figure 3-1 shows an illustration of the RDM. RDM is a symbolic link from a VMDK file on a VMFS to a raw LUN.

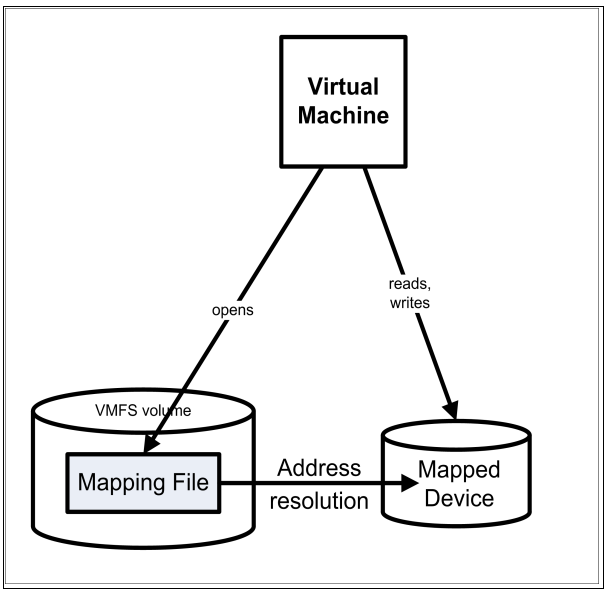


Figure 3-1 Illustrating Raw Device Mapping

An RDM offers a number of benefits, but it should not be used in every situation. In general, virtual-disk files are preferred over RDMs for manageability purpose.

The most common use case for RDM is Microsoft Cluster Server (MSCS). In an MSCS clustering scenario that spans multiple hosts, which can be a mix of virtual and physical clusters, the cluster data and quorum disk are to be configured as RDMs.

Use of storage area network management agents within a virtual machine

The two RDM modes are as follows:

- Virtual: With virtual mode, the RDM appears to be the same as a virtual disk in a VMFS and the VMKernel sends its reads and writes to the mapping file instead of accessing the physical device directly.
- Physical: Physical mode provides more control over the physical LUN to access it directly, but it also has a downside; VMware snapshots are not supported. Additionally, in physical mode, you cannot convert the RDM disk to a VMFS virtual disk by using storage vMotion.

3.1.3 VMware vSphere Virtual Volume

Before vVols, a VMDK was presented to a VM in the form of a file. This file represents a disk to the VM, which is then accessed by the guest operating system in the same way as a physical disk is accessed on a physical server. This VMDK is stored on a VMware file system (VMFS) formatted data store.

The VMFS data store is hosted by a single volume on a storage system, such as the IBM FlashSystem 9200. A single VMFS data store can have hundreds or even thousands of VMDKs.

vVols provide a one-to-one mapping between the VM's disks and the volumes that are hosted by the storage system. These vVols are wholly owned by the VM. Making the vVols available at the storage level enables storage system-based operations at the granular VM level. For example, capabilities such as compression and encryption can be applied to an individual VM. Similarly, IBM FlashCopy® can be used at the vVol level when you perform snapshot and clone operations.

The integration of vVols with IBM Spectrum Virtualize storage systems is dependent upon the vSphere application programming interfaces (APIs) for Storage Awareness (VASA). These APIs facilitate VM-related tasks that are initiated at the vSphere level to be communicated down to the storage system.

IBM support for VASA is provided by IBM Spectrum Connect. IBM Spectrum Connect is an out-of-band VASA Provider, which enables the communication between vSphere and the storage system along the control plane.

IBM FlashSystem manages vVols at the storage level and enables the flexible and dynamic provisioning of VM storage that is required of a truly software-defined storage environment.

For information about how to implement vVols with IBM FlashSystem, see Chapter 4, "Integrating with VMware by using IBM Spectrum Connect" on page 41.

Storage system considerations

Important factors must be considered when planning an implementation of vVols. The goal is to highlight the decisions that need to be made to maximize the potential of vVols and meet the needs of your environment.

Because the storage for vVols is allocated as a child pool, it is important to consider the structure of the parent pools from which these child pools are allocated. The following sections describe the two contrasting approaches to defining parent pools and a description of how their usage might influence the vVols environment.

Drive class based

You might want to define parent pools based on the underlying drive class. This approach enables the allocation of child pools from a specific tier of storage. At the vSphere level, these child pools serve as the backing-storage container for distinct vVols data stores. This approach enables you to determine the class of storage for individual vVols when provisioning VMs.

We can use vSphere policies (for example, Gold, Silver, and Bronze) to select the appropriate class of storage when we provision VMs.

Figure 3-2 shows an example of this configuration.

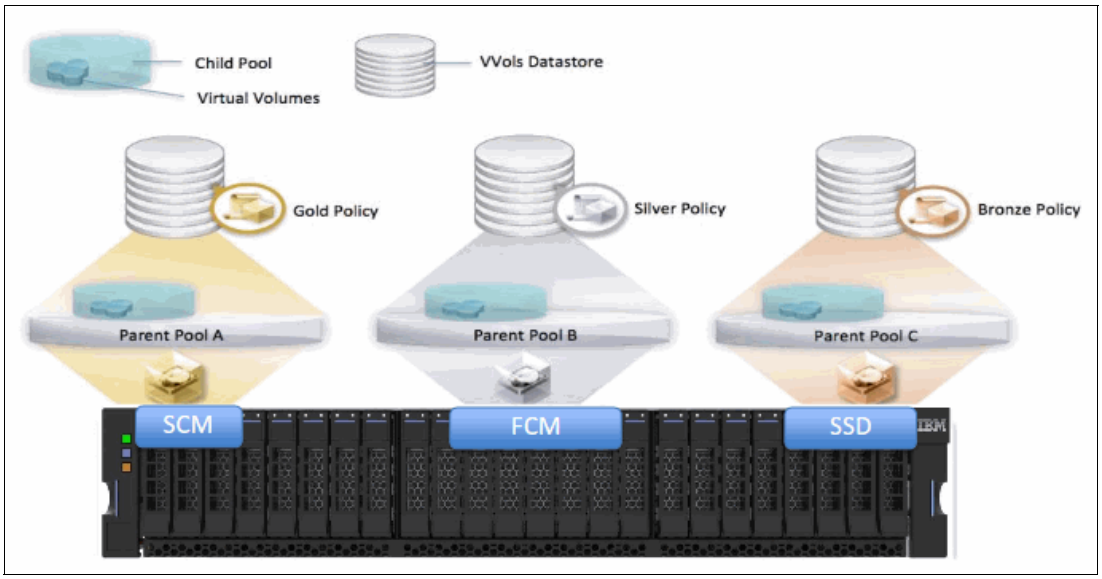


Figure 3-2 A vVols environment where parent and child pools are segregated by drive class

With the introduction of vVols, by defining a range of storage services on IBM Spectrum Connect, policies can become far more interesting and useful than the simple Gold, Silver, and Bronze model.

Each of the policies (gold, silver, and bronze) can be further subdivided. For example, we might divide our solid-state drive (SSD) parent pool into two distinct child pools. One child pool is linked to an encrypted storage service, and the other is associated with an unencrypted storage service. This approach provides the vSphere administrators with the flexibility to provision VMs on storage that matches the requirements of the application, on a per-VM basis.

IBM Easy Tier based

An alternative to the drive-class based parent pools would be to define parent pools with a combination of drive classes, and enable the IBM Easy Tier feature. By monitoring the heatmap of a volume's extents, Easy Tier can intelligently optimize the use of storage by automatically migrating these extents onto the most appropriate storage tier.

Because vVols are a special volume, Easy Tier can manage their extents in an identical fashion.

- A *hot* (frequently used) extent of a vVol is promoted to faster storage, such as SSD.
- A *cold* (infrequently used) extent of a vVol is moved onto slower drives.

A vVols implementation that takes advantage of Easy Tier can provide greater simplicity for the storage administrator. By defining a child pool within an Easy Tier enabled parent pool, the storage system is enabled to flexibly manage the extents of any vVols created therein.

This flexibility removes the requirement for a choice of storage class when the vSphere administrator initially provisions the VM. Such an approach can also minimize the need for Storage vMotion tasks because Easy Tier eliminates the requirement to manually migrate vVols onto faster or slower storage as the needs of an application change.

Figure 3-3 on page 33 demonstrates a vVols configuration, based on a single parent pool, with Easy Tier enabled.

Note: Easy Tier also provides benefits within a single-tiered pool. When enabled, Easy Tier automatically balances the load between managed disks (MDisks) to optimize performance.

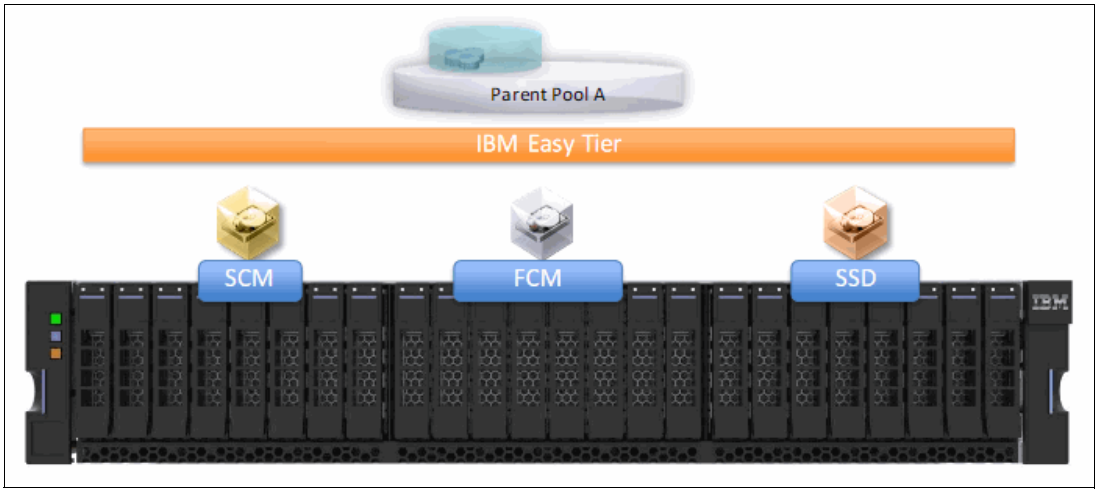


Figure 3-3 The simplified approach to vVols provisioning that can be implemented by enabling Easy Tier

3.2 VMware vSphere Storage APIs – Array Integration

vSphere Storage APIs – Array Integration (VAAI), also known as *hardware acceleration*, is an API that allows the Elastic Sky X integrated (ESXi) host to offload resource-intensive I/O operations to the storage array, for example, by copying the VM files.

- Without VAAI, the VM files are copied by using the host.
- With VAAI, the data is copied within the same storage array.

VAAI helps the ESXi performance because the storage area network (SAN) fabric is not used and fewer central processing unit (CPU) cycles are needed because the copy does not need to be handled by the host.

In vSphere 5.x and later releases, these extensions (VAAI operations) are implemented as T10 Small Computer System Interface (SCSI) commands. As a result, with the devices that support the T10 SCSI standard (such as the IBM FlashSystem) your ESXi host can communicate directly and does not require the VAAI plug-ins.

The following types of operations are supported by the VAAI hardware acceleration for IBM FlashSystem:

- Atomic Test and Set (ATS), which is used during creation and locking of files on the VMFS volume
- Clone Blocks, Full Copy, and extended copy (XCOPY), which is used to copy or migrate data within the same physical array
- Zero Blocks/Write Same, which is used when creating VMDKs with an eager-zeroed thick provisioning profile
- SCSI **UNMAP**, which is used to reclaim storage space

3.2.1 Atomic Test and Set / SCSI Compare and Write

ATS, also known as *hardware-assisted locking*, intelligently relegates resource-access serialization down to the granularity of the block-level during VMware metadata updates. ATS uses this approach rather than using a mature SCSI2 reservation, which serializes access to the adjacent ESXi hosts with a minimum scope of an entire LUN.

ATS is a standard T10 SCSI command with opcode *0x89* (SCSI Compare and Write (CAW)). The **ATS** primitive has the following advantages where LUNs are used by multiple applications or processes at one time:

- Significantly reduces SCSI reservation contentions by locking a range of blocks within an LUN rather than issuing a SCSI reservation on the entire LUN.
- Enables parallel storage processing.
- Reduces latency for multiple ESXi hosts accessing the same LUN during common.
- Increases cluster scalability by greatly extending the number of ESXi hosts and VMs that can viably reside simultaneously on a VMFS data store.

Our recommendation is to enable ATS hardware-accelerated locking and set to **ATS-only public**. For more information about how to perform this task, see Configuring the ESXi operating system.

Note: All newly formatted VMFS5 and VMFS6 data stores use the ATS-only mechanism if the underlying storage supports it. SCSI reservations are never used.

VMFS3 volumes that are upgraded to VMFS5 must be manually upgraded to ATS-only so that it is easier to redeploy the data store and migrate the VMs to the data store.

VMware ATS heartbeating

VMware ESXi uses the SCSI **Compare and Write** command (VMware refers to this command as **ATS**) to heartbeat periodically to data stores.

Note: The use of ATS heartbeating is not supported on the following platforms:

- ESXi hosts that run version 5.5 update 2 or later
- ESXi version 6.0 before update 3

During high-latency events, ATS heartbeats might timeout, which results in ATS miscompare errors. If multiple heartbeat attempts fail, the ESXi host might lose access to the data store in which timeouts are observed.

ATS heartbeating increases the load on the system and can lead to access issues on busy systems, particularly during maintenance procedures. To reduce this load, ATS heartbeats can be disabled.

- For VMware vSphere versions 5.5 and 6.0, the recommendation is to disable ATS heartbeating because of host-disconnect issues.

  To disable ATS heartbeats, run the following command:

  ```
  # esxcli system settings advanced set -i 0 -o /VMFS3/UseATSForHBOnVMFS5
  ```

- For VMware vSphere versions 6.5, 6.7 and 7.0, the recommendation is to enable ATS heartbeating.

To enable ATS heartbeats, run the following command-line interface (CLI) command:

```
# esxcli system settings advanced set -i 1 -o /VMFS3/UseATSForHBOnVMFS5
```

3.2.2 Extended copy

XCOPY, also known as a hardware-accelerated move, offloads copy operations from VMware ESXi to the IBM FlashSystem. This process allows for rapid movement of data when performing copy or move operations within the IBM storage system. XCOPY reduces CPU cycles and host bus adapter (HBA) workload of the ESXi host.

Similarly, XCOPY reduces the volume of traffic moving through the SAN when a VM is deployed. It does so by synchronizing individual VM-level or file system operations (including clone and migration activities) with the physical storage-level operations at the granularity of individual blocks on the devices. The potential scope in the context of the storage is both within and across LUNs.

The **XCOPY** command has the following benefits:

- Expedites copy operations including the following tasks:
  - Cloning of VMs, including deploying from template
  - Migrating VMs from one data store to another (storage vMotion)
- Minimizes host processing and resource allocation: Copies data from one LUN to another without reading and writing through the ESXi host and network.
- Reduces SAN traffic.

The SCSI opcode for XCOPY is **0x83**. As a best practice, set the XCOPY transfer size to 4096, as shown in Example 3-1.

Example 3-1 XCOPY transfer size 4096

```
# Get-VMHost  | Get-AdvancedSetting -Name DataMover.MaxHWTransferSize | select Entity, name, value
Entity                         Name                         Value
------                         ----                         -----
vmlab11c2.ssd.hursley.ibm.com DataMover.MaxHWTransferSize  4096
```

3.2.3 WRITE_SAME

Block Zeroing, Write_Same (Zero), or *hardware-accelerated initialization* use the **WRITE_SAME 0x93** SCSI command to issue a chain of identical write transactions to the storage system. This command almost entirely eliminates server processor and memory use by eliminating the need for the host to run repetitive identical write transactions. It also reduces the volume of host HBA and SAN traffic when repetitive block-level write operations are performed within VM disks to the IBM FlashSystem.

The **WRITE_SAME 0x93** SCSI command allows the IBM FlashSystem to minimize internal bandwidth consumption. For example, when provisioning a VMDK file with the **eagerzeroedthick** specification, the Zero Block's primitive issues a single **WRITE_SAME** command that replicates zeros across the capacity range that is represented by the difference between the provisioned capacity of the VMDK and the capacity that is consumed by actual data. The alternative to using the **WRITE_SAME** command requires the ESXi host to issue individual writes to fill the VMDK file with zeros. The same applies when cloning or performing storage vMotion of a VM with eager-zeroed thick VMDKs.

The scope of the Zero Block's primitive is the VMDK creation within a VMFS data store. Therefore, the scope of the primitive is generally within a single LUN on the storage subsystem, but it can potentially span LUNs backing multi-extent data stores.

Block Zeroing offers the following benefits:

- Offloads initial formatting of Eager Zero Thick (EZT) VMDKs to the storage array
- Assigns zeros to large areas of storage without writing zeros from the ESXi host
- Speeds up creation of new VMs
- Reduces elapsed time, server workload, and network workload

Note: In thin-provisioned volumes, IBM FlashSystem further augments this benefit by flagging the capacity as "zeroed" in metadata without the requirement to physically write zeros to the cache and the disk, which implies even faster provisioning of the eager-zeroed VMDKs.

3.2.4 SCSI UNMAP command

IBM FlashSystem that is built with IBM Spectrum Virtualize software supports the SCSI **UNMAP** command since Version 8.1.0, which enables hosts to notify the storage controller of capacity that is no longer required, which might improve savings. Reclaiming storage space can provide higher host-to-flash I/O throughput and improve flash endurance.

When an IBM FlashSystem receives a SCSI **UNMAP** command, it overwrites the relevant region of the volume with all-zero data, which allows thin-provisioned storage controllers (such as the IBM FlashSystem) to reclaim physical capacity through garbage collection.

The main benefit is that this action helps prevent a thin-provisioning storage controller from running out of free capacity for write I/O requests, which means that when thin-provisioned storage controllers are used, SCSI Unmap should normally be left enabled.

With lower-performing storage (such as nearline arrays), extra I/O workload can be generated, which can increase response times.

To enable SCSI **UNMAP**, run the following command on IBM FlashSystem:

```
chsystem -hostunmap on
```

Enabling SCSI **UNMAP** does not affect data on older volumes that are created before using this command. You must create data stores on newly created volumes, migrate data through storage vMotion, and delete old volumes.

SCSI UNMAP effects on standard storage pool and on data reduction pool

Hosting **UNMAP** commands in a standard storage pool results in data being zeroed, but does not increase the free capacity that is reported by the storage pool. When an array is composed by IBM FlashCore Module (FCM) modules, the **UNMAP** command will increase the free physical capacity on that array.

Host **UNMAP** commands can increase the free capacity that is reported by the data reduction pool (DRP) when received by thin-provisioned or compressed volumes. SCSI **UNMAP** commands also are sent to internal FlashCore Modules (FCMs) to free physical capacity.

For more information, see SCSI Unmap support in IBM Spectrum Virtualize systems.

For more information about DRPs, see *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430.

Storage space reclamation

ESXi supports the SCSI **UNMAP** command (also known as the space reclamation command) that originates from a VMFS data store or a VM guest operating system.

Inside the VM, storage space is freed when you delete files on the thin virtual disk. Storage space that is left by deleting or removing files from a VMFS data store can be freed up within the file system. This free space is allocated to a storage device until the file system releases or unmaps it. This operation helps the storage array to reclaim unused free space.

Space reclamation requests from VMFS data stores

VMFS5 and earlier file systems do not unmap free space automatically, but you can use the **esxcli storage vmfs unmap** command to reclaim space manually. When you use the command, be aware that it might send many unmap requests simultaneously. This action can lock some of the resources during the operation.

On VMFS6 data stores, ESXi supports the automatic asynchronous reclamation of free space. VMFS6 can run the **UNMAP** command to release free storage space in the background on thin-provisioned storage arrays that support unmap operations. Asynchronous unmap processing has several advantages:

- Unmap requests are sent at a rate (that can be throttled in vSphere), which helps to avoid any instant load on the backing array.
- Freed regions are batched and unmapped together.
- I/O performance of other workloads is not impacted by the **UNMAP** command.

For information about the space-reclamation parameters for VMFS6 data stores, see Space Reclamation Requests from VMFS Datastores.

Space reclamation requests from guest operating systems

ESXi supports the **UNMAP** commands that are issued directly from a guest operating system to reclaim storage space. The level of support and requirements depend on the type of data store in which your VM resides.

The guest operating system notifies VMFS about freed space by sending the **UNMAP** command. The **UNMAP** command that is sent from the guest operating system releases space within the VMFS data store. The command proceeds to the array so that the array can reclaim the freed blocks of space.

VMs on top of VMFS5 typically cannot pass the **UNMAP** command directly to the array; you must run the **esxcli storage vmfs unmap** command to trigger unmaps from the IBM FlashSystem. However, for a limited number of the guest operating systems, VMFS5 supports the automatic space reclamation requests.

To send the unmap requests from the guest operating system to the array, the VM must meet the following prerequisites:

- The virtual disk must be thin-provisioned.
- VM hardware must be version 11 (ESXi 6.0) or later.
- The advanced **EnableBlockDelete** setting must be set to 1.
- The guest operating system must be able to identify the virtual disk as thin.

VMFS6 generally supports automatic space-reclamation requests that generate from the guest operating systems, and passes these requests to the array. Many guest operating systems can send the **UNMAP** command and do not require any additional configuration. The guest operating systems that do not support automatic unmaps might require user intervention.

The following considerations apply when you use space reclamation with VMFS6:

- VMFS6 processes the unmap request from the guest operating system (OS) only when the space to reclaim equals 1 MB or is a multiple of 1 MB. If the space is less than 1 MB or is not aligned to 1 MB, the unmap requests are not processed.
- For VMs with snapshots in the default SEsparse format, VMFS6 supports the automatic space reclamation only on ESXi hosts version 6.7 or later.

Space reclamation affects only the top snapshot and works when the VM is powered on.

IBM FlashSystem and VAAI

To verify which VAAI operations are supported by your storage device, issue a command as shown in Example 3-2.

Example 3-2 Verifying device VAAI support where "naa.xxx" stands for device identifier

```
[root@ESX1-ITSO:~] esxcli storage core device vaai status get -d naa.xxx
naa.xxx
   VAAI Plugin Name:
   ATS Status: supported
   Clone Status: supported
   Zero Status: supported
   Delete Status: unsupported
```

You can verify and change your VAAI settings in host Advanced System Settings (Figure 3-4). A value of 1 means that the feature is enabled. If the setting is host-wide, it is enabled if the connected storage supports it.

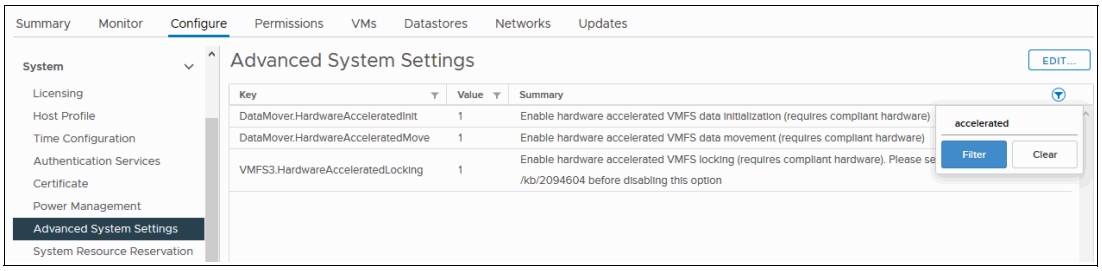


Figure 3-4 VAAI settings

Table 3-2 lists the VAAI settings and parameter descriptions.

Table 3-2 Parameters description

Parameter name	Description
DataMover.HardwareAcceleratedInit	Zero Blocks/Write Same
DataMover.HardwareAcceleratedMove	Clone Blocks/XCOPY
VMFS3.HardwareAcceleratedLocking	Atomic Test and Set (ATS)

It is advisable to keep all VAAI operations enabled when using IBM FlashSystem storage systems so that as much work as possible is offloaded to storage.

Example 3-3 shows how to evaluate the VAAI status by using the PowerCLI command.

Example 3-3 PowerCLI command that is used to evaluate the VAAI status

```
# Get-VMHost  | Get-AdvancedSetting -name *HardwareAccelerated* | select Name, value
Name                              Value
----                              -----
DataMover.HardwareAcceleratedMove     1
VMFS3.HardwareAcceleratedLocking      1
DataMover.HardwareAcceleratedInit     1
```

Hardware offloading is not used if following conditions occur:

- The source and destination VMFS data stores have different block sizes. (This situation can happen when using existing VMFS3 data stores.)
- The source disk is RDM and the destination is non-RDM (regular VMDK).
- The source VMDK type is `eagerzeroedthick` but the destination VMDK type is *thin*.
- The source or destination VMDK is in a sparse or hosted format.
- The source VM has a snapshot.
- The logical address and transfer length in the requested operation are not aligned to the minimum alignment required by the storage device. (All data stores created with the vSphere Web Client are aligned automatically.)
- The VMFS has multiple LUNs or extents, and they are on different arrays.
- Hardware cloning between arrays, even within the same VMFS data store, does not work. (This situation is not the case if arrays are managed by IBM FlashSystem by using external virtualization.)

Note: You might decide to increase a block size (`MaxHWTransferSize` for XCOPY), which is processed by storage globally. Although we do not recommend changing default values, you might notice a small improvement in the performance during Data Mover operations (typically around 10%). This change is global and affects all your VAAI-enabled storage devices that are connected to the ESXi. Therefore, changing the default values can have unpredictable impact on different storage arrays.

4

Integrating with VMware by using IBM Spectrum Connect

This chapter describes the configuration steps to integrate VMware and IBM Spectrum Connect, best-practice considerations, and troubleshooting tips.

This chapter includes the following sections:

4.1 Overview of IBM Spectrum Connect

IBM Spectrum Connect is a dedicated Linux-based application that provides centralized management of IBM storage platforms for multiple virtualization, cloud, and container interfaces. By implementing advanced storage-provisioning techniques, IBM Spectrum Connect allows storage and hybrid-cloud administrators to supply and integrate IBM storage with a range of VMware solutions, Kubernetes container clusters, and Microsoft PowerShell automation methods.

IBM Spectrum Connect provides a web-based user interface (UI) that makes this entire administration easy and straightforward. Use of the UI saves a significant amount of time in setting up, connecting, and integrating the required storage resources into your cloud environment.

Through its user credential, storage system, storage space, and service management options, IBM Spectrum Connect facilitates the integration of IBM storage system resources with the supported virtualization, cloud, and container platforms.

The following storage services, which can be considered as storage profiles, are defined in IBM Spectrum Connect and delegated for use in VMware for simplified profile-based volume provisioning:

- VMware vSphere Web Client (vWC)
- VMware vSphere Storage APIs - Storage Awareness (VASA)
- VMware vRealize Operations (vROps) Manager
- VMware vRealize Automation, and VMware vRealize Orchestrator (vRO)
- Microsoft PowerShell

Note: The VASA for VMware vSphere Virtual Volumes (vVols)) function is also available through the Embedded VASA Provider in IBM Spectrum Virtualize 8.5.1.0 or later. For more information about using VASA or vVols with the Embedded VASA Provider, see Chapter 6, “Embedded VASA Provider for Virtual Volumes” on page 129.

4.1.1 Supported cloud interfaces

This book focuses on the following cloud interfaces that are compatible with VMware integrations that use IBM Spectrum Connect:

- IBM Storage Provider for VMware VASA
- IBM Storage Enhancements for VMware vSphere Web Client
- IBM Storage Plug-in for VMware vRO
- IBM Storage Management Pack for VMware vROps Manager

In addition, the IBM Spectrum Connect lifecycle and compatibility matrix on IBM Documentation details the IBM Spectrum Connect lifecycle with compatible storage-system microcodes and supported cloud interfaces.

4.1.2 Installation considerations

You can install the IBM Spectrum Connect software on a compatible version of Red Hat Enterprise Linux (RHEL) or CentOS. For more information about supported operating systems, see IBM Spectrum Connect 3.7.0.

As shown in Figure 4-1, the IBM Spectrum Connect application communicates with IBM storage systems by using command-line interface (CLI) commands over Secure Shell (SSH). VMware also issues application programming interface (API) calls directly to IBM Spectrum Connect over Transmission Control Protocol/Internet Protocol (TCP/IP). Therefore, IP connectivity must exist between the IBM Spectrum Connect server and the Management IP address (sometimes referred to as a *Cluster IP*) of the storage system. For security, some network infrastructures might be segmented into virtual local area networks (VLANs) or have isolation preventing the management of the storage system from being accessible from virtual machines (VMs) within a vSphere environment. Check with your network administrator to ensure that IP connectivity exists between the different components.

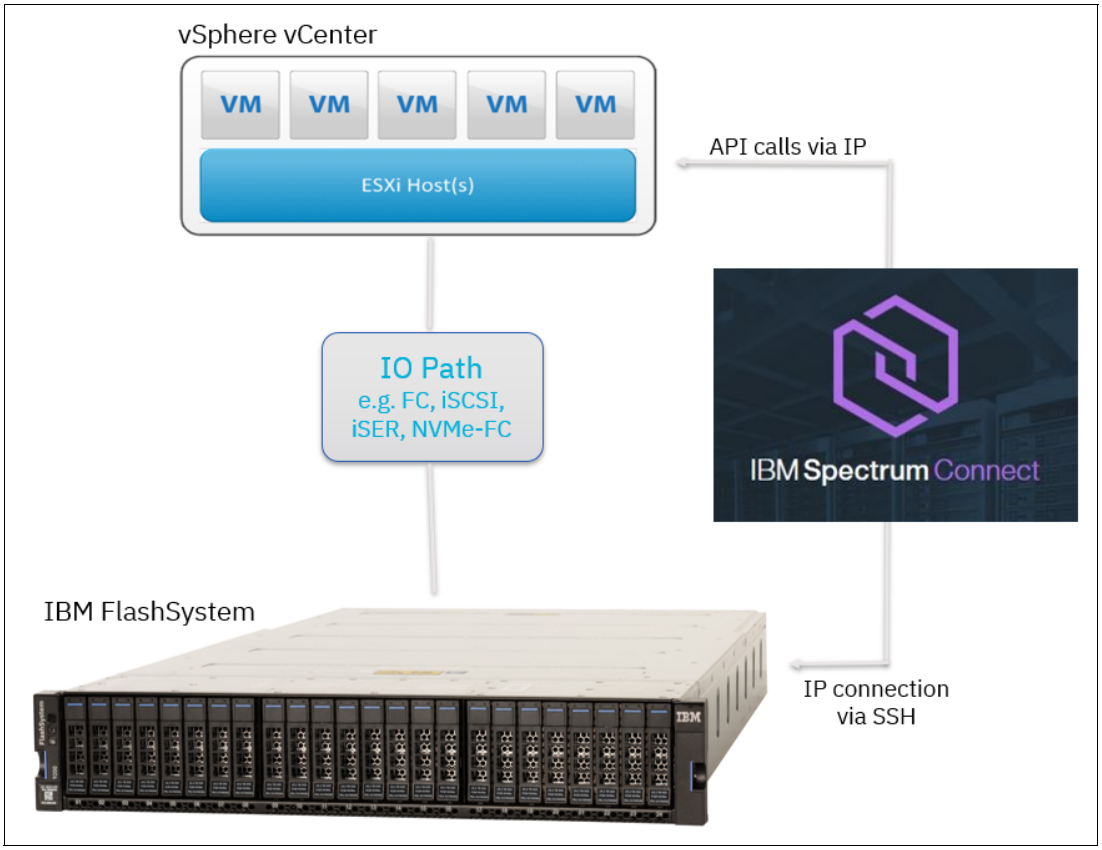


Figure 4-1 IBM FlashSystem and vSphere vCenter integration architecture

Communication between VMware vRO, vSphere, or vCenter and the IBM storage system by using IBM Spectrum Connect is “out-of-band” and is therefore separate from I/O traffic between a host (for example an Elastic Sky X integrated (ESXi) host) and the storage system.

If network connectivity issues occur, which prevent IBM Spectrum Connect from communicating with either the cloud interface or the storage system, the I/O workload that is running on the hosts is unaffected.

Note: When you use vVols, IBM Spectrum Connect can operate in a high availability (HA) model, where two IBM Spectrum Connect servers can be configured to run in Active/Standby. For more information about this feature, see *IBM Spectrum Connect Version 3.7.0 User Guide*.

VM resource requirements might depend on the roles that are performed by the IBM Spectrum Connect Server. The IBM Spectrum Connect server periodically (by default, it is configured to run in 10-minute intervals) queries the storage system for an inventory of objects such as VDisks, hosts, and FlashCopy mappings. This query populates a local cache of the configuration, which can be used for cache can be used by the following services:

- *vROps* for performance data
- The *IBM Storage Enhancements plug-in*, which maintains constant awareness of host objects that are defined on the storage system

Depending on the size and complexity of the IBM Spectrum Virtualize configuration, the population task might take some time.

Tip: In large environments, consider increasing the population interval to allow sufficient time for the task to complete. For more information about IBM Spectrum Connect, see IBM Documentation at Working with multiple storage systems.

4.1.3 Downloading and installing IBM Spectrum Connect

IBM Spectrum Connect is available to download from IBM Fix Central.

Note: Before you install the IBM Spectrum Connect application, it is a best practice to configure Network Time Protocol (NTP) client on the Linux operating system (OS). Given the multiple components in the end-to-end infrastructure any time-skew between IBM Spectrum Connect, IBM Spectrum Virtualize, and the VMware platforms can complicate debugging issues when reviewing logs.

For more information about installation instructions and minimum requirements, see *IBM Spectrum Connect Version 3.7.0 User Guide*.

Make note of the installation summary screen (Figure 4-2 on page 45), which provides details about additional steps to configure firewall rules and SELinux, if required.

```
SECURITY NOTES:
===============
The following ports must be opened on this host:
- Port 5672 for rabbitmq on the internal interface (lo).
- Port 4369 for ampq on the internal interface (lo).
- Port 8440 on the external interface.

If you are using the linux default firewall, you can use the following commands to open the
port:
firewall-cmd --permanent --zone=trusted --add-interface=lo
firewall-cmd --permanent --add-port=8440/tcp
firewall-cmd --permanent --zone=trusted --add-port=4369/tcp
firewall-cmd --permanent --zone=trusted --add-port=5672/tcp
firewall-cmd --reload
If you are using a different firewall software please refer to the software documentation for
help.

If SELinux is enabled on this machine, nginx must be allowed to bind network interfaces and
connect to ibmsc socket. This can be done using the following commands:
semodule -i /opt/ibm/ibm_spectrum_connect/conf.d/selinux/rhel7/ibmsc.pp
systemctl nginx restart
To display ibmsc selinux policy:
cat /opt/ibm/ibm_spectrum_connect/conf.d/selinux/rhel7/ibmsc.te

If the rabbitmq-server service is reported as not running it can be restarted by the
following command:
systemctl restart rabbitmq-server

IMPORTANT: To avoid unauthorized access to the IBM Spectrum Connect,
the password for this username should be changed as soon as possible.
You can control IBM Spectrum Connect services using the
'service ibm_spectrum_connect {start|stop|status}' command.

Installation completed successfully.
```

Figure 4-2 Installation summary screen

4.1.4 Initial configuration

When the IBM Spectrum Connect application is installed, you can access the management web interface (Figure 4-3), by connecting to `http://<IP or FQDN>:8440`. The default credentials are:

- Name: `admin`
- Password: `admin1!`

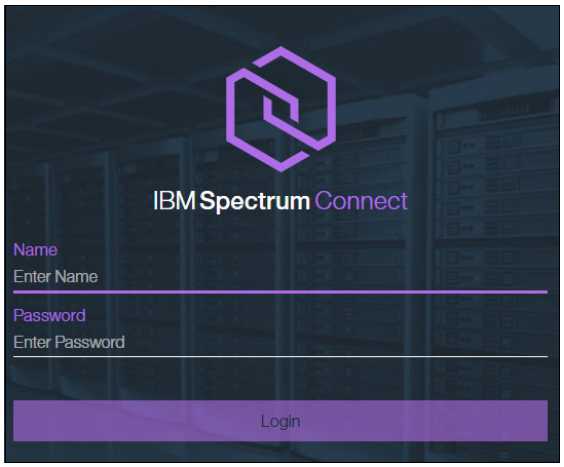


Figure 4-3 Management web interface

After successful login, complete the initial setup wizard, which includes the following mandatory configuration steps:

1. Set up an HA group and IBM Spectrum Connect server identity.
2. Provide details for the SSL certificate.
3. Define storage system credentials.
4. Change the default IBM Spectrum Connect credentials.

Notes:

- When specifying the storage-system credentials, it is a best practice to create a dedicated user account on the storage-system, which is easily identifiable as being from an IBM Spectrum Connect server. This account is the User account that is used when IBM Spectrum Connect issues CLI commands to the storage system. Having an easily recognizable username assists with tasks (for examples, reviewing audit logs within IBM Spectrum Virtualize) and make it clear that CLI commands were issued by IBM Spectrum Connect.
- The storage-system credentials are global and apply to all storage systems being registered in IBM Spectrum Connect. When possible, consider using Lightweight Directory Access Protocol (LDAP) authentication on the storage system to simplify user account management.

Ensure that the associated user account is created on every storage system that is to be registered and that the account was granted a suitable role.

For example, when using VMware vSphere Virtual Volumes, the storage-system user account must be assigned the “VASA Provider” role on each storage system. To do this task, you must first create a User Group with the “VASA Provider” role in IBM Spectrum Virtualize, as follows:

1. Open the IBM Spectrum Virtualize management interface and click **Access option** in the navigation menu. Click **Create User Group** at the bottom of the window (Figure 4-4).

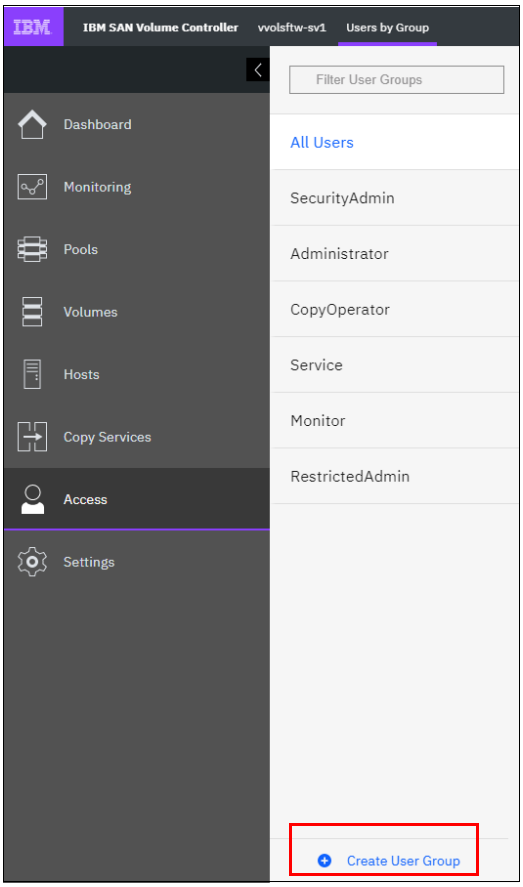


Figure 4-4 Create User Group: 1

2. Enter a name for the User Group to be created, and select **VASA Provider**. Click **Create** (Figure 4-5).

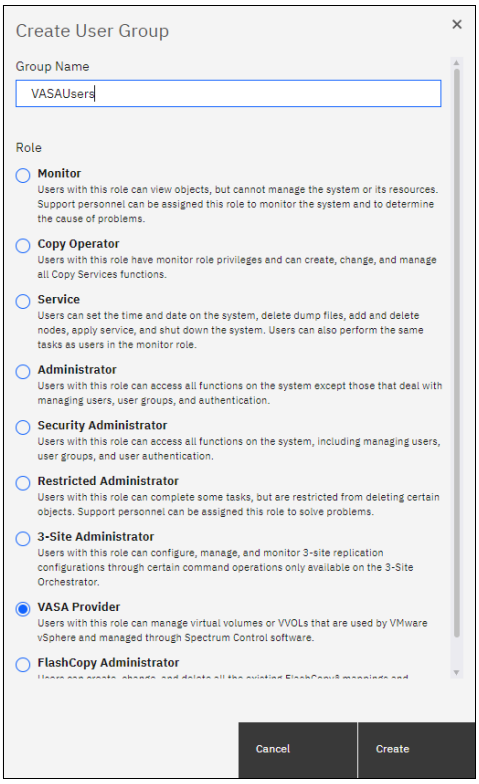


Figure 4-5 VASA Provider option

3. Click **Create User** (Figure 4-6).

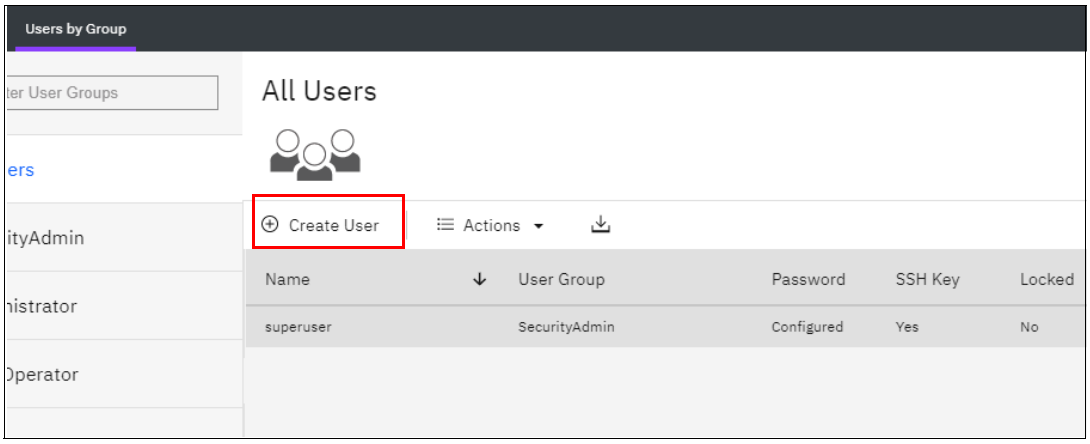


Figure 4-6 Create User: 1

4. Enter a suitable username and select the **VASAUsers** group created previously. Enter and confirm the password for this user account then click **Create**, as shown in Figure 4-7 on page 49.

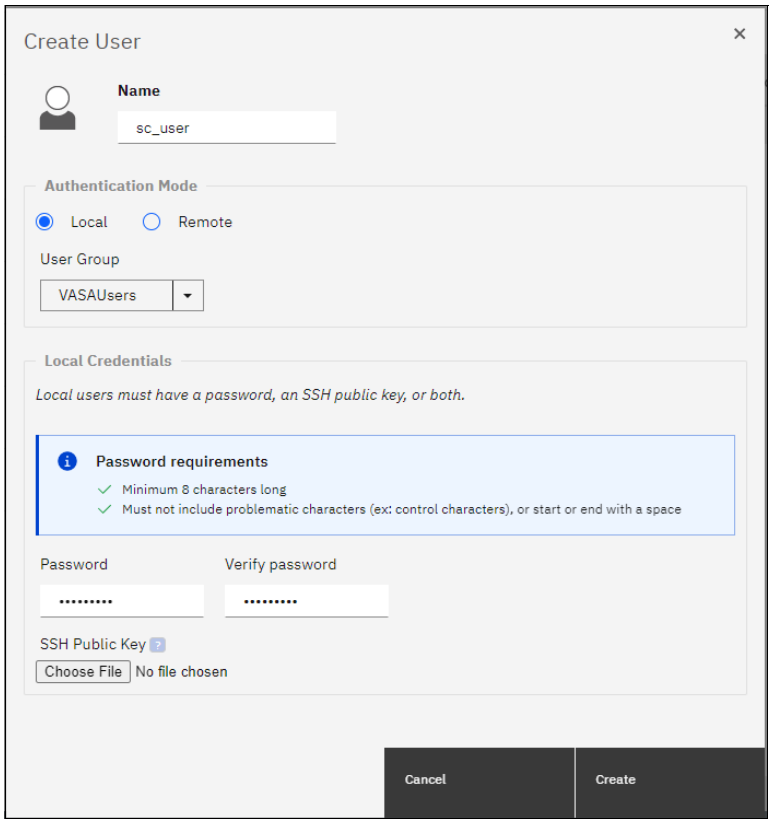


Figure 4-7 Create User: 2

5. When the initial configuration wizard is complete, in the IBM Spectrum Connect management interface, you see a “Guided tour” that provides a brief overview of the interface. When the tour is completed, empty inventory of IBM Spectrum Connect is displayed (Figure 4-8).

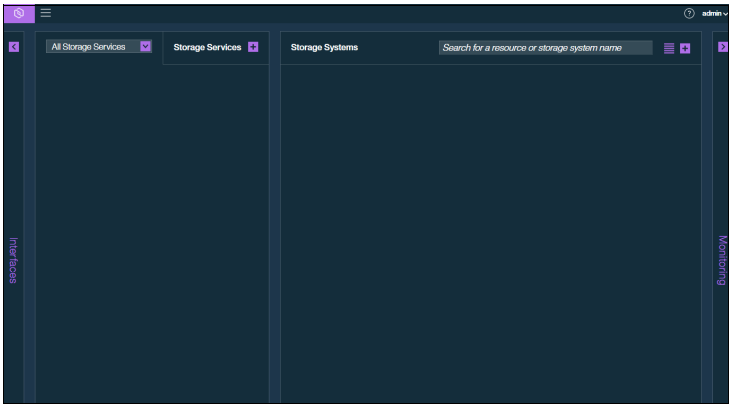


Figure 4-8 Empty inventory of IBM Spectrum Connect

6. Click the Interfaces pane, on the left side of the screen, to view and configure the Cloud Interfaces. You can configure items such as a vCenter server for the IBM Storage Enhancements or the connector for vRO.

   Alternatively, click the Monitoring pane, on the right side of the screen, to configure the vROps Manager integration.

4.1.5 Registering a Storage System into IBM Spectrum Connect

On the main page of the IBM Spectrum Connect management interface, click the plus (+) icon, and enter the fully qualified domain name (FQDN) or IP address for the storage system (Figure 4-9).

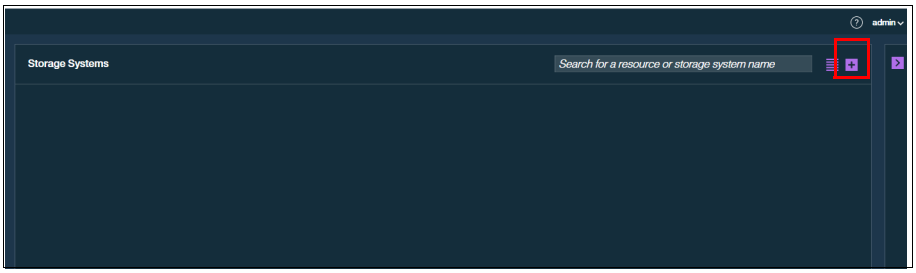


Figure 4-9 Clicking the plus icon

You are prompted to enter only the IP or hostname of the storage system because the Storage Credentials were defined in the initial configuration wizard (Figure 4-10).

Note: If you use VMware vSphere Virtual Volumes, ensure that the Storage Credentials are configured with the “VASA Provider” role.

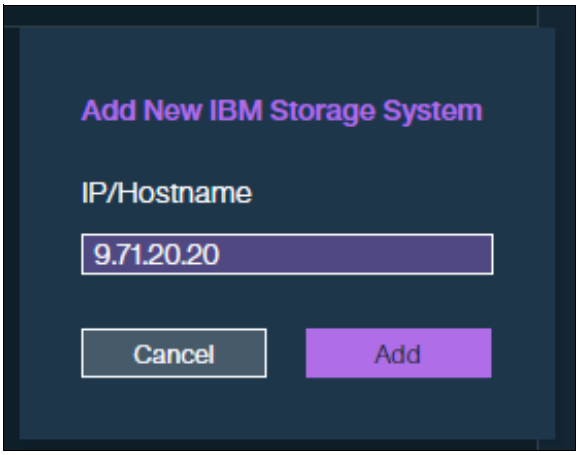


Figure 4-10 Entering the IP or hostname

7. Click **Add**. The Storage System is now represented in the IBM Spectrum Connect UI (Figure 4-11 on page 51).

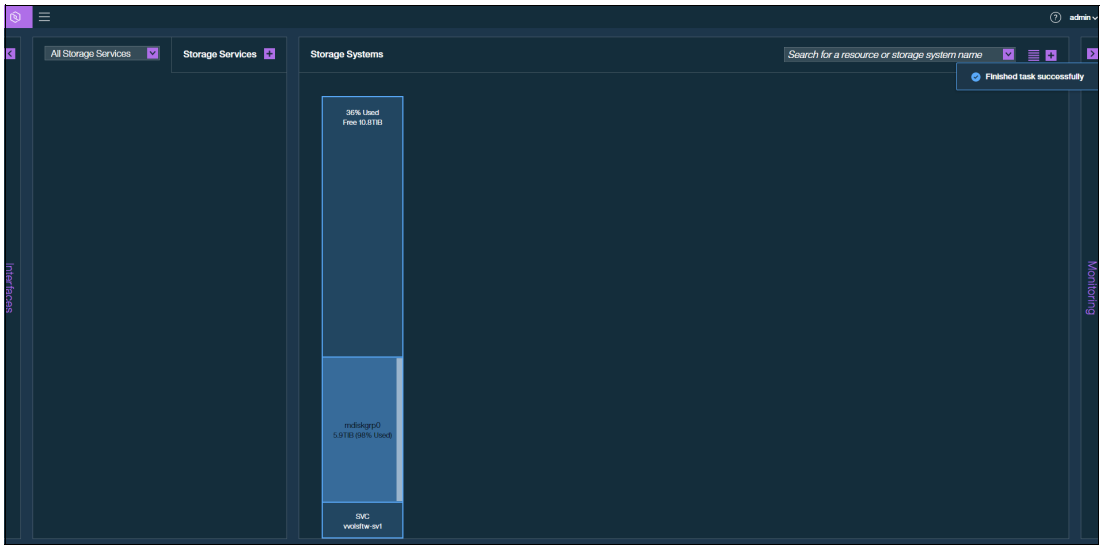


Figure 4-11 Storage System shown in the IBM Spectrum Connect UI

4.2 Understanding Storage Spaces and Storage Services

To allow a cloud interface the ability to create objects on the storage system, relationships must be created between the specific interface and a specific storage system. Also, you must define how volumes should be created without exposing every configuration option to the VMware interface. For example, should the volumes be Space Efficient, Deduplicated, or Encrypted? Perhaps a storage administrator would like to enforce that every volume is mirrored between two different external storage controllers, or that every volume must be thin-provisioned to ensure the maximum use of the available storage capacity.

IBM Spectrum Connect uses the concept of a *Storage Service* for simpler and more flexible storage management.A Storage Service can also be described as a *Storage Provisioning Profile*. If a storage administrator defines the capabilities on the service (or profile), all volumes that are created in that service are created the same way and inherit the same attributes.

Multiple Storage Services can be presented to the VMware interface to present multiple provisioning options to the vSphere administrator (Figure 4-12).

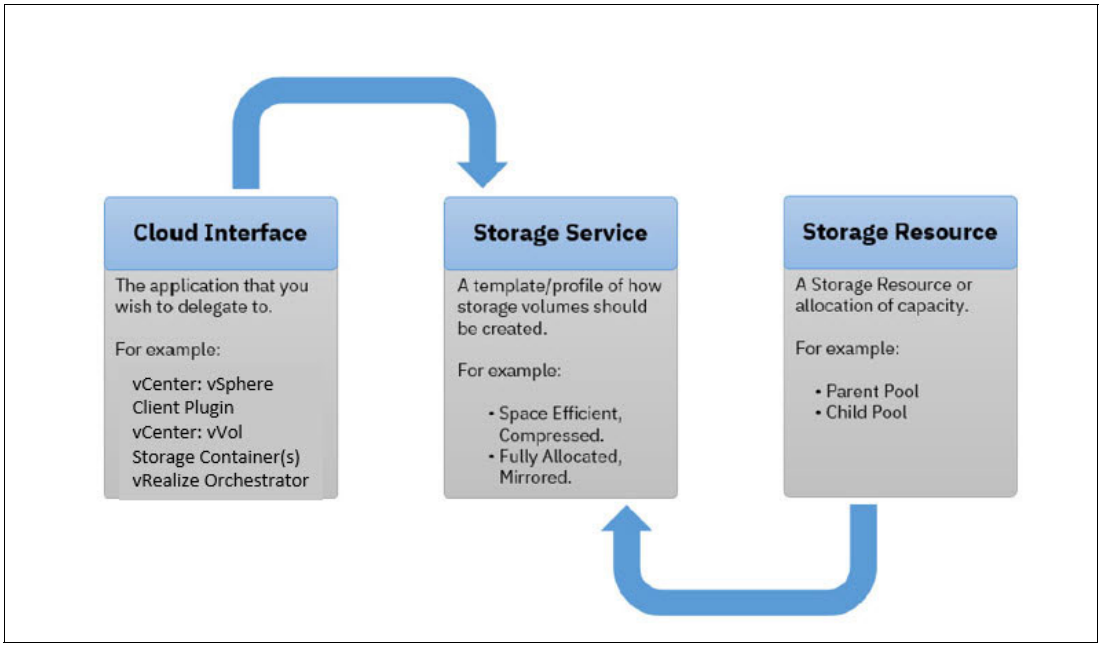


Figure 4-12 Storage Spaces and Storage Services

However, a storage administrator might be reluctant to grant full access to a storage system because if too many volumes are created, the capacity is difficult to manage. In this scenario, the storage administrator can create child pools to present ring-fenced allocations of storage capacity to a Storage Service. A vSphere administrator can then create volumes on-demand within that storage allocation.

4.2.1 Creating Storage Services

To create Storage Services, complete the following steps:

1. Identify a suitable Storage Space and click the plus icon (**+**) to create a new Storage Service (Figure 4-13 on page 53). Enter text into the Name and Description fields to later identify the intended use and select the capabilities that best suit the application requirements.

   Multiple Storage Services, with different capabilities can be associated to the vCenter interface, which allows the vSphere administrator to select which *profile* of storage should be used for a situation.

Tips:

- When a Storage Service is dedicated to use with VMware vSphere Virtual Volumes, ensure that the **vVol Service** checkbox is selected.
- Consider a mix of configuration options, such as Space Efficiency, Encryption, Tier, and Data Reduction.

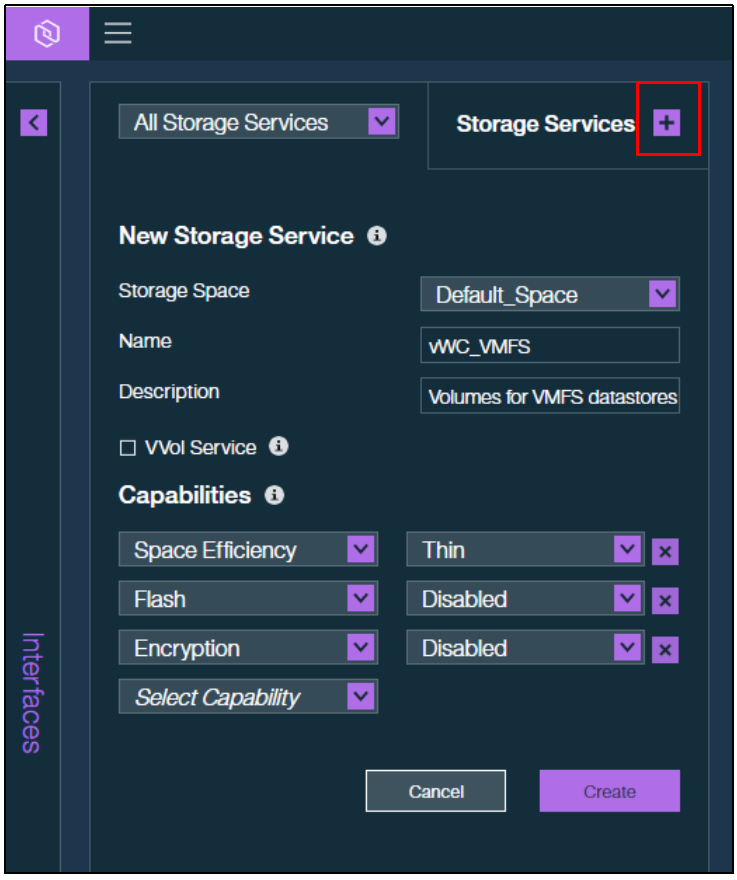


Figure 4-13 Creating Storage Services

Note: When you create the Storage Service and select the capabilities, Storage Systems that are not compatible with the current selection are unavailable, such as:

- When **Synchronous Replication** is selected, but the Storage System that is registered in IBM Spectrum Connect is not yet configured for HyperSwap.
- When Flash storage is requested, but storage pools do not exist in the Flash devices.

2. After you define the required capabilities of the storage profile, click **Create**.

   The newly created Storage Service is now listed. Notice that allocated storage capacity does not exist for this Storage Service. A storage resource needs to be associated to the Storage Service.

4.2.2 Allocating capacity to Storage Services

To allocate storage capacity to the Storage Services, complete the following steps:

1. Right-click the Storage Service and select **Manage Resources** (Figure 4-14).

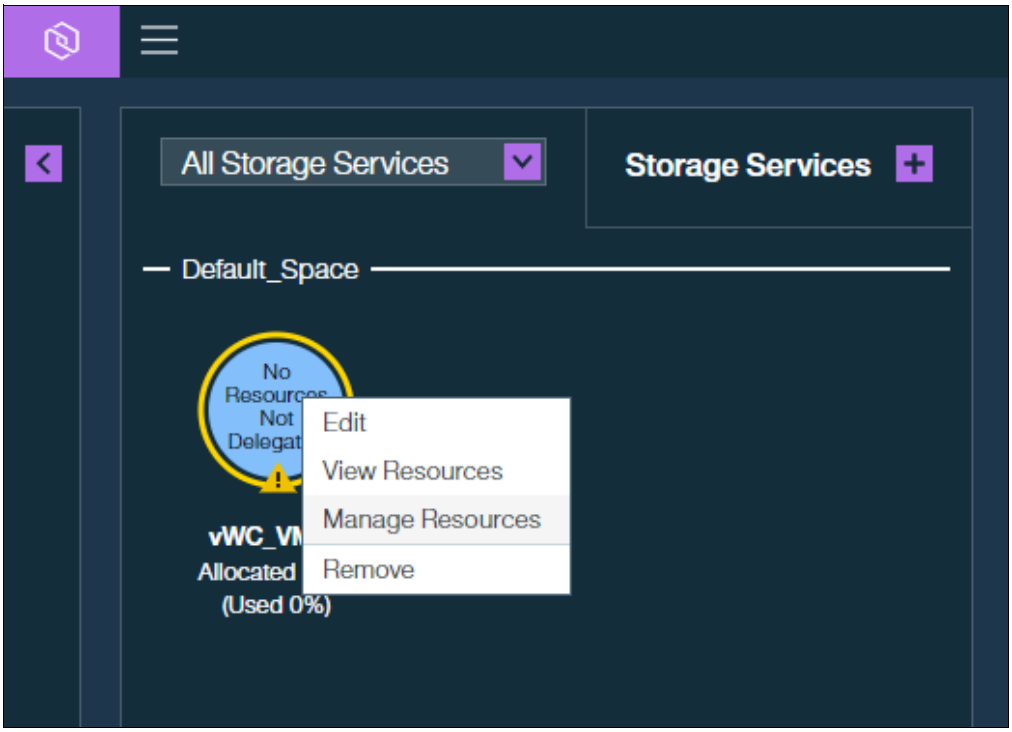


Figure 4-14 Selecting Manage Resources

The Manage Resources window opens (Figure 4-15 on page 55).

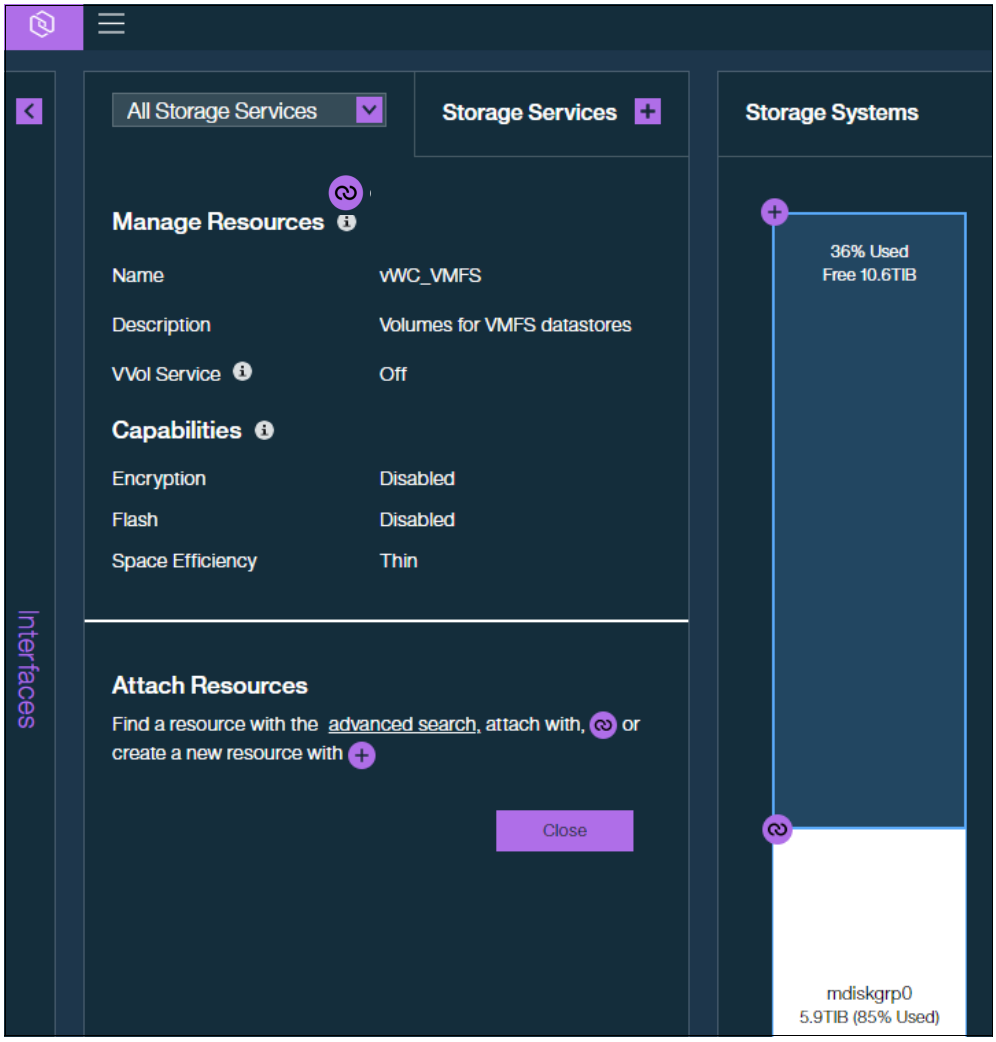


Figure 4-15 Manage Resources window

Note: A Storage Resource can be either an existing parent pool within the IBM Spectrum Virtualize storage system, or a new or existing child pool. A child pool is a ring-fenced allocation of storage capacity that is taken from an existing parent pool.

2. To associate an existing parent pool to the selected Storage Service, click the Delegate icon ( ) associated with the specific parent pool.

   Alternatively, to create and associate a new child pool, click the **Plus** icon (**+**) at the top of the Storage System to create a Storage Resource, as shown in Figure 4-16.

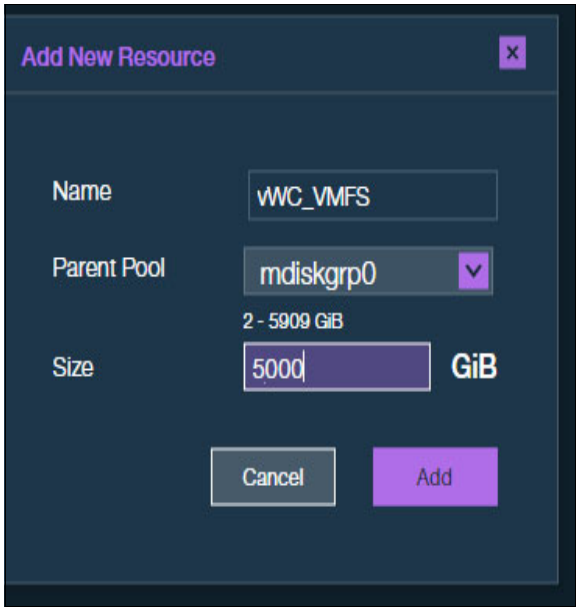


Figure 4-16 Add New Resource

This step creates a child pool of a defined name and capacity within a specified parent pool.

Note: When using a HyperSwap configuration, you are asked to specify the parent pool at both sites (Figure 4-17). This specification creates a paired child pool with the same capacity that is defined at each site.

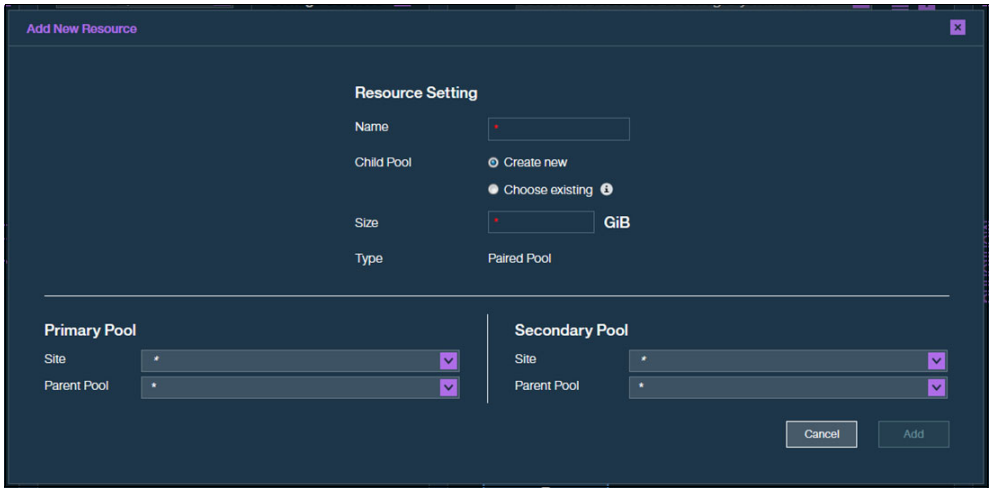


Figure 4-17 HyperSwap configuration

3. Verify that the Storage Resource allocation for the Storage Service is correct and click **Close** (Figure 4-18 on page 57).

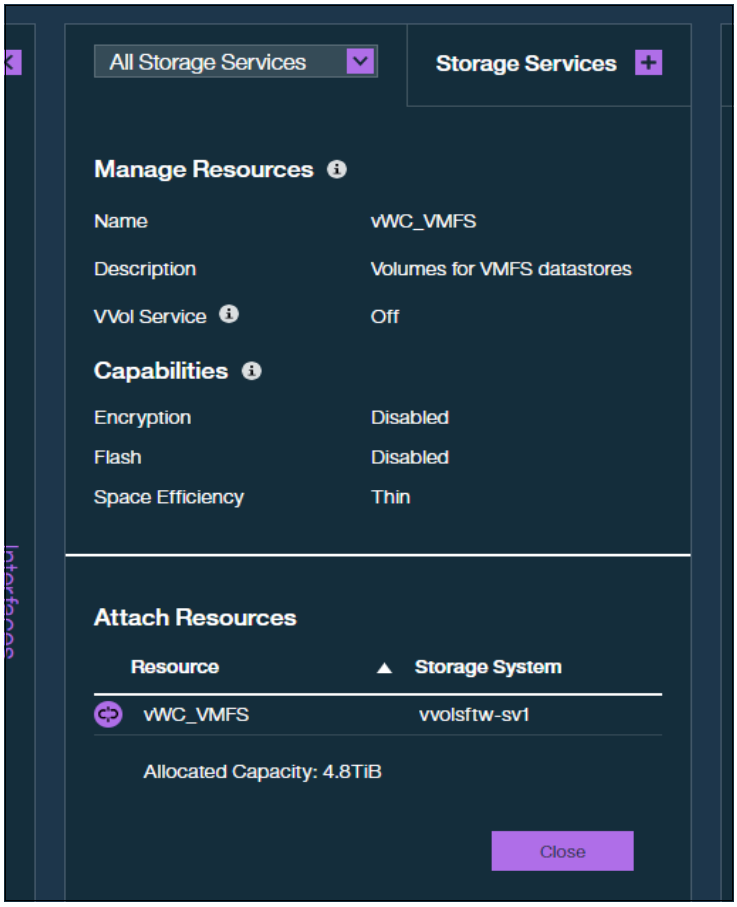


Figure 4-18 Verifying the Storage Resource allocation

4.2.3 Delegating Storage Services to vCenter

The Storage Service is created and some storage capacity is allocated. You can now delegate the Storage Service to the vCenter interface.

Tip: When you use vVol-enabled Storage Services, this step is optional. However it is a best practice because it provides more visibility of the individual vVol-to-VM relationships.

To delegate Storage Services to vCenter:

1. In the Interfaces window, select the **vCenter interface** to which you want the Storage Service to have access (Figure 4-19). Click the Delegate icon ( ).

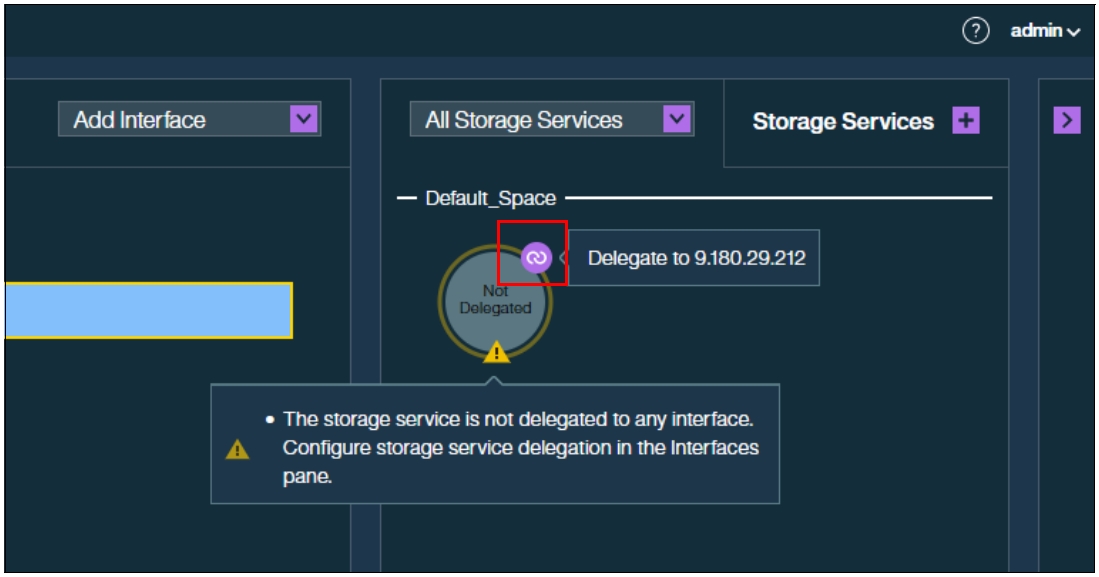


Figure 4-19 Storage Service delegation

Note the changes that occur when the Storage Service is delegated:

- The color changes to yellow on the Storage Service (Figure 4-19).
- The Delegate icon changes from to (Figure 4-20).
- The Allocated Storage capacity, which is described just under the selected vCenter interface, is updated to reflect the new Storage Service delegation (Figure 4-20).

Figure 4-20 New Storage Service delegation

2. If you want to remove the delegation, click the Delegate icon ( ) icon to disassociate the Storage Service mapping (Figure 4-20).

> **Note:** The creation and removal of delegation of Storage Service to Interface does not impact the existing volumes or host mappings on the storage system.

4.3 VMware vSphere Virtual Volumes

This section provides an overview of VMware vSphere Virtual Volumes and describes how to configure IBM Spectrum Virtualize to support vVols.

With the introduction of IBM Spectrum Virtualize 8.5.1.0 (or later), the vVol function can be provided in either of two ways:

- IBM Spectrum Connect: An isolated application that runs on a separate VM within the vSphere environment that converts API calls from vSphere into CLI commands that are issued to the IBM FlashSystem array.
- Embedded VASA Provider: This application runs natively as a service on the IBM FlashSystem configuration node, communicates directly with vSphere, and is not dependent on any other external application or server.

The two models *cannot* be run in parallel, so you must choose which one to implement in your infrastructure.

To learn more about the Embedded VASA Provider feature and to see whether it is appropriate for your environment, see Chapter 6, “Embedded VASA Provider for Virtual Volumes” on page 129.

4.3.1 VMware vSphere Virtual Volumes overview

IBM Spectrum Connect delivers comprehensive storage-virtualization support that uses vVol technology.

The vVol architecture, introduced in VMware vSphere 6.0 with VASA 2.0, preserves the concept of a traditional data store, maintaining familiarity and functions and also offers some benefits of legacy Raw Device Mappings.

Virtual Machine File System data stores

A large volume from an IBM Spectrum Virtualize storage system can be formatted as a Virtual Machine File System (VMFS) data store and shared between a number of hosts in the vSphere cluster. On traditional VMFS data stores, each virtual machine disk (VMDK) that is associated to a VM is essentially a file on a file system, effectively a layer of abstraction apart from how the storage system would process the workload.

VM-level snapshots provide an excellent way of providing a point-in-time copy of a VM. However, the read/write overhead when using snapshots for anything beyond a 24 - 72 hour period can greatly impact the performance of the VM.

Also, in this scenario there are several VMs, distributed across many ESXi hosts, all performing I/O operations to that volume generate a high workload. When investigating these performance issues, the storage administrator might observe high IOPS measurements, and high response times against that volume. In this example, the storage system would have limited knowledge of where the I/O workload was coming from or the cause of the performance issues. Similarly, if multiple VMs are performing sequential workloads to their VMDK disks, the ESXi layer confuses the stream of I/O and cannot use the advanced caching features of IBM Spectrum Virtualize.

Finally, when performing logical unit number (LUN)-level operations on the storage system (for example, volume-level snapshots that use a FlashCopy map), all VMs that are on the VMFS data store that is backed by that volume are affected.

Raw Device Mappings

Raw Device Mappings (RDMs) provide a more traditional method of storage provisioning where a single volume from a storage system could be dedicated to a VM's VMDK disk. This method provides a direct mapping of logical block addressing from a VMDK file to the volume as presented by the storage system.

However, RDMs do not offer the same level of functions when compared with VMFS data stores, specifically regarding VM cloning and snapshots. RDMs also require a storage administrator to present dedicated volumes for each VMDK file, which increases storage-provisioning management overhead.

VMware vSphere Virtual Volumes

vVols can be viewed as taking the benefits from both VMFS and RDM storage concepts, with none of the associated limitations. The vVols model provides a 1:1 association of VMDK files to Volumes in the storage system. For every VMDK file that is created for a VM, a volume is created on the storage array similar to how RDMs are configured.

The main advantages of vVols are:

- Automated storage provisioning: vVols are created, deleted, and mapped or unmapped automatically. As a VMware administrator performs tasks in vCenter, such as creating VMs or adding more VMDKs to an existing VM, the associated API calls are sent to the IBM Spectrum Connect server. The API calls are then processed and converted into CLI commands to run against the storage system by using SSH. This action reduces the administrative overhead of managing VM storage when compared to RDM volumes because action is not required from the storage administrator.
- Improved VM snapshots, storage migration, and clones: Because each VMDK file exists natively as a volume within IBM Spectrum Virtualize, VM-level snapshots, storage migrations, and VM clone operations are offloaded to the storage system in the form of FlashCopy operations that are processed asynchronously.
- Storage Policy Based Management: To aid storage provisioning, Storage Policies can be created in vCenter and allocated on a per-VMDK basis, which can then be associated to Storage Services within IBM Spectrum Connect.
- Granularity: More granular visibility of VM storage consumption and performance. Because each VMDK exists as a volume within IBM Spectrum Virtualize, the storage administrator can see the precise workload for an application rather than a combination of all I/O from multiple VMs to a volume.
- No shared storage: No per-LUN IO contention or VMFS overhead. Sequential I/O from multiple VMs can be easily identified and processed in a more efficient manner on the storage system.

With vVol, the IBM storage systems become aware of individual VMDKs, which allows data operations, such as snapshots, to be performed directly by the storage system at the VM level. The storage system uses IBM Spectrum Connect (the VASA provider) to present vVols to the ESXi host and inform the vCenter of the availability of vVol-aware storage by using Protocol Endpoints (PEs).

One PE (also known as *Administrative LUN* or *PE*) is presented at each node in the IBM Spectrum Virtualize cluster. A PE presents itself as a traditional storage device and is detected by an ESXi server like a normal volume mapping. However, PEs offer a more efficient method for vVols to be mapped and detected by ESXi hosts when compared to traditional volume mappings and do not require a rescan of host bus adapter (HBA).

PEs are automatically presented to all hosts that are defined within the IBM Spectrum Virtualize storage system and configured with:

- The vVol host type, when the UI is used to perform configuration
- The **adminlun** option, when the CLI is used to perform configuration

Storage services are configured on the IBM Spectrum Connect server by the storage administrator. Storage services are then used to configure various storage containers with specific capabilities that can later be configured as vVol data stores in the vSphere Client.

4.3.2 Configuring IBM Spectrum Virtualize to support vVols

To enable vVols on the IBM Spectrum Virtualize storage system, complete the following steps:

1. Go to the Settings window of the management interface, and select **vVol** (Figure 4-21).

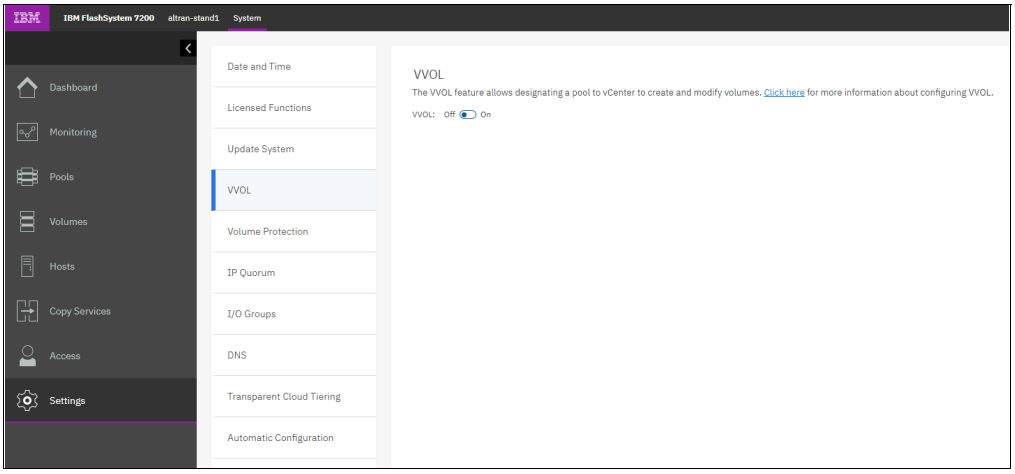


Figure 4-21 Settings window: VVOL

Note: To keep time synchronized between all components of the vVol infrastructure, ensure that an NTP server is defined on the storage system, IBM Spectrum Connect server, vCenter, and ESXi hosts.

2. Set the VVOL switch to **ON** to display more options (Figure 4-22).

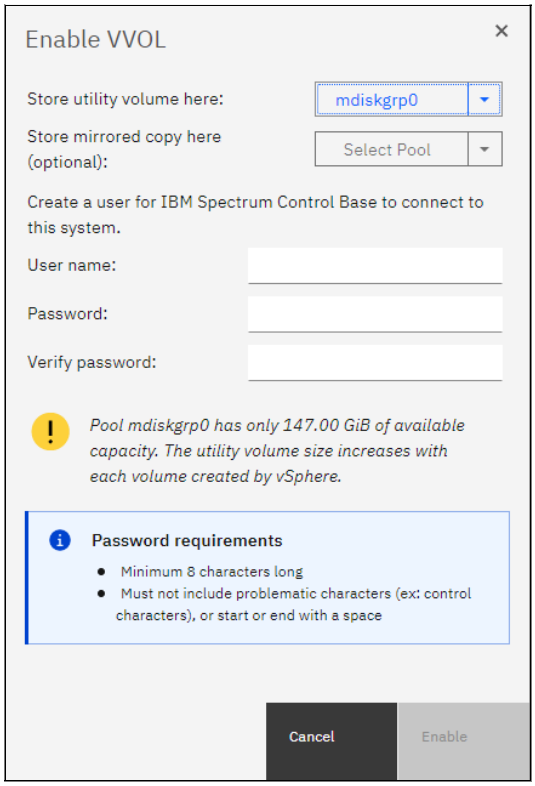


Figure 4-22 Enable VVOL:1

3. Select a parent pool to store the Utility Volume. This volume is configured as a space-efficient 2 TB volume, which stores the associated metadata database that is required by IBM Spectrum Connect to store information that is required to manage the vVol environment.

   Although the volume is created with a capacity of 2 TB, it is unlikely to require more than 2 GB of capacity.

   If possible, consider creating a mirrored copy of this Utility Volume by selecting a second pool to store the additional copy.

4. Define credentials for a newly dedicated user account, which enables the IBM Spectrum Control server to connect to the CLI of the storage system.

   This process initially creates a User Group within the IBM Spectrum Virtualize Storage system that is assigned with the VASAProvider role, and then creates the User Account by using the specified credentials within that user group.

5. Click **Enable** (Figure 4-22).

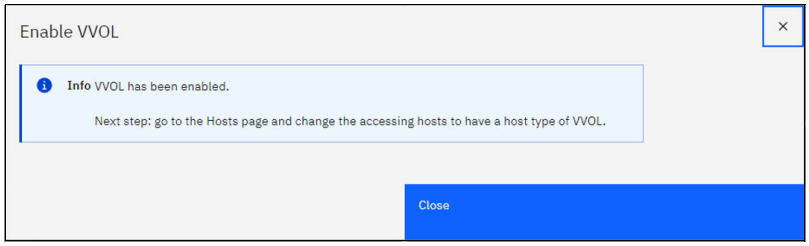


Figure 4-23 Enable VVOL: 2

6. Configure host objects to be vVol-enabled.

   You can modify host types on either an individual host or a host-cluster basis, as follows:

   – Individual hosts:

     i. Go to the Hosts object. Right-click the hosts that you want to enable and select **Modify Type** (Figure 4-24).

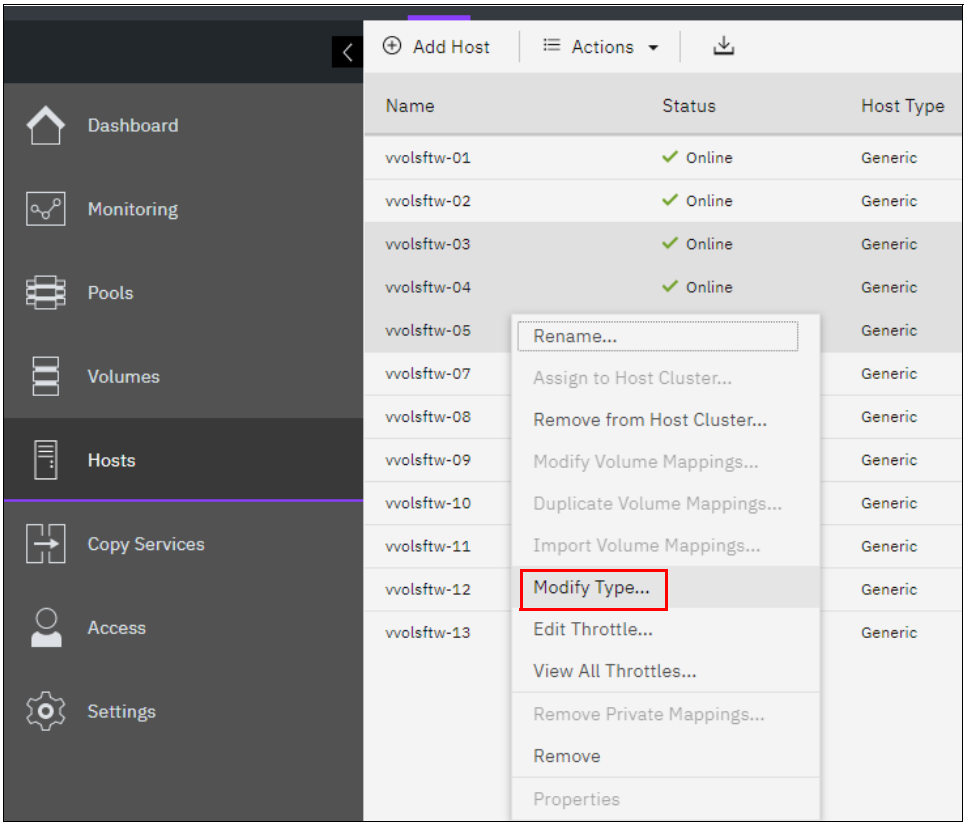


Figure 4-24 Selecting Modify Type

     ii. Select **VVOL** and click **Modify** (Figure 4-25).

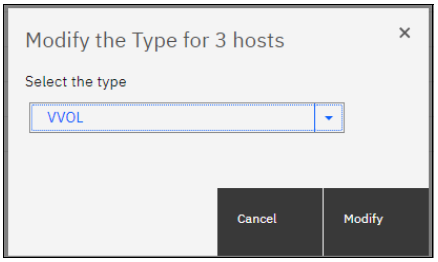


Figure 4-25 Clicking Modify

– Configure host clusters, as follows:

i. Go to the Host Clusters object. Right-click the host cluster that you want to enable and select **Modify Host Types**, as shown in Figure 4-26.

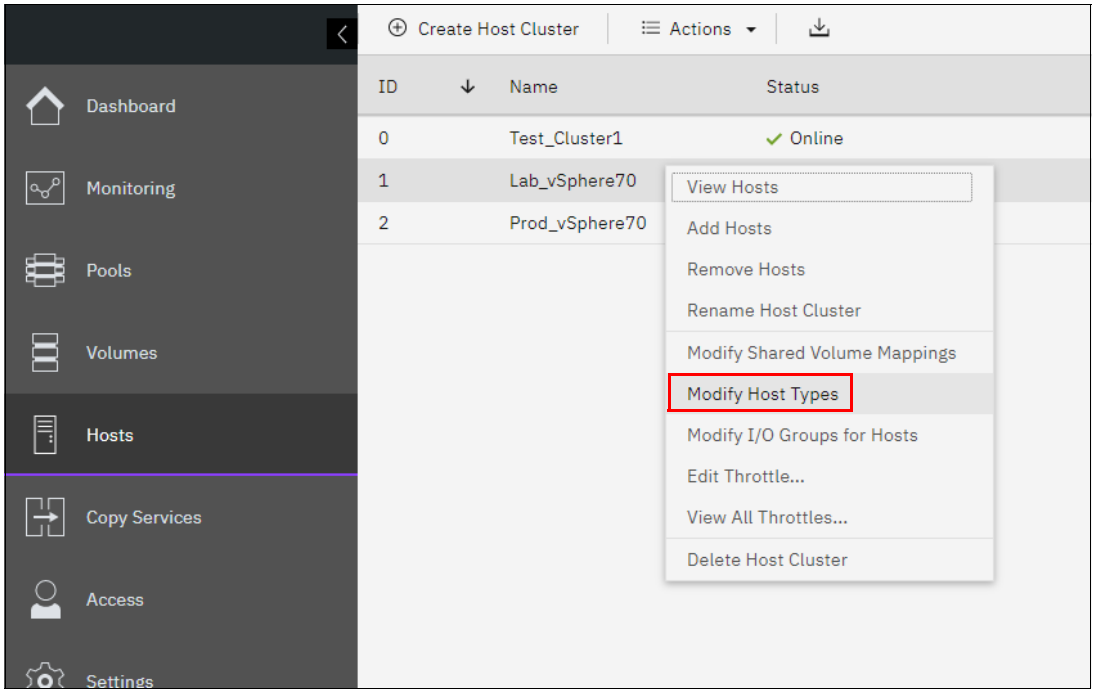


Figure 4-26 Selecting Modify Types

ii. Select **VVOL** and click **Modify** (Figure 4-27).

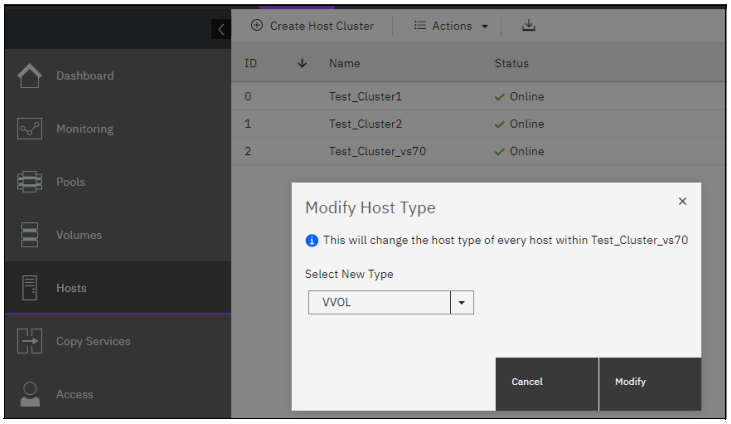


Figure 4-27 Clicking Modify

A warning is displayed (Figure 4-28 on page 65).

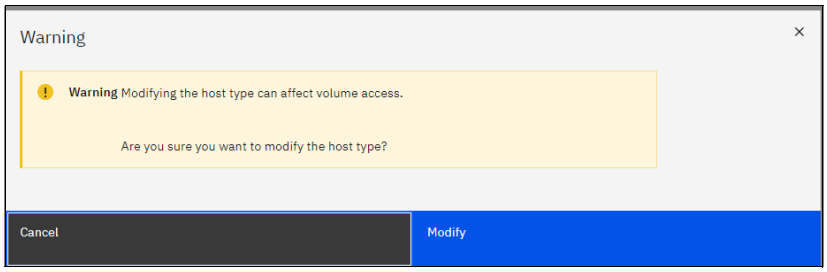


Figure 4-28 Warning message

Hosts with a generic host-type allow volumes to be mapped with a Small Computer System Interface (SCSI) ID of 0 - 2048. However, ESXi detects only SCSI IDs 0 - 1023. Hosts with a vVol (or `adminlun`) host type are limited to SCSI IDs 0 - 511. This warning refers to existing volumes that might be mapped with a SCSI ID above 511 that are lost when changing the host type to support vVols. Normally, the lowest available SCSI ID is used when a volume is mapped to a host. Therefore, it is only in rare circumstances where either:

- More than 511 volumes are mapped to a host or host cluster.
- A SCSI ID above 511 is specified when a previous mapping was created.

7. Verify the existing SCSI IDs by reviewing the existing host mappings for a host or host cluster, as follows:

   a. Select **Hosts** → **Mappings** in the navigation menu, and select **All host mappings**. Sort the mappings in descending order by the SCSI ID column to show the highest SCSI IDs in use (Figure 4-29).

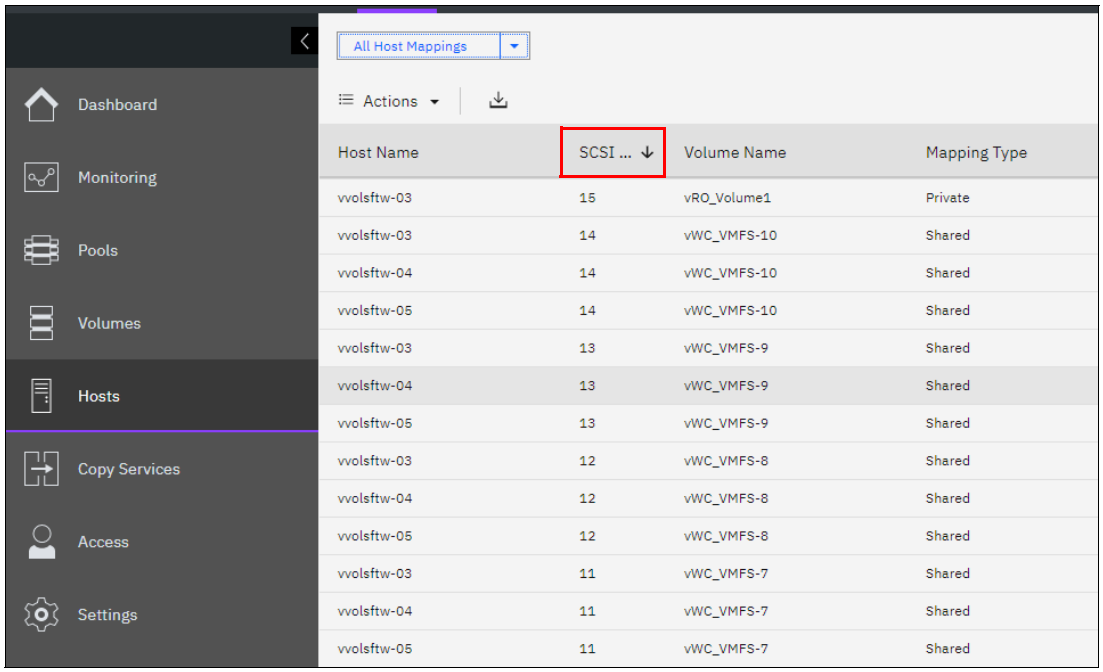


Figure 4-29 Sorting the mappings in descending order

   b. Ensure that SCSI IDs above 511 are not in use before you change the host type.

8. Confirm the identifiers for the PEs by connecting to the CLI of the storage-system management interface by using SSH and run the following CLI command: **`IBM_FlashSystem:FS9200-CL:superuser>svcinfo lsadminlunv`** (Example 4-1).

Example 4-1 The svcinfo lsadminlun command

```
IBM_FlashSystem:FS9200-CL:superuser>svcinfo lsadminlun

id SCSI_id          UID
0  0300000000000000 600507680CD00000DC000000C0000000
1  0301000000000000 600507680CD00000DC000000C0000001
```

Note: In Example 4-1:

- The `SCSI_id` column contains the SCSI ID as a hexadecimal value.

  The corresponding value, as reported by either vSphere or the ESXi, is a decimal number.
- The `UID` column appears similar to a traditional LUN UUID. However, the “C” character distinguishes it from a traditional volume.

c. Rescan the HBAs of the hosts to detect the PEs.

d. Verify that the PEs are visible from the ESXi servers.

In vSphere, select a host from the inventory and select **Configure** → **Storage Devices** (Figure 4-30).

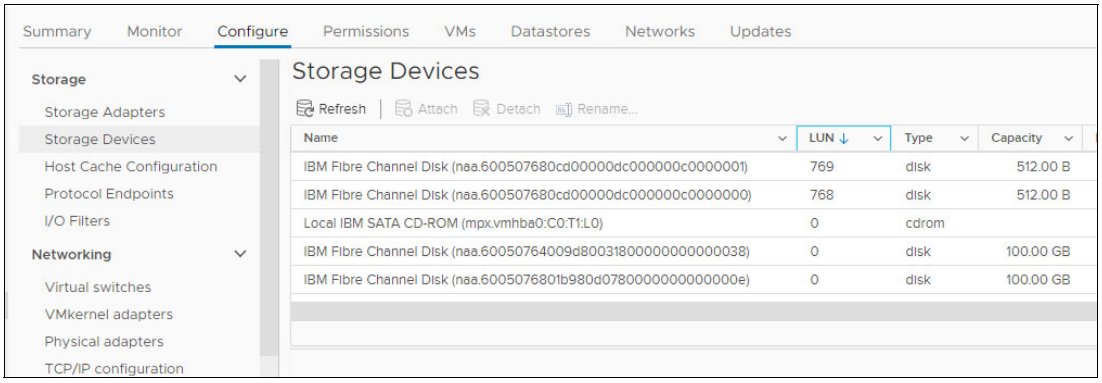


Figure 4-30 Storage Devices

e. Sort the LUN column in descending order and confirm that a PE exists for each node in the storage system.

In a standard two-node (single I/O-group) cluster, two PEs are presented. SCSI IDs that are used for the first PEs start at 768-769, and increase for each additional node in the storage system.

4.3.3 Configuring IBM Spectrum Connect

VASA Credentials, which vCenter uses to connect to the IBM Spectrum Connect server, must be defined.

To define VASA credentials, complete the following steps:

1. Log in to the management interface of IBM Spectrum Connect.
1. Select **VASA Provider settings** from the **Settings** menu (Figure 4-31).

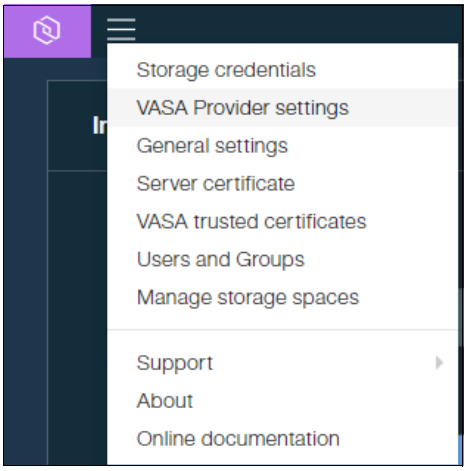


Figure 4-31 VASA Provider settings

2. Define the credentials that vCenter uses to connect to IBM Spectrum Connect when the storage provider is registered (Figure 4-32).

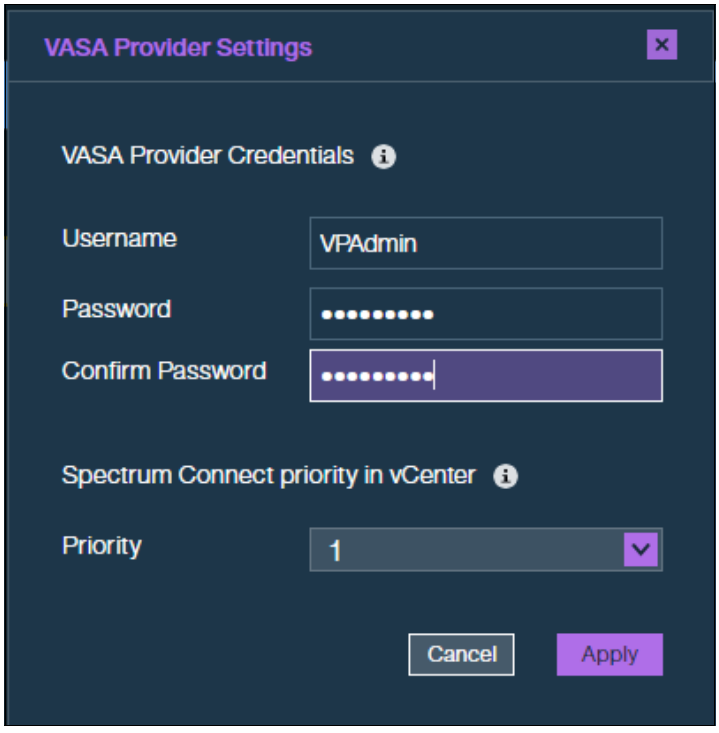


Figure 4-32 VASA Provider Credentials

3. Click **Apply**.

4.3.4 Configuring VMware vSphere vCenter

To register the IBM Spectrum Connect server in vSphere as a Storage Provider, complete the following steps:

1. In the vSphere client, select the vCenter server in the inventory on the left pane, click the **Configure** tab, and select **Storage Providers** under **Security** (Figure 4-33).

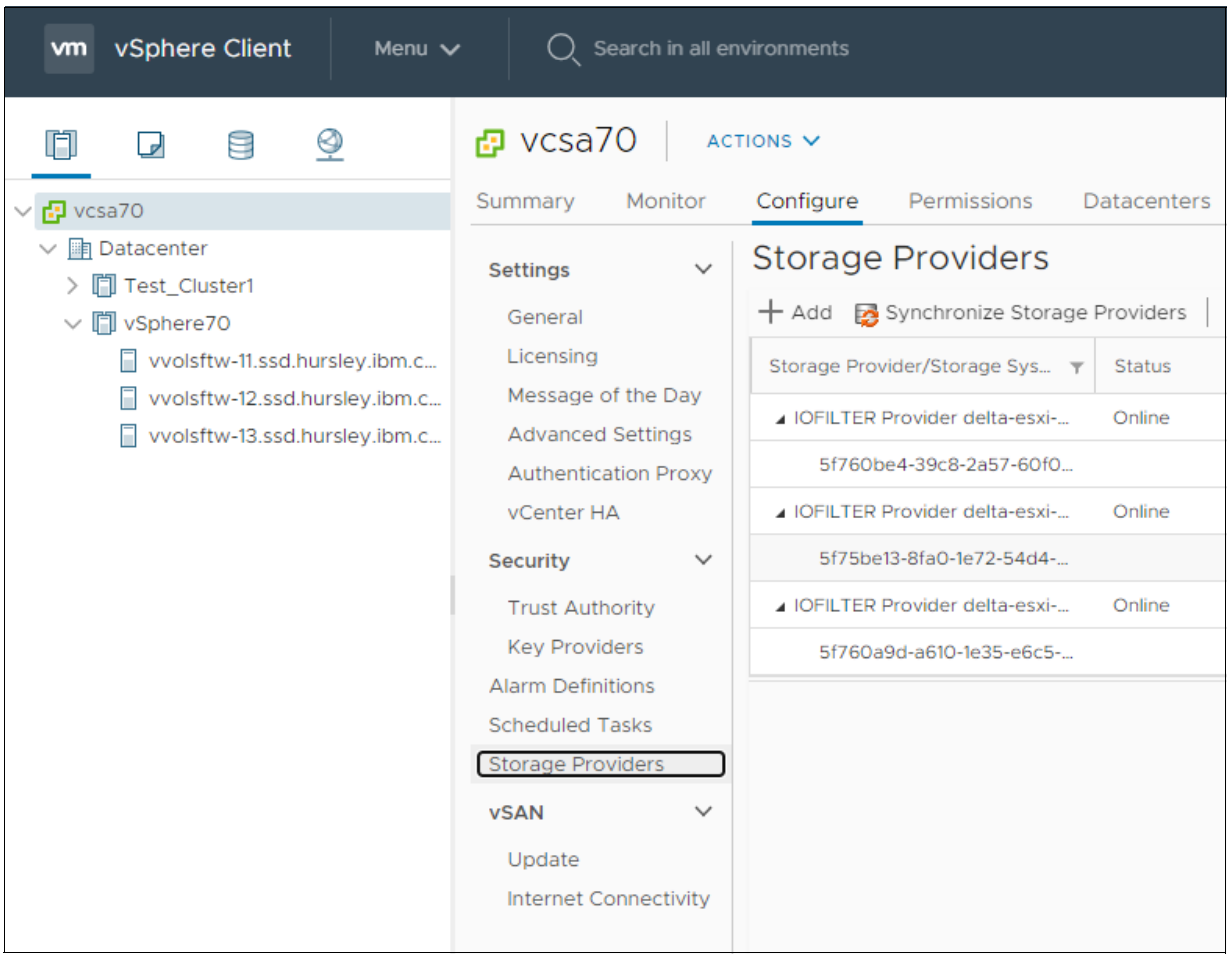


Figure 4-33 Storage Providers

2. Click **Add** to register a new Storage Provider.
3. Enter the name, URL of the IBM Spectrum Connect server, and the VASA Provider credentials (Figure 4-34 on page 69) as configured in the “Configuring IBM Spectrum Connect” on page 67.

> **Note:** The URL must be in the format of `https://<IP or FQDN of Spectrum Connect server>:8440/services/vasa`.

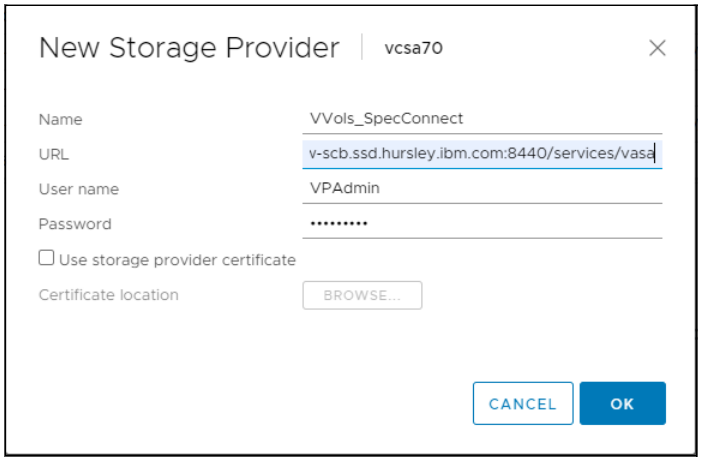


Figure 4-34 New Storage Provider

4. Click **OK**. The Storage Providers list now show the newly registered storage provider (Figure 4-35).

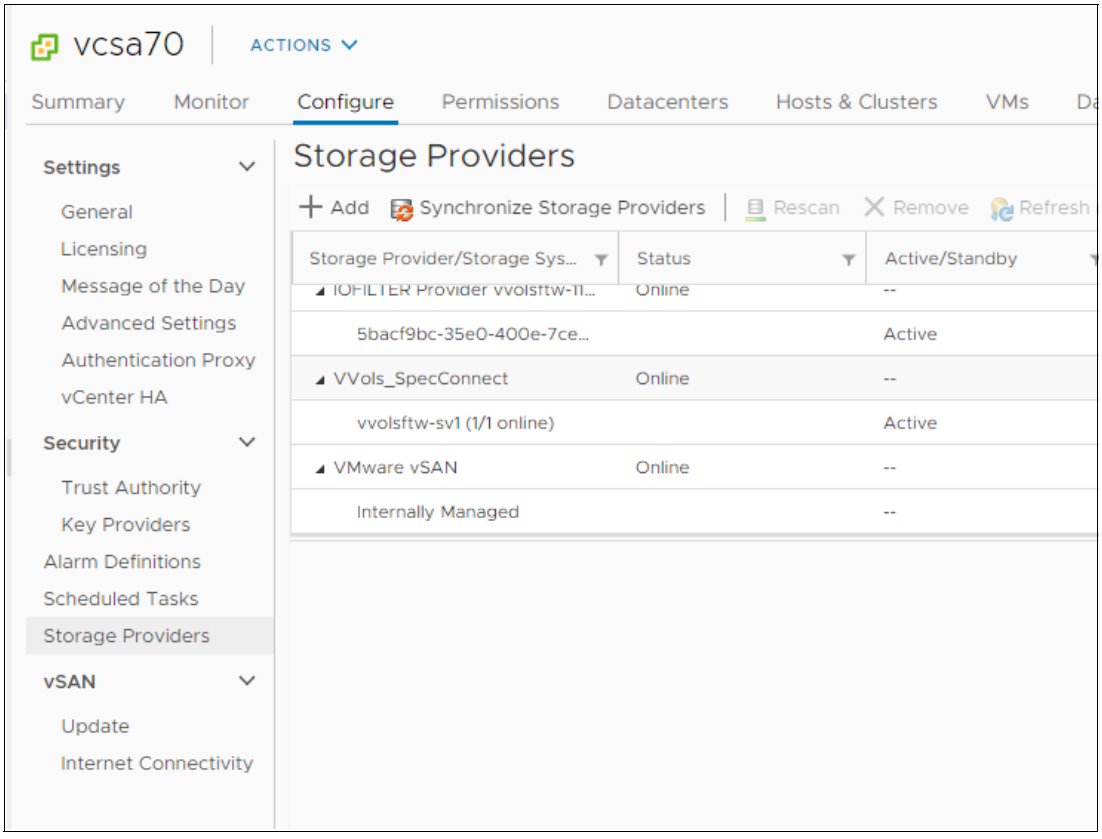


Figure 4-35 Newly registered storage provider is shown

4.3.5 Creating a vVol data store

Before you create a vVol data store, you must ensure that you created vVol-enabled Storage Services as described in 4.2.1, “Creating Storage Services” on page 52.

In this example, we created and delegated various vVol-enabled storage services, such as:

- Production systems and applications.
- Development or test environments.
- An application environment that requires encryption or extra security considerations and must be isolated from other applications.

Each storage service can be configured with unique capabilities that are associated to Space Efficiency, Encryption, or Storage Tier (Figure 4-36).

Figure 4-36 vVol-enabled Storage Services

Each Storage Services was allocated a storage resource and associated to the vCenter interface (Figure 4-37) as described in 4.2.3, “Delegating Storage Services to vCenter” on page 57.

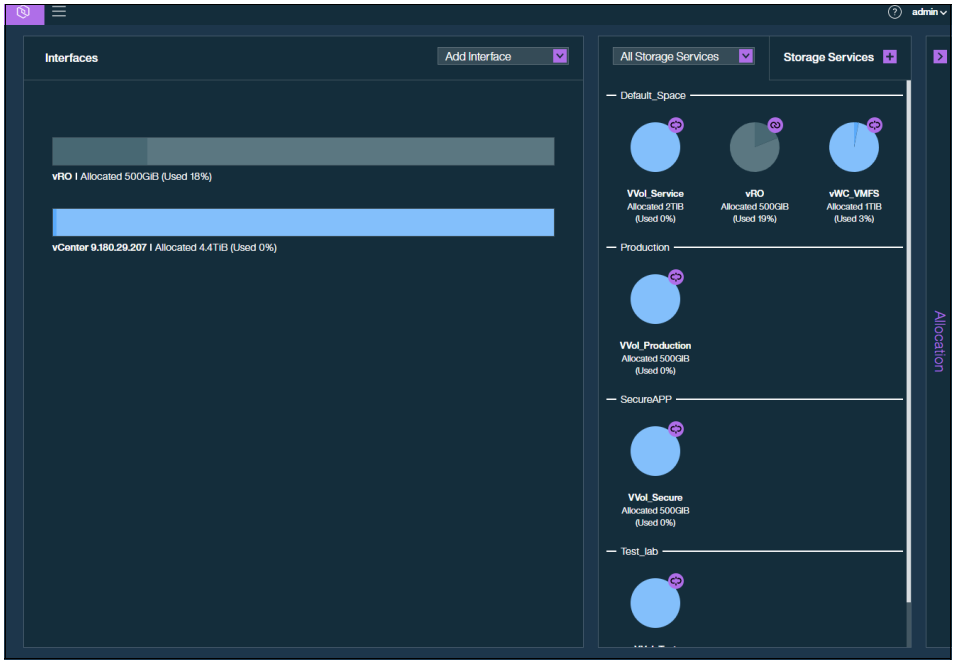


Figure 4-37 vVol-enabled Storage Services delegated to a vCenter interface

After the Storage Containers are configured, create the vVol data store by completing the following steps:

1. In the vSphere Client window, identify the host or host-cluster for which you want to configure the vVol data store by right-clicking the object in the vSphere inventory and selecting **Storage** → **New Datastore** (Figure 4-38).

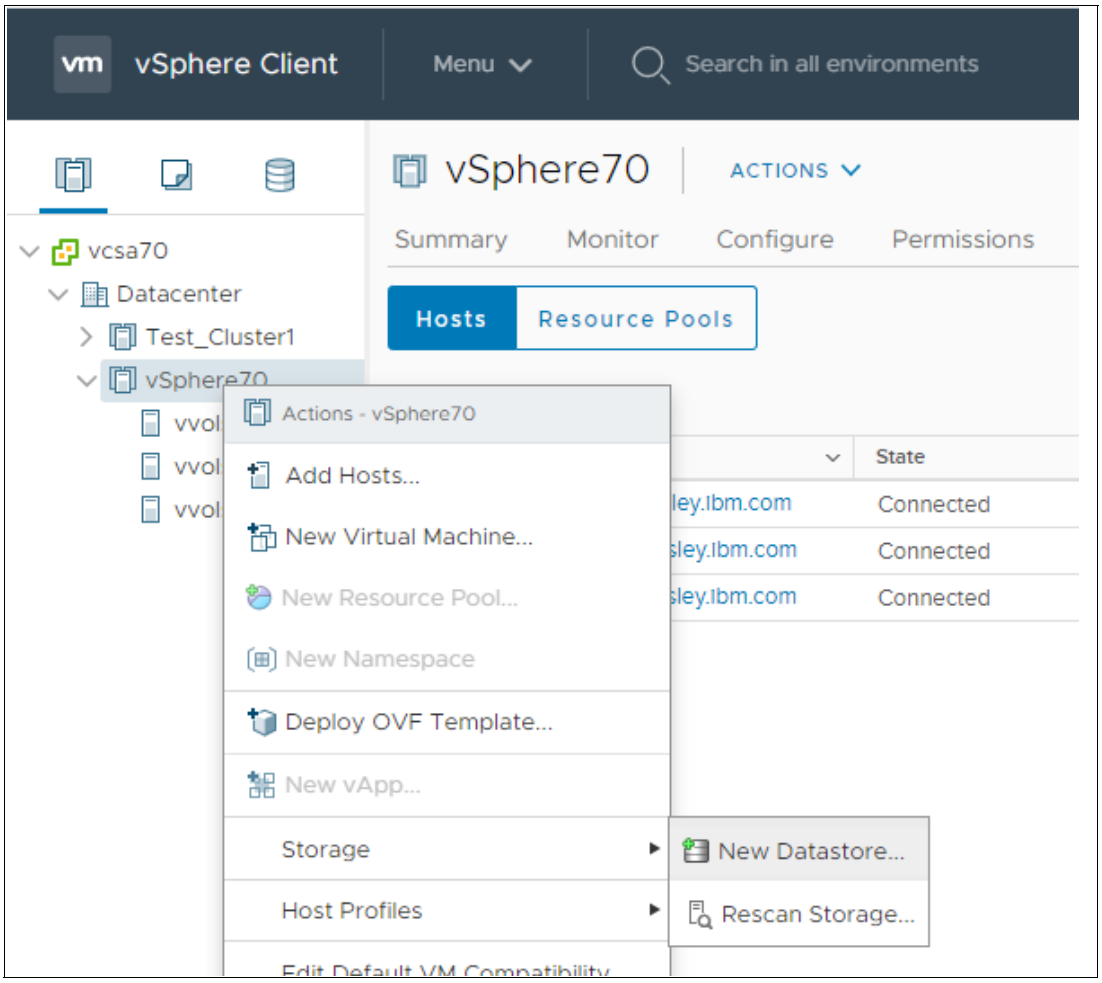


Figure 4-38 New Datastore:1

2. Select **vVol**, and click **NEXT** (Figure 4-39).

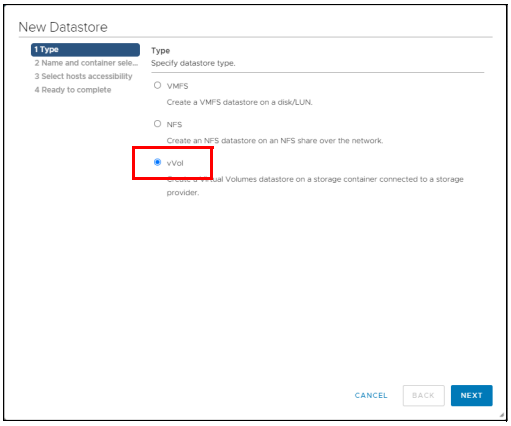


Figure 4-39 New Datastore: 2

3. Enter a name for the VVol data store, select the related Storage Container from the list, and click **NEXT** (Figure 4-40).

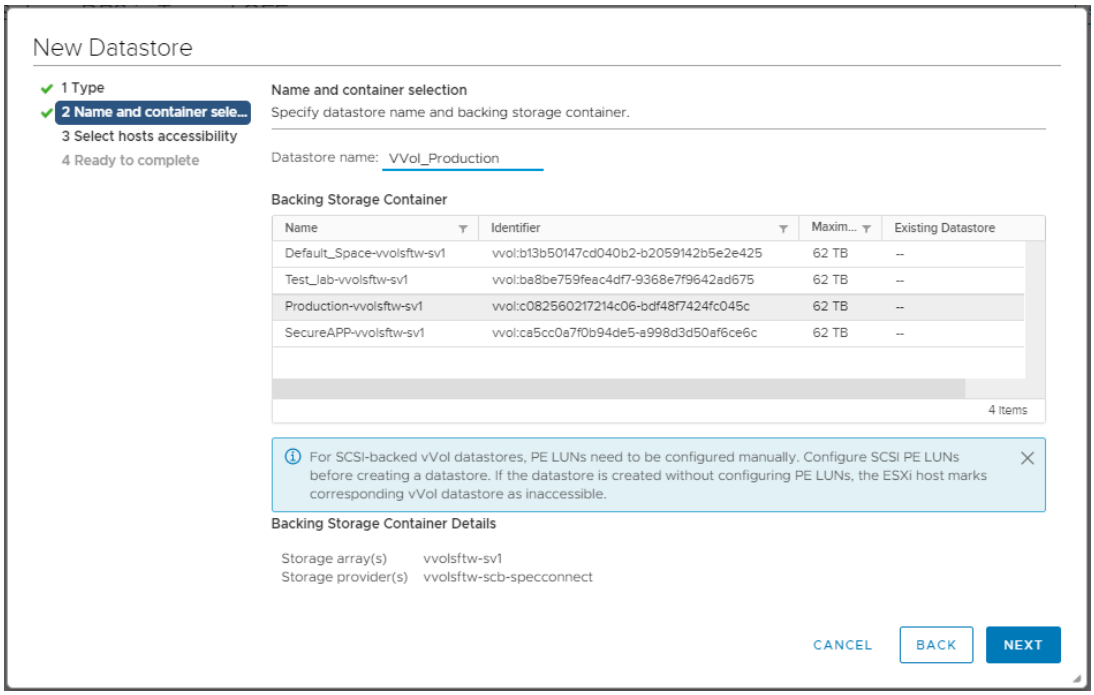


Figure 4-40 New Datastore: 3

4. Select the hosts that require access to the data store (Figure 4-41) and click **NEXT**. Review the summary and click **FINISH**.

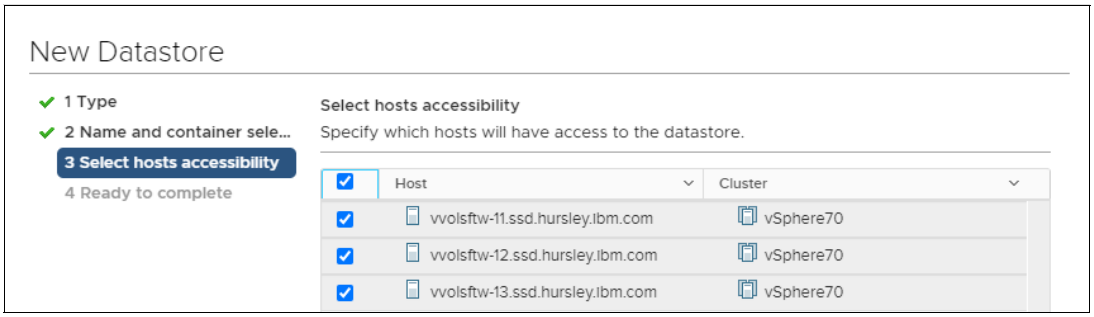


Figure 4-41 New Datastore: 4

5. Repeat this process for any additional vVol data stores (Figure 4-42). When the process is finished, the data stores are ready for use.

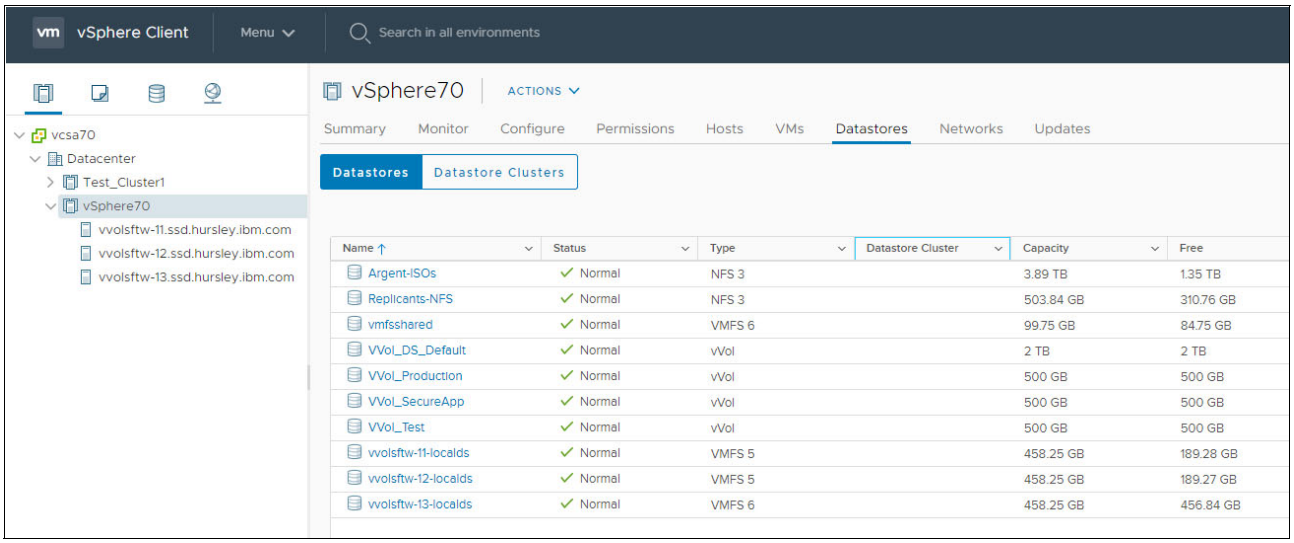


Figure 4-42 Repeating the process for any additional vVol data stores

4.4 Best-practice considerations and troubleshooting

This section describes best-practice considerations and troubleshooting tips.

4.4.1 Protecting the IBM Spectrum Connect server

The IBM Spectrum Connect server is the key component in facilitating integration between multiple VMware platforms and IBM Spectrum Virtualize storage systems.

Therefore, it is a best practice to protect the IBM Spectrum Connect server to enable optimal functions:

- Where possible, ensure that VMware vSphere Fault Tolerance (FT) is used to ensure that if there is an outage to the ESXi host that is running the IBM Spectrum Connect server, then the wider infrastructure is still able to access the IBM Spectrum Connect integrations. If FT is not available, then ensure that vSphere HA is used to minimize downtime of the server hosting the IBM Spectrum Connect application.
- Allocate appropriate compute resources to the IBM Spectrum Connect VM. Because the applications run as a conduit between the multiple VMware interfaces and multiple storage systems, performance of the end-to-end system is impacted if resources are limited.
- Most importantly, ensure that all VMs associated with providing the vVol infrastructure are not stored on a VMware vSphere Virtual Volume data store, which include the vCenter Service Appliance (vCSA), the IBM Spectrum Connect server, and other servers that these applications require to function.

  If an outage occurs, the IBM Spectrum Connect server must start before any vVol operations function.

  If the IBM Spectrum Connect server depends on an external LDAP or NFS server that is on a vVol data store, then it will fail to successfully start IBM Spectrum Connect services, which means that vVol data stores will be inaccessible and VMs on the vVol data store will be unavailable. If this situation occurs, contact IBM Support.

4.4.2 Viewing the relationships between vVol and VM

Since the vVols are created with a unique identifier, it can be difficult to establish to which VM a vVol belongs and to which vVols a VM is associated.

By creating the association between the vVol-enable storage service and the vCenter interface in IBM Spectrum Connect, the IBM Storage Enhancements plug-in can assimilate information from both systems.

The relationships between vVols and VMs can then be displayed for the vSphere administrator.

To access display of vVol and VM relationships, complete the following steps:

1. Locate the top-level vCenter object in the vSphere Client inventory, and select the **More Objects** tab (Figure 4-43).

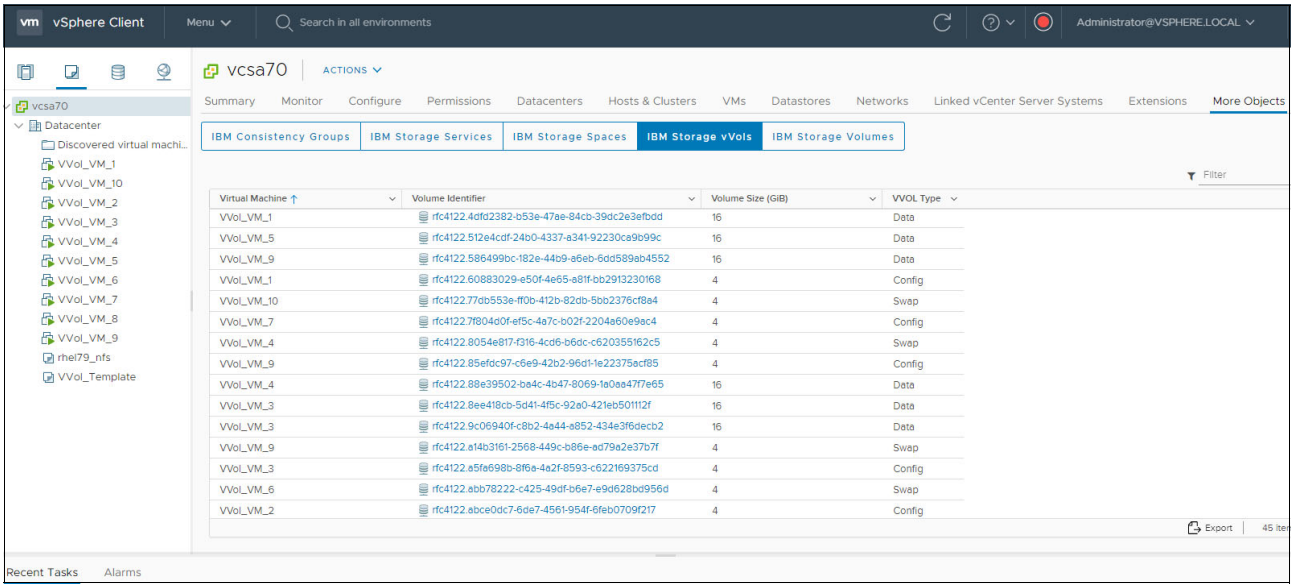


Figure 4-43 Selecting IBM Storage vVols

2. Click **IBM Storage vVols** to list the detected vVols on the storage array and the vVols size and type.
3. If required, export the list as a CSV file to generate a report.

   For more information about the IBM Storage Enhancements plug-in, see 4.5, “IBM Storage Enhancements for VMware vSphere Web Client” on page 82.

4.4.3 Mirroring the utility volume

A utility volume is created on the storage system as part of the configuration of vVols. IBM Spectrum Connect creates and manages a database on the utility volume. The VM metadata that is stored in this database is critical to both IBM Spectrum Connect operations and the vVols environment.

Because the availability and integrity of the utility volume is fundamental to the vVols environment, store a mirrored copy of the volume for redundancy. If possible, store the mirrored copy in a second storage pool that is in a separate failure domain. For example, use a storage pool that is created with managed disks (MDisks) that are from different storage systems or separate I/O groups.

4.4.4 Performing an upgrade on a storage system with vVols enabled

When performing a code upgrade, the storage system intentionally limits the tasks that a user can run both by using the GUI and CLI. This measure is a protective one to ensure that the upgrade is not disrupted. Because managing VMs on vVols data stores requires the running of CLI commands on the system, the same restrictions apply in the vVols environment.

Therefore, VM-management tasks (for example, powering off a VM) fail when an upgrade is running on the storage system. Automated services, such as VMware HA and Distributed Resource Scheduler (DRS), also are affected because they send system commands by using IBM Spectrum Connect.

Therefore, it is important to plan an upgrade with your vSphere administrator. To ensure a smooth upgrade in your vVols environment, consider the following suggestions:

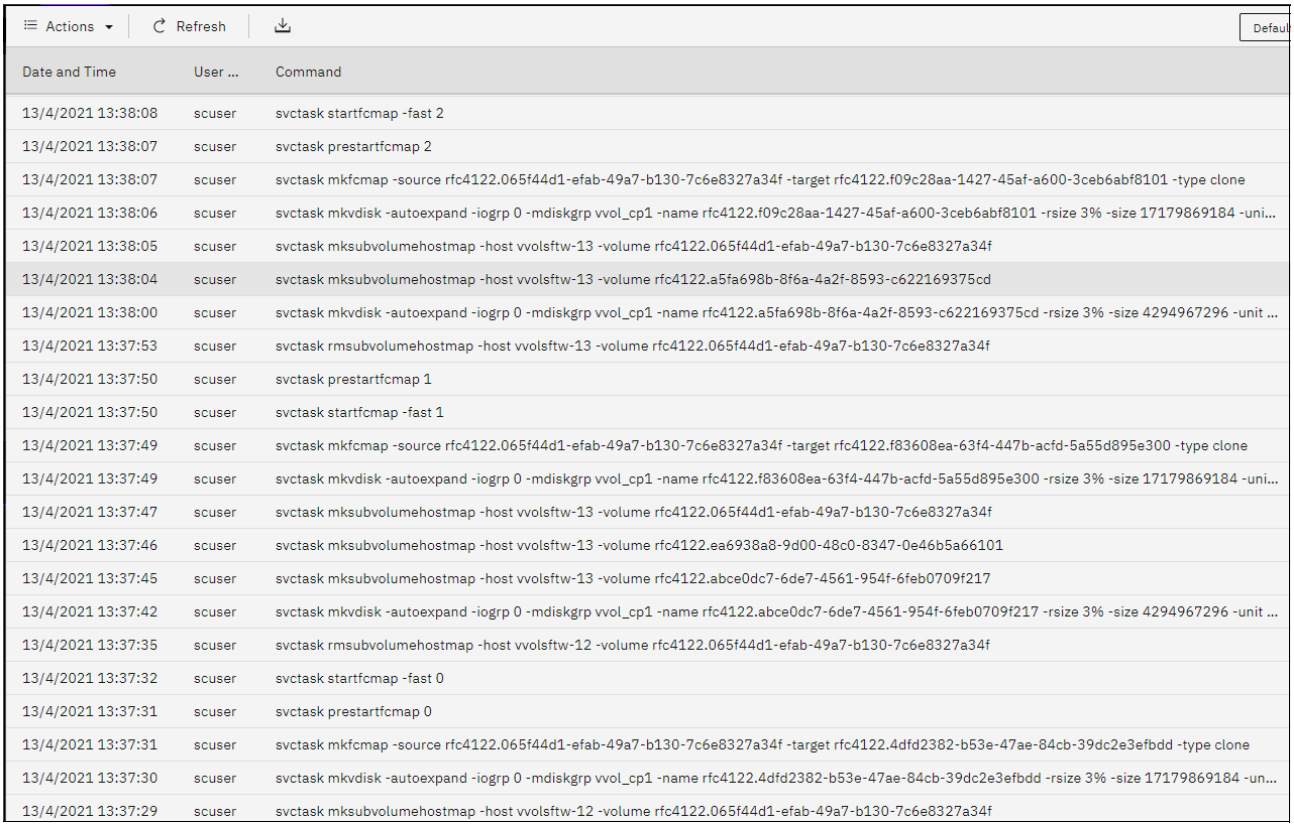
- Plan your upgrade for a time when VM management is not required.
- Warn your vSphere administrator not to run management tasks on VMs stored on vVols data stores during an upgrade because it results in task failures on the vSphere Client.
- Be aware of automated backup schedules that use VM snapshots.

Tip: After you perform an upgrade, it is possible that the vSphere Web Client will mark cold VMs as *inaccessible*, which means that ESXi hosts were unable to start a new binding to these VMs (expected during a code upgrade) and should not cause alarm.

To recover management of these VMs, the vSphere administrator should remove the affected VMs from the inventory and then add them again.

4.4.5 Understanding audit log entries

For illustration purposes, we created a VM template on a vVol data store, and created 10 VMs from this template. Figure 4-44 on page 77 shows a section of the audit log of the storage system where you can see the CLI commands that are issued from the IBM Spectrum Connect server.



Actions | Refresh | Default

Date and Time	User ...	Command
13/4/2021 13:38:08	scuser	svctask startfcmap -fast 2
13/4/2021 13:38:07	scuser	svctask prestartfcmap 2
13/4/2021 13:38:07	scuser	svctask mkfcmap -source rfc4122.065f44d1-efab-49a7-b130-7c6e8327a34f -target rfc4122.f09c28aa-1427-45af-a600-3ceb6abf8101 -type clone
13/4/2021 13:38:06	scuser	svctask mkvdisk -autoexpand -iogrp 0 -mdiskgrp vvol_cp1 -name rfc4122.f09c28aa-1427-45af-a600-3ceb6abf8101 -rsize 3% -size 17179869184 -uni...
13/4/2021 13:38:05	scuser	svctask mksubvolumehostmap -host vvolsftw-13 -volume rfc4122.065f44d1-efab-49a7-b130-7c6e8327a34f
13/4/2021 13:38:04	scuser	svctask mksubvolumehostmap -host vvolsftw-13 -volume rfc4122.a5fa698b-8f6a-4a2f-8593-c622169375cd
13/4/2021 13:38:00	scuser	svctask mkvdisk -autoexpand -iogrp 0 -mdiskgrp vvol_cp1 -name rfc4122.a5fa698b-8f6a-4a2f-8593-c622169375cd -rsize 3% -size 4294967296 -unit ...
13/4/2021 13:37:53	scuser	svctask rmsubvolumehostmap -host vvolsftw-13 -volume rfc4122.065f44d1-efab-49a7-b130-7c6e8327a34f
13/4/2021 13:37:50	scuser	svctask prestartfcmap 1
13/4/2021 13:37:50	scuser	svctask startfcmap -fast 1
13/4/2021 13:37:49	scuser	svctask mkfcmap -source rfc4122.065f44d1-efab-49a7-b130-7c6e8327a34f -target rfc4122.f83608ea-63f4-447b-acfd-5a55d895e300 -type clone
13/4/2021 13:37:49	scuser	svctask mkvdisk -autoexpand -iogrp 0 -mdiskgrp vvol_cp1 -name rfc4122.f83608ea-63f4-447b-acfd-5a55d895e300 -rsize 3% -size 17179869184 -uni...
13/4/2021 13:37:47	scuser	svctask mksubvolumehostmap -host vvolsftw-13 -volume rfc4122.065f44d1-efab-49a7-b130-7c6e8327a34f
13/4/2021 13:37:46	scuser	svctask mksubvolumehostmap -host vvolsftw-13 -volume rfc4122.ea6938a8-9d00-48c0-8347-0e46b5a66101
13/4/2021 13:37:45	scuser	svctask mksubvolumehostmap -host vvolsftw-13 -volume rfc4122.abce0dc7-6de7-4561-954f-6feb0709f217
13/4/2021 13:37:42	scuser	svctask mkvdisk -autoexpand -iogrp 0 -mdiskgrp vvol_cp1 -name rfc4122.abce0dc7-6de7-4561-954f-6feb0709f217 -rsize 3% -size 4294967296 -unit ...
13/4/2021 13:37:35	scuser	svctask rmsubvolumehostmap -host vvolsftw-12 -volume rfc4122.065f44d1-efab-49a7-b130-7c6e8327a34f
13/4/2021 13:37:32	scuser	svctask startfcmap -fast 0
13/4/2021 13:37:31	scuser	svctask prestartfcmap 0
13/4/2021 13:37:31	scuser	svctask mkfcmap -source rfc4122.065f44d1-efab-49a7-b130-7c6e8327a34f -target rfc4122.4dfd2382-b53e-47ae-84cb-39dc2e3efbdd -type clone
13/4/2021 13:37:30	scuser	svctask mkvdisk -autoexpand -iogrp 0 -mdiskgrp vvol_cp1 -name rfc4122.4dfd2382-b53e-47ae-84cb-39dc2e3efbdd -rsize 3% -size 17179869184 -un...
13/4/2021 13:37:29	scuser	svctask mksubvolumehostmap -host vvolsftw-12 -volume rfc4122.065f44d1-efab-49a7-b130-7c6e8327a34f

Figure 4-44 Audit log entries

In Figure 4-44, the following CLI commands are used:

- **`mkvdisk`**: This command is used to create the vVols in the allocated storage pool. vVols will be created with a unique identifier (compliant with Request For Comment (RFC) 4122) to distinguish themselves from traditional volumes. A volume is created for each VMDK on the vVol data store.

  If VM-level snapshots are created for any VMDK on the vVol data store, more **`mkvdisk`** commands are issued to create associated volumes on the storage array. Also, if the VM Home folder is stored on a vVol data store, then more vVols are required to store the VM configuration files or logs. When a VM is powered on, an extra volume is created and dedicated to memory swap. This volume is deleted when a VM is powered off.

- **`mksubvolumehostmap`** and **`rmsubvolumehostmap`**: These commands are used to map (and remove the map when no longer required) the vVols to a specific ESXi host. These commands differ from the more traditional **`mkvdiskhostmap`** and **`rmvdiskhostmap`** commands, which are used to map traditional SCSI devices, and require an HBA rescan to detect. The **`subvolumehostmap`** commands are mapped by using the PE and provide a more dynamic environment with vVols being mapped and removed more frequently.

  ESXi periodically requires access to the VM metadata, so you might observe these commands being used frequently, even when active administration or management tasks are not being performed. This situation is normal and can be ignored.

- **`mkfcmap`**, **`prestartfcmap`**, and **`startfcmap`**: FlashCopy is used mainly by vVols to offload VM-level snapshots to the storage system. However, it is also used when you perform storage migrations by using Storage vMotion or when you perform VM-clone operations of a powered-off VM. When a VM is powered on, ESXi uses its own data-copy operator.

4.4.6 IBM Spectrum Virtualize GUI

It is not expected that a storage administrator will need to interact with vVols individually, but rather rely on the vSphere administrator's actions within vCenter to facilitate management tasks. Therefore, the vVols are not visible through the IBM Spectrum Virtualize GUI. Instead, they are only visible to the storage administrator by using the CLI over SSH. The Utility Volume (or **metadatavdisk**) is visible in both the GUI and CLI.

To prevent an accidental action in the IBM Spectrum Virtualize from impact the IBM Spectrum Connect objects, the storage system prevents manipulation of these objects unless the manipulation is performed by a user with the VASA-Provider role. When a “superuser” or similar account is used to change objects that IBM Spectrum Connect created, the following error is reported:

```
CMMVC8652E The command failed as a volume is owned and has restricted use.
```

In some circumstances, there might be a requirement to make manual changes to these objects. Therefore, the storage administrator is advised to log in to the Storage System with a user account that is assigned with the VASA Provider role.

Warning: Do not make manual changes unless advised by a member of IBM Support because it might cause more issues.

4.4.7 Metadata VDisk

The metadata VDisk (or Utility Volume) is a dedicated volume within the IBM Spectrum Virtualize storage system that is mounted locally onto the configuration node of the cluster. This volume stores the metadata database that is required to maintain awareness of the individual components of the vVols environment.

When configured in an HA configuration, both Active and Standby IBM Spectrum Connect servers use the same database to ensure infrastructure consistency.

Tip: For more information about the HA configuration, see the *IBM Spectrum Connect User Guide* available at IBM Fix Central.

If the IBM Spectrum Connect server is permanently unavailable (and no backup exists), a new IBM Spectrum Connect server can be commissioned to recover the metadata database on the metadata VDisk, and the environment can be resurrected. For more information about this recovery process, contact IBM Support.

To query the active state of the metadata VDisk, connect to the Storage System by using the CLI management interface and run the following command (Example 4-2).

Example 4-2 The svcinfo lsmetadatavdisk command

```
IBM_FlashSystem:FS9200-CL:superuser>svcinfo lsmetadatavdisk
vdisk_id 16
vdisk_name vdisk0
status online
```

If the status of the **metadatavdisk** reports `corrupt`, verify that the underlying VDisk is in an operational state by checking the detailed view of the specific vdisk_id (vdisk_id **16** in Example 4-3 on page 79). Verify that the underlying VDisk is in an operational state.

Example 4-3 Verifyin gthat the underlying VDisk is in an operational state

```
IBM_2145:vvolsftw-sv1:superuser> lsvdisk 16 | grep ^status
status online
status online
```

In rare circumstances, the output from the **lsmetadatavdisk** command shows `corrupt`, and there are messages in the IBM Spectrum Connect `hsgsvr.log` file reporting the following error:

```
CMMVC8580E The action failed because the metadata utility volume is corrupt.
```

This error might occur if the configuration node experienced issues when attempting to access the internal mounted volume. To resolve this issue, put the configuration node into service state by using the service assistant. After a few minutes, bring the node back in to the cluster, and retry the **lsmetadatavdisk** command again. The metadata VDisk should now report online status.

Note: If you are unsure of the status of the metadata VDisk, contact IBM Support.

4.4.8 Enabling debugging in IBM Spectrum Connect

By default, the logging in IBM Spectrum Connect is limited to informational events and does not always provide the user with the necessary information that is required to debug specific issues. More verbose logging is available when required.

To enable the debug logging, complete the following steps:

1. Log in to the IBM Spectrum Connect server by using SSH or console.
2. Locate the `hsgsvr.ini` file located in the `/opt/ibm/ibm_spectrum_connect/conf.d/hsgsvr` directory.
3. Change the `debug=False` setting to `debug=True`.
4. Restart IBM Spectrum Connect Services by running the following command:

   ```
   systemctl restart ibm_spectrum_connect
   ```

5. Wait a few minutes for the services to restart. Additional debug logging should appear in the `/var/log/sc/hsgsvr.log` file.

4.4.9 Certificates

Certificates are used by vSphere, IBM Spectrum Connect, and IBM Spectrum Virtualize to secure communication between the separate components. Therefore, ensure that the IBM Spectrum Connect certificate is configured correctly.

A common symptom of certificate issues is where the Storage Provider is registered successfully in vCenter, but the ESXi hosts might report errors in `vvold.log` citing errors in communicating with the IBM Spectrum Connect server because the hostname does not match the names that are defined in the certificate (Example 4-4).

Example 4-4 The vvold.log file

```
2021-04-13T12:06:20.483Z error vvold[1053651] [Originator@6876 sub=Default] Initialize: Unable
to init session to VP vVols_SpecConnect state: 0
2021-04-13T12:06:25.487Z info vvold[1053615] [Originator@6876 sub=Default]
VasaSession::GetEndPoint: with url https://vvolsftw-scb.ssd.hursley.ibm.com:8440/services/vasa
2021-04-13T12:06:25.490Z warning vvold[1053615] [Originator@6876 sub=Default]
VasaSession::GetEndPoint: failed to get endpoint, err=SSL Exception: Verification parameters:
--> PeerThumbprint: A1:61:6B:C9:11:72:DF:0B:5D:BA:9D:3B:C2:49:1E:FB:3B:64:84:9D
--> ExpectedThumbprint:
--> ExpectedPeerName: vvolsftw-scb.ssd.hursley.ibm.com
--> The remote host certificate has these problems:
-->
--> * Hostname does not match the subject names in certificate, using default
```

If this certificate-issue occurs, regenerate the certificate in IBM Spectrum Connect with the correct common name and fully qualified domain name, as follows:

1. Go to the IBM Spectrum Connect management interface window and click **Server Certificate** (Figure 4-45).

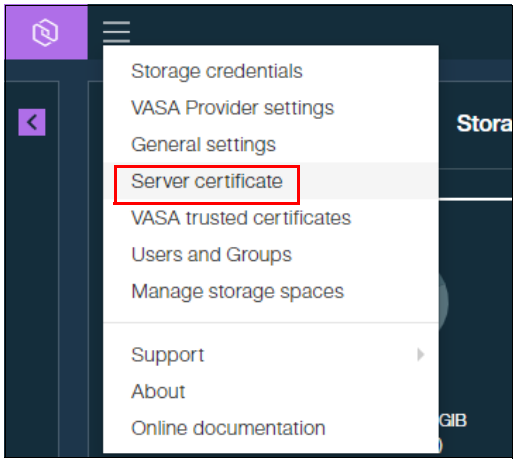


Figure 4-45 Server Certificate: 1

2. In the Server Certificate window, click **Generate** (Figure 4-46 on page 81).

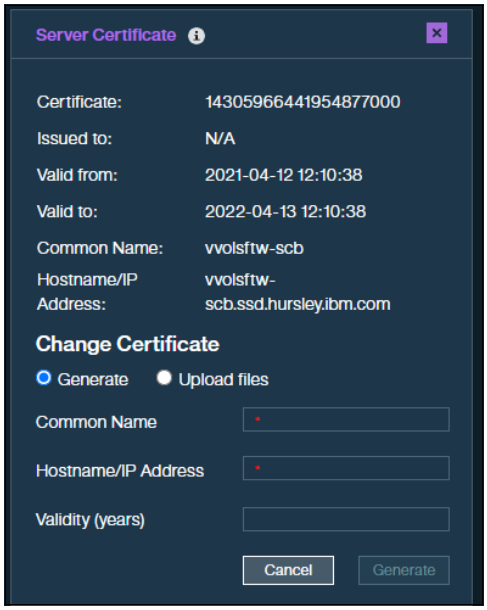


Figure 4-46 Server Certificate: 2

3. After the certificate regenerates, you must remove and re-register the Storage Provider.

4.4.10 Prerequisites, limitations, and restrictions

This section describes the prerequisites, limitations, and restrictions.

Prerequisites

The following items are the prerequisites:

- NTP: NTP client must be configured on the base Linux OS under the IBM Spectrum Connect application, the IBM Spectrum Virtualize storage system, and vSphere (vCenter and ESXi hosts). Given the multiple components in the end-to-end infrastructure, any time-skew between IBM Spectrum Connect, IBM Spectrum Virtualize, and the VMware platforms can complicate debugging issues when logs are reviewed.
- Supported versions: Check IBM Documentation for the interoperability matrix for supported versions of IBM Spectrum Connect, IBM Spectrum Virtualize, and vSphere. For IBM Spectrum Connect V3.7, which was the latest version at the time of writing, see Compatibility and requirements.

Restrictions

The following are the restrictions:

- Array-based replication for vVol is not currently supported by IBM Spectrum Connect or IBM Spectrum Virtualize. This items is planned for a future release.
- vVols are not supported in a HyperSwap configuration.

Limitations

The following are the limitations:

- At the time of writing, The number of vVols in an IBM Spectrum Virtualize storage system is limited to 10,000 per IBM Spectrum Virtualize cluster. Depending on the scale of the vSphere environment, the number of VMsmight not be suitable for a vVol implementation, especially given the number of volumes that can be consumed by a single VM.

  With traditional VMFS data stores, a single LUN could host hundreds or even thousands of VMs, and the best-practice guidance on the number of VMs per data store is more focused on the workload being generated, rather than the number of VMs.

  In a vVol environment, a single VM requires a minimum of three volumes:

  - Configuration vVol
  - Swap vVol
  - Data vVol

  Note: For more information about the types of vVols, see Virtual Volume Objects.

- If more VMDKs are configured on the VM, then associated more vVols are created. If a VM-level snapshot is taken of the VM, then an additional vVol is created for each VMDK configured on the VM.

  Note: A conservative assumption would be to assume that a single VM could require 10 vVols, and so consider the scale of the VMware environment being used to evaluate whether a vVol solution is suitable.

For the specific version of IBM Spectrum Virtualize, see the “Configuration Limits and Restrictions for IBM Spectrum Virtualize” page in IBM Documentation. For example, see V8.4.0.x Configuration Limits and Restrictions for IBM FlashSystem 9200.

4.5 IBM Storage Enhancements for VMware vSphere Web Client

The *IBM Storage Enhancements for VMware vSphere* plug-in integrates directly into the vSphere Client and enables VMware administrators the ability to provision storage resources to hosts or clusters that are registered within vCenter.

The storage administrator can use the Storage Spaces and Storage Services objects in IBM Spectrum Connect to complete the following actions:

- Create a preset of specific storage capabilities.
- Allocate pools of storage capacity (either parent pools or child pools) in which volumes that are created by using the IBM Storage Enhancements vSphere plug-in will be located.
- Delegate those presets to vCenter so they can be consumed by a vSphere administrator as either VMFS data stores or RDMs.

The IBM Storage Enhancements for VMware vSphere Web Client plug-in is automatically deployed and enabled for each vCenter server that is registered in the Interfaces window of IBM Spectrum Connect.

4.5.1 Installating the vSphere plug-in

Before you can use the IBM Storage Enhancements for VMware vSphere plug-in on the vCenter client side, you need to define, in IBM Spectrum Connect, the vCenter servers for which you want to provide storage resources. Then, you can attach storage services that you want to make available to each vCenter server.

The storage services that you attach on the IBM Spectrum Connect side become visible on vSphere Client, and can be used for volume creation by using the IBM Storage Enhancements for vSphere Client.

Before you begin, log out of vSphere Client browser windows on the vCenter server to which you want to add IBM Spectrum Connect. If you stay logged in, you will be able to use the extension only after you log out and log in to vCenter after IBM Spectrum Connect is added?

You need to add the vCenter servers for which you can later attach storage services that would be visible and accessible on the vSphere Client side. You can add a single vCenter server at a time.

When entering the vCenter credentials on the IBM Spectrum Connect side, verify that the vCenter user has sufficient access level in vCenter to complete this procedure.

To add a vCenter server, complete the following steps:

1. In the IBM Spectrum Connect UI, click **Add Interface** on the Interfaces window and then select **Add vCenter**. The Add New vCenter Server for vWC window opend (Figure 4-47).

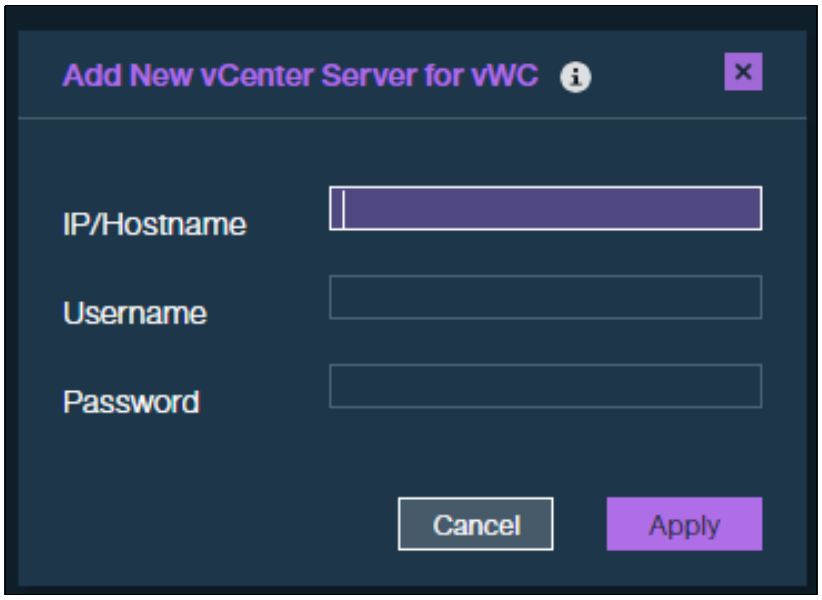


Figure 4-47 Add New vCenter Server for vWC

2. Enter the IP address or fully qualified domain name (FQDN) of the vCenter server, and the username and password for logging in to that vCenter server (Figure 4-48).

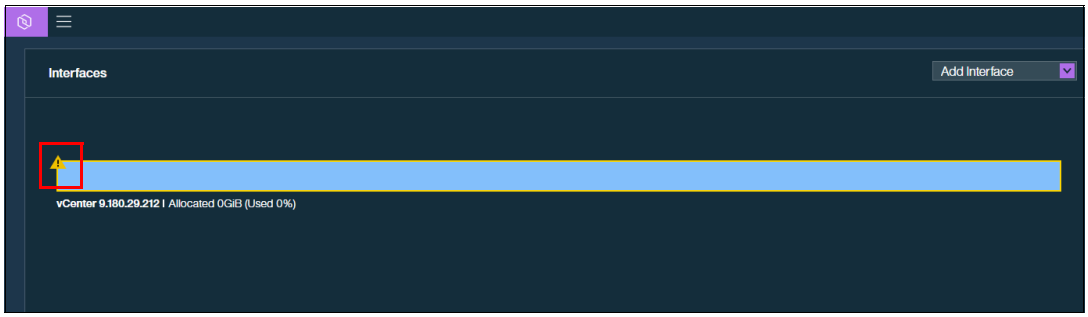


Figure 4-48 vCenter interface was added to IBM Spectrum Connect

If the provided IP address and credentials are accepted by the vCenter server, it is added to the list of servers on the Interfaces window. The yellow frame and the exclamation mark in Figure 4-48 indicate that storage services are not yet delegated to interface.

Notes:

- If you want to use the vSphere Web Client extension on all vCenter servers that operate in linked mode, each server instance must be added to IBM Spectrum Connect, which ensures that the extension is registered on all linked servers properly.
- The same vCenter server cannot be added to more than one IBM Spectrum Connect instance. Any attempt to add an already registered vCenter server to another IBM Spectrum Connect overrides the primary connection.

3. If you need to update the vCenter credentials that are being used by IBM Spectrum Connect or to remove the vCenter interface, right-click the **vCenter** as displayed in the Interfaces window, and click either **Modify** or **Remove** (Figure 4-49).

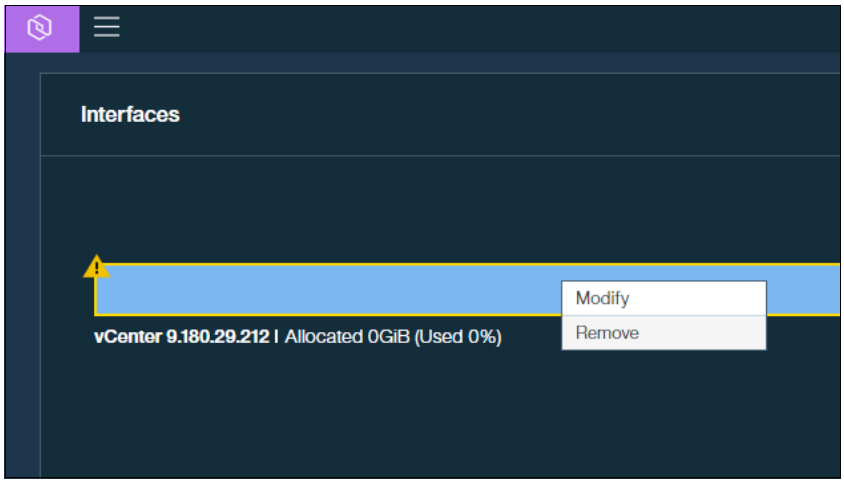


Figure 4-49 Updating the vCenter credentials

4.5.2 Provisioning storage from within vCenter

To enable the provisioning of storage from the vSphere Client, you must first create a Storage Service with an associated storage resource. This Storage Service can then be delegated to the vCenter interface (as described in 4.2.3, “Delegating Storage Services to vCenter” on page 57). Volumes that are created from within the vSphere Client will be created in the associated pool on the IBM Spectrum Virtualize storage system.

4.5.3 Using the IBM Storage Enhancements vSphere plug-in

When the vCenter interface is registered in IBM Spectrum Connect, the IBM Storage plug-in is automatically installed into vCenter. This plug-in enables the extra UI functions to interact with the IBM Spectrum Virtualize storage system and perform storage provisioning operations.

To provision volumes by using the IBM Storage Enhancements plug-in, complete the following steps:

1. In the vSphere Client, select the **Hosts & Clusters** tab. Right-click the host or cluster to which you want to provision volumes and select **IBM Storage** → **Create new IBM Storage volume** (Figure 4-50).

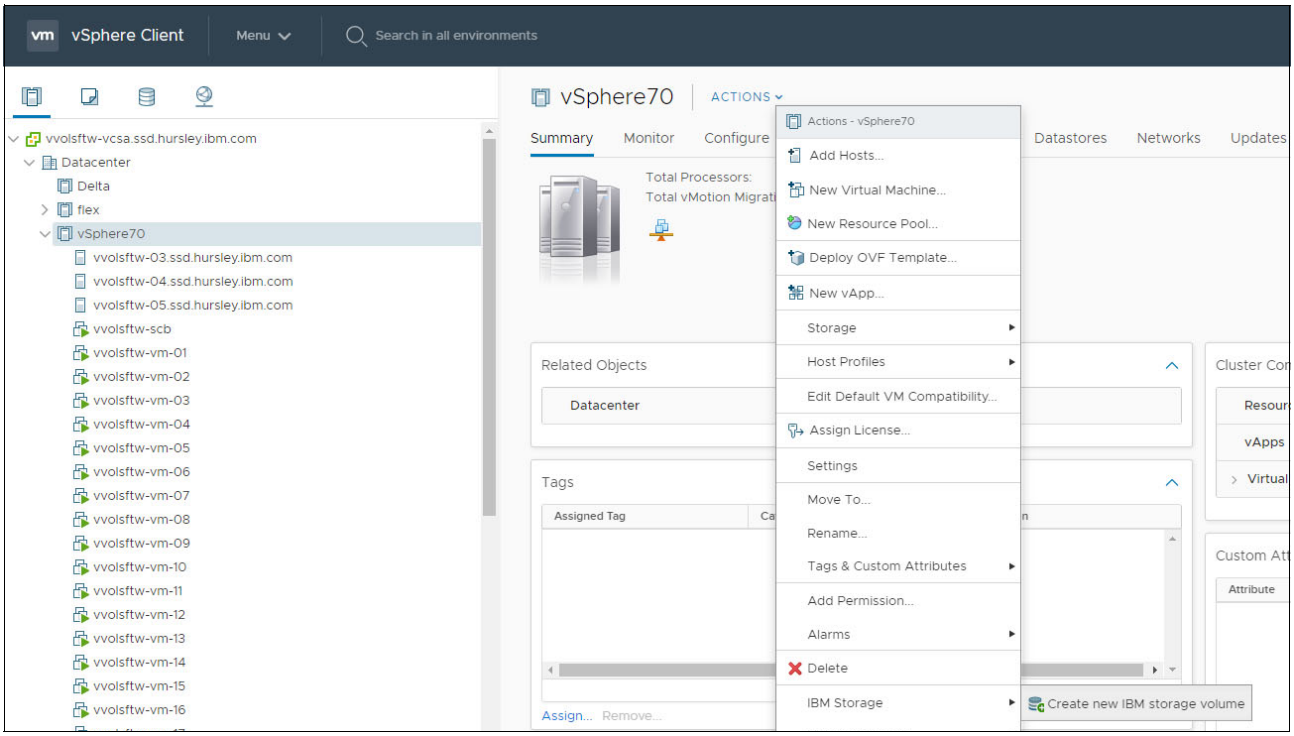


Figure 4-50 Creating an IBM Storage volume

The New Volume window is displayed (Figure 4-51).

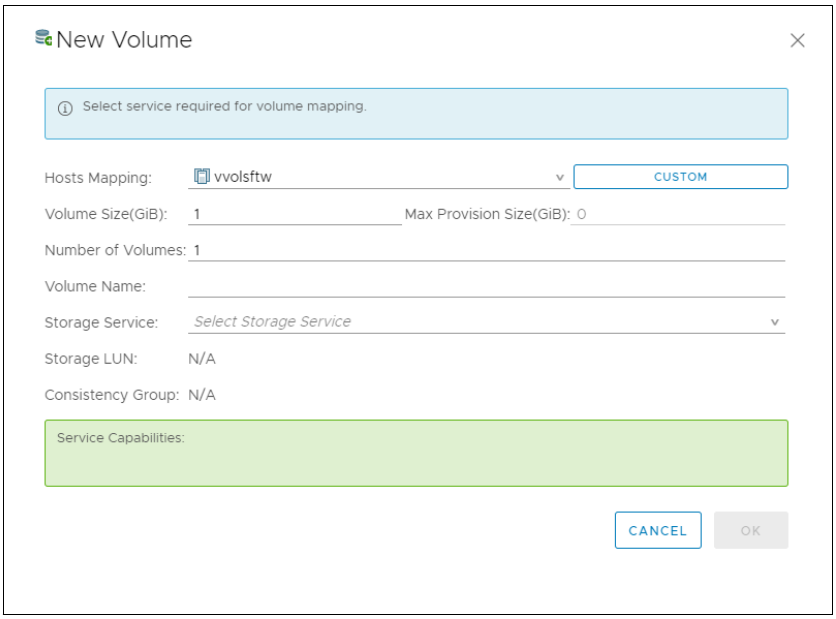


Figure 4-51 New Volume window

2. In the Hosts Mapping field, click the arrow and select the hosts to which you want to map the volumes. If you select a vSphere Cluster from the list, the volumes are mapped by using the Host Cluster feature in IBM Spectrum Virtualize (see 2.1.1, “IBM FlashSystem host clusters” on page 8). This mapping ensures that if more hosts are added to the Host Cluster on the IBM Spectrum Virtualize system, they automatically inherit existing Host Cluster mappings.
3. If a custom-volume mapping is required, click **Custom**, and select the boxes for each ESXi host or cluster to which you want to map the volumes and click **OK** (Figure 4-52).

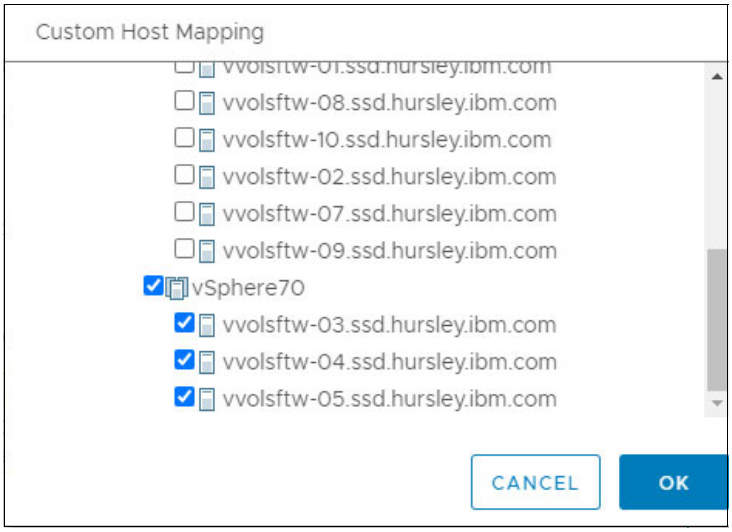


Figure 4-52 Custom Host Mapping

4. Enter the required size, quantity, and name for the volumes to be created. When you create multiple volumes simultaneously, the text box next to the Volume Name entry displays `vol_{1}` by default. The number between the brackets ({ }) is incremented for each volume being created (Figure 4-53).

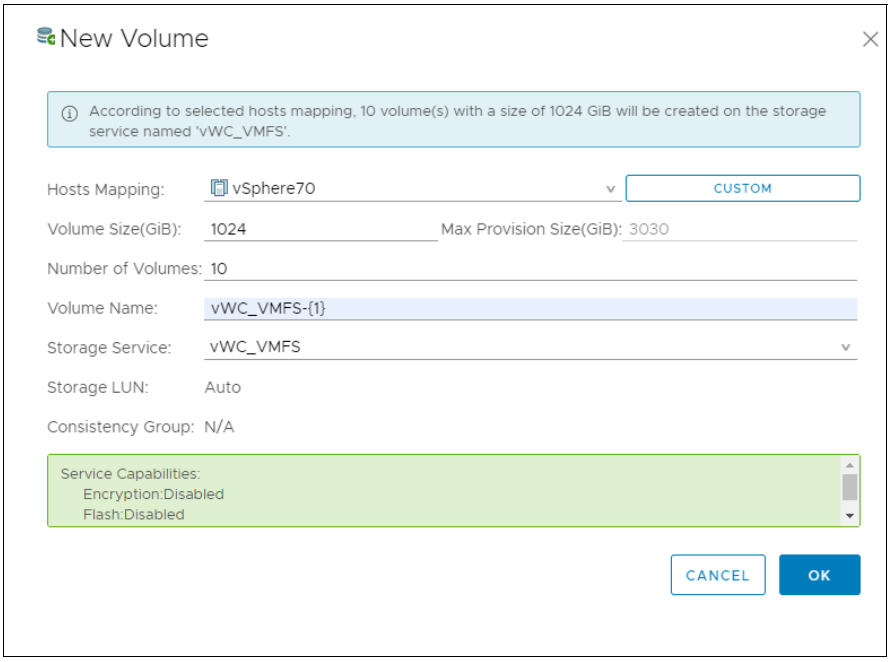


Figure 4-53 New Volume

5. Select the Storage Service in which you want to create the volumes. If multiple Storage Services exist, they are included in the list.
   - Storage capabilities that were defined on the Storage Service are listed in the green area of the window.
   - A summary of the task is shown in the blue area of the window.
6. The Storage LUN value defines the SCSI ID to be used when mapping the volume to the host or host cluster. Unless for a specific requirement, select **Auto** for the Storage LUN value so that the SCSI ID can be automatically determined by the system.
7. Click **OK** to begin the storage provisioning task.

   The status of the operation is displayed in the Recent Tasks tab of the vSphere Client window (Figure 4-54).

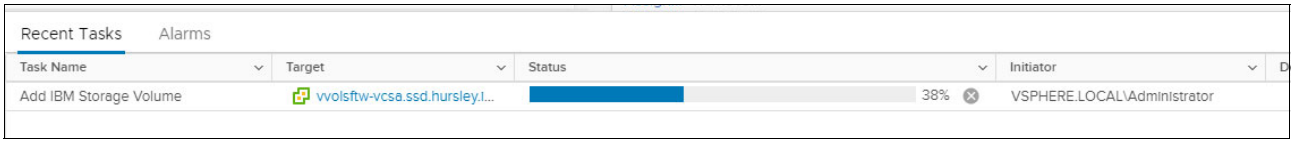


Figure 4-54 Status of the operation

Hosts that were involved in the storage-provisioning operation automatically scan for storage changes. Also, the display names for the volumes also reflect the names that are defined in the previous step (Figure 4-55).

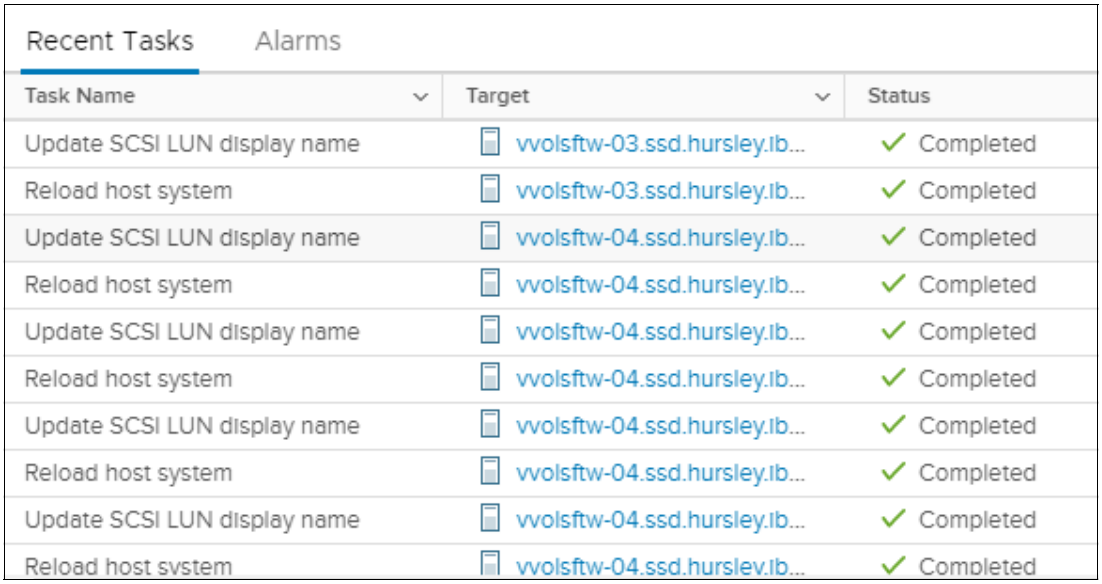
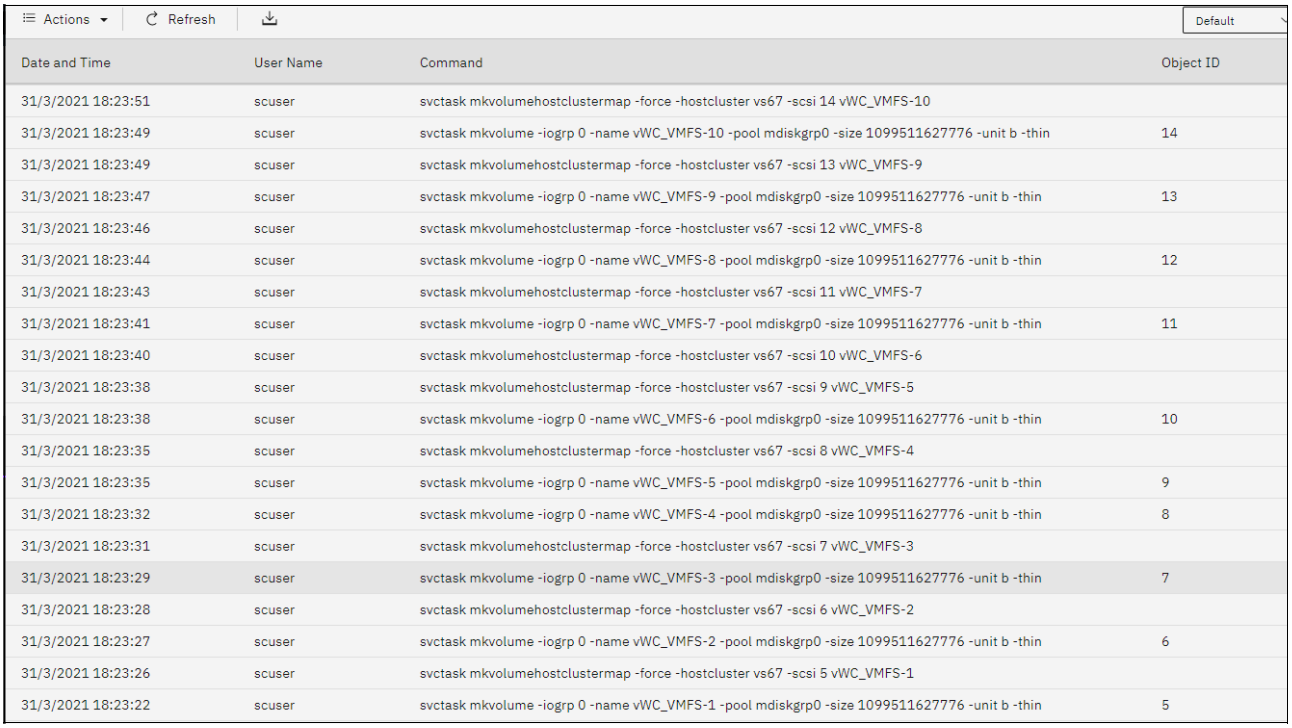


Figure 4-55 Recent Tasks

8. Verify that the commands were issued correctly by checking the Audit Log (Figure 4-56) in the IBM Spectrum Virtualize storage system. To access the Audit Log, log in to the web interface of the Storage System and select **Access** → **Audit Log**. You can also use the CLI to run the **`cataudlog`** command when you are logged in using SSH.



Date and Time	User Name	Command	Object ID
31/3/2021 18:23:51	scuser	svctask mkvolumehostclustermap -force -hostcluster vs67 -scsi 14 vWC_VMFS-10	
31/3/2021 18:23:49	scuser	svctask mkvolume -iogrp 0 -name vWC_VMFS-10 -pool mdiskgrp0 -size 1099511627776 -unit b -thin	14
31/3/2021 18:23:49	scuser	svctask mkvolumehostclustermap -force -hostcluster vs67 -scsi 13 vWC_VMFS-9	
31/3/2021 18:23:47	scuser	svctask mkvolume -iogrp 0 -name vWC_VMFS-9 -pool mdiskgrp0 -size 1099511627776 -unit b -thin	13
31/3/2021 18:23:46	scuser	svctask mkvolumehostclustermap -force -hostcluster vs67 -scsi 12 vWC_VMFS-8	
31/3/2021 18:23:44	scuser	svctask mkvolume -iogrp 0 -name vWC_VMFS-8 -pool mdiskgrp0 -size 1099511627776 -unit b -thin	12
31/3/2021 18:23:43	scuser	svctask mkvolumehostclustermap -force -hostcluster vs67 -scsi 11 vWC_VMFS-7	
31/3/2021 18:23:41	scuser	svctask mkvolume -iogrp 0 -name vWC_VMFS-7 -pool mdiskgrp0 -size 1099511627776 -unit b -thin	11
31/3/2021 18:23:40	scuser	svctask mkvolumehostclustermap -force -hostcluster vs67 -scsi 10 vWC_VMFS-6	
31/3/2021 18:23:38	scuser	svctask mkvolumehostclustermap -force -hostcluster vs67 -scsi 9 vWC_VMFS-5	
31/3/2021 18:23:38	scuser	svctask mkvolume -iogrp 0 -name vWC_VMFS-6 -pool mdiskgrp0 -size 1099511627776 -unit b -thin	10
31/3/2021 18:23:35	scuser	svctask mkvolumehostclustermap -force -hostcluster vs67 -scsi 8 vWC_VMFS-4	
31/3/2021 18:23:35	scuser	svctask mkvolume -iogrp 0 -name vWC_VMFS-5 -pool mdiskgrp0 -size 1099511627776 -unit b -thin	9
31/3/2021 18:23:32	scuser	svctask mkvolume -iogrp 0 -name vWC_VMFS-4 -pool mdiskgrp0 -size 1099511627776 -unit b -thin	8
31/3/2021 18:23:31	scuser	svctask mkvolumehostclustermap -force -hostcluster vs67 -scsi 7 vWC_VMFS-3	
31/3/2021 18:23:29	scuser	svctask mkvolume -iogrp 0 -name vWC_VMFS-3 -pool mdiskgrp0 -size 1099511627776 -unit b -thin	7
31/3/2021 18:23:28	scuser	svctask mkvolumehostclustermap -force -hostcluster vs67 -scsi 6 vWC_VMFS-2	
31/3/2021 18:23:27	scuser	svctask mkvolume -iogrp 0 -name vWC_VMFS-2 -pool mdiskgrp0 -size 1099511627776 -unit b -thin	6
31/3/2021 18:23:26	scuser	svctask mkvolumehostclustermap -force -hostcluster vs67 -scsi 5 vWC_VMFS-1	
31/3/2021 18:23:22	scuser	svctask mkvolume -iogrp 0 -name vWC_VMFS-1 -pool mdiskgrp0 -size 1099511627776 -unit b -thin	5

Figure 4-56 Audit Log

The names of the created volumes were carried through from the previous New Volume step (Figure 4-57).

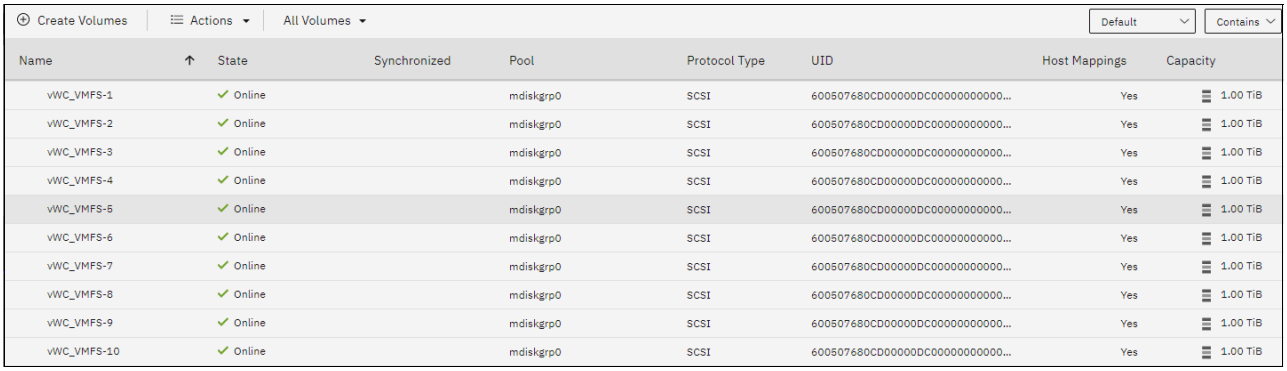


Figure 4-57 Notice the names of the created volumes

The volumes are created and mapped to the selected Hosts or Clusters, and the vSphere administrator is now able to create VMFS data stores or RDMs from these volumes by using the normal vSphere workflow (Figure 4-58).

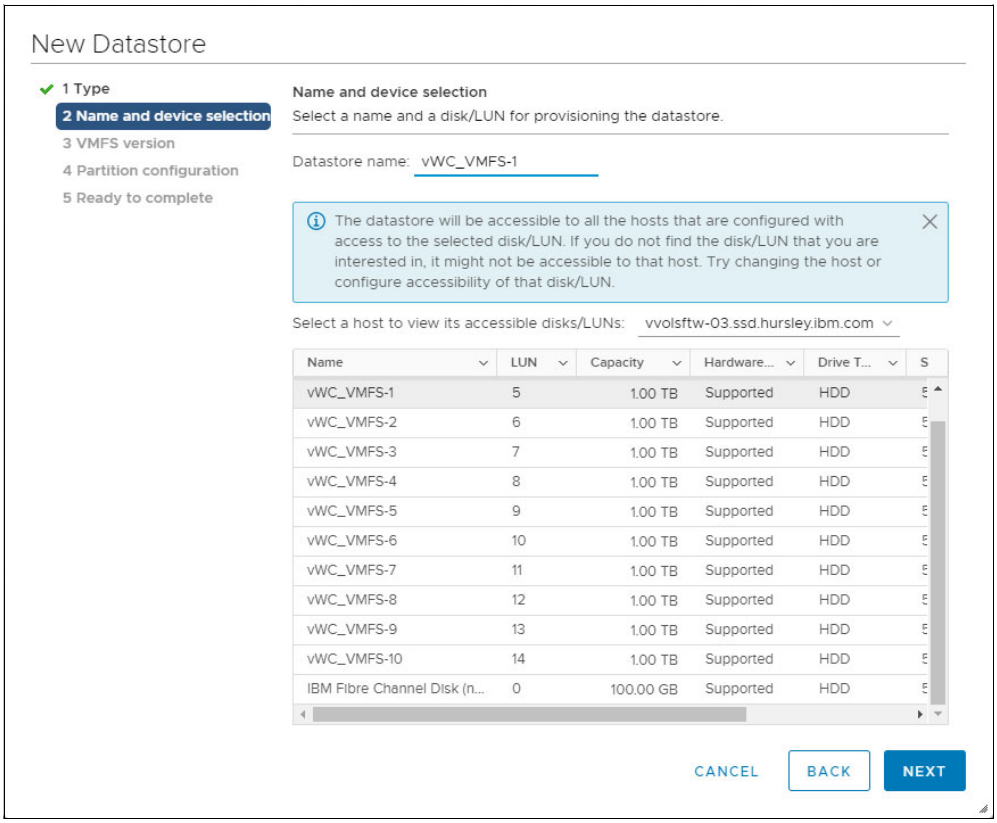


Figure 4-58 New Datastore

4.5.4 Viewing more storage information from within the vSphere Client

When logged in to the vSphere Client, the vSphere administrator can use the IBM Storage Enhancements plug-in to view additional information about the storage objects that were delegated to the vCenter Interface.

For each vCenter server, the following IBM Storage categories are available to view for IBM Spectrum Virtualize platforms:

- Storage services
- Storage spaces
- Storage volumes
- Storage vVols

Important: You might notice references to IBM consistency groups. However, this integration applies only to IBM FlashSystem A9000/R storage systems.

To view additional information about the storage objects, complete the following steps:

1. Go to the vCenter Server under the Resources list from the Global Inventory Lists view in the vSphere Client, and open an IBM Storage category to view additional information about the objects that are currently delegated to the selected vCenter server (Figure 4-59).

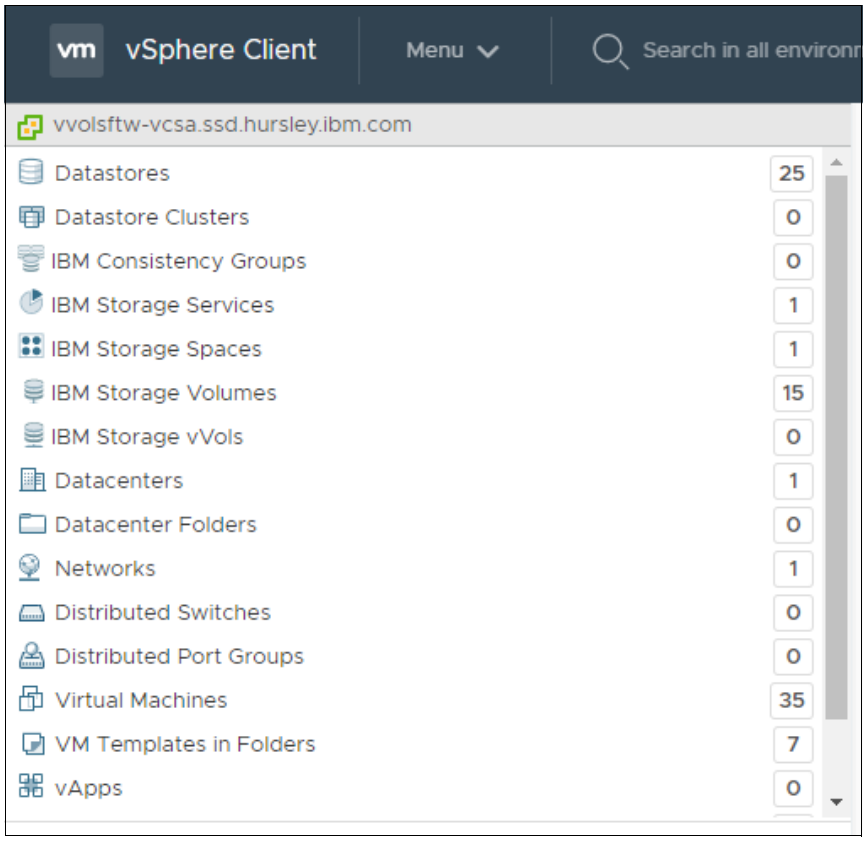


Figure 4-59 Global Inventory Lists

2. To view the capabilities names that are defined on a Storage Service, select **IBM Storage Services** in the menu on the left. Beneath the menu, select the required Storage Service (Figure 4-60).

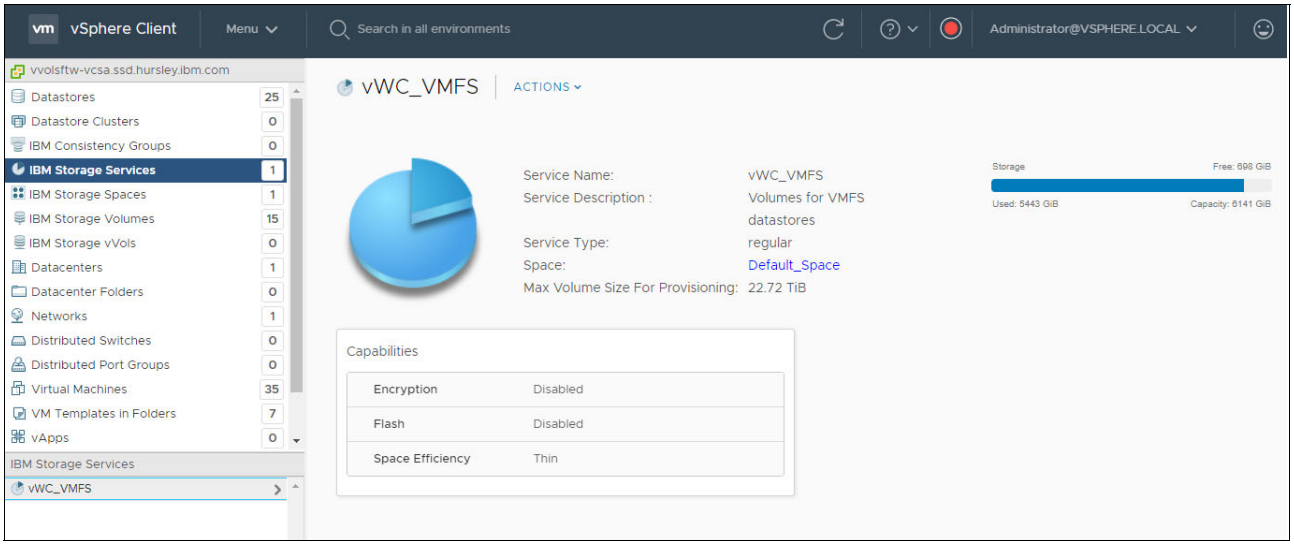


Figure 4-60 Viewing the capabilities that are defined on a Storage Service

3. To find specific information about a particular volume, select **IBM Storage Volumes** from the menu on the left. Beneath the menu, select the specific storage volume (Figure 4-61).

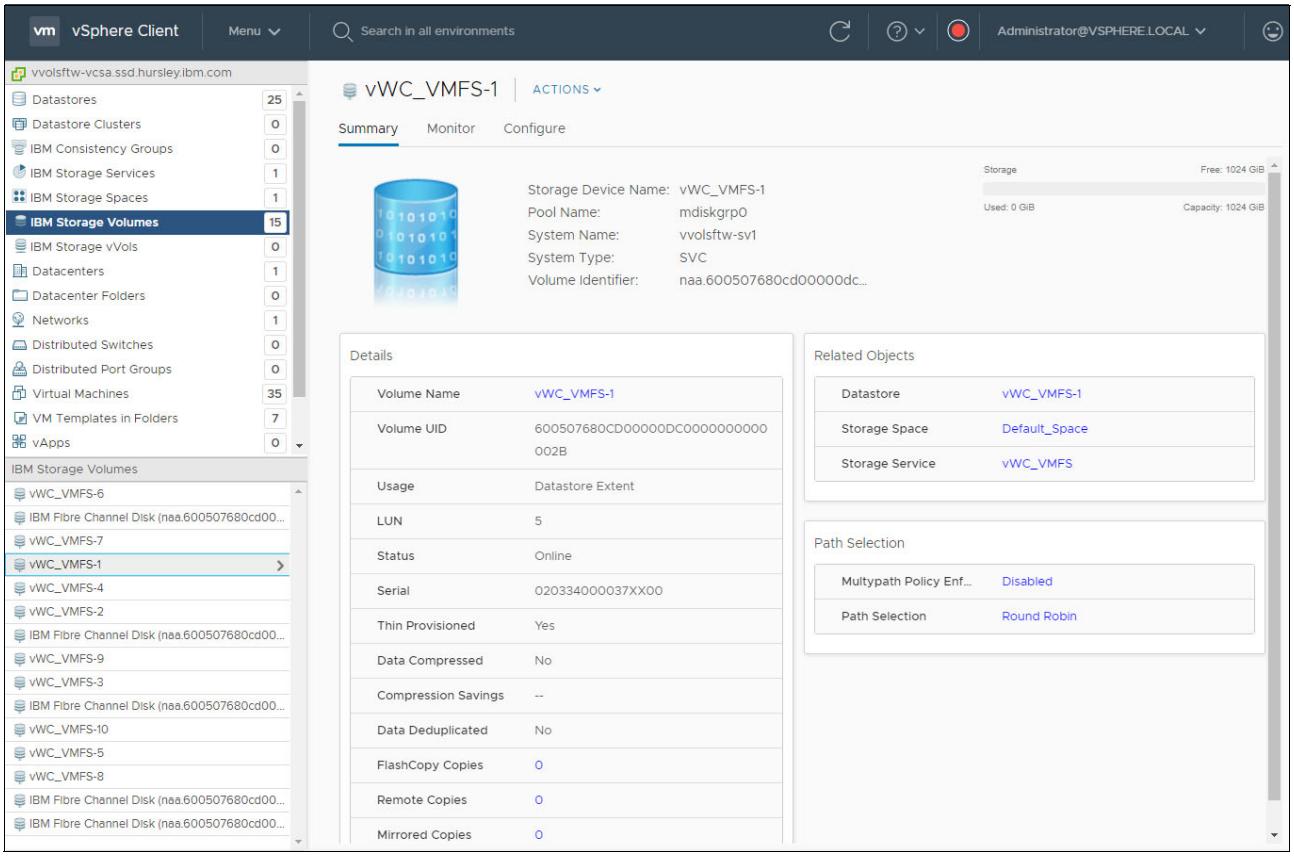


Figure 4-61 Finding specific information about a particular volume

4.6 Performing more storage volume management tasks

When viewing the additional information for a storage volume by using the Global Inventory List view in the vSphere Client, the vSphere Administrator can perform more tasks, such as:

- Rename the volume.
- Delete the volume.
- Define the Multipath Policy.
- Create more host mappings.
- Remove existing host mappings.
- Extend the volume size.

4.6.1 Considerations

Volume protection is an IBM Spectrum Virtualize feature that prevents volumes from being inadvertently deleted or unmapped from a host or host cluster. When attempting to delete a volume or remove existing host mappings, this task might fail if volume protection is enabled on the storage system, and the volume recently processed I/O operations. When this setting is enabled, volumes must be idle before they can be deleted from the system or unmapped from a host or host cluster. Volumes can be deleted only if they have been idle for the specified interval (by default this interval is set to 15 minutes).

When volume protection is disabled, volumes can be deleted even if they recently processed I/O operations. In the management GUI, select **Settings** → **System** → **Volume Protection** to manage volume protection values on the system and on specific pools. You must have the *SecurityAdmin* or *Administrator* role to change volume protection settings.

The Extend volume size task might fail if a thick-provisioned (fully allocated) volume is created on the storage system, and a fast-format task is still in progress. If this situation occurs, wait for the fast-formatting process to complete, and then run the command again.

When the Extend volume size task successfully completes, the LUN that is backing the data store increases in size but the VMFS file system does not change. You must rescan the HBA for each host that accesses the data store, so that the host can detect the change in volume size. Then, you must expand the VMFS file system by right-clicking a data store and selecting **Increase Datastore Capacity** to take advantage of the additional capacity.

4.7 IBM Storage Plug-in for VMware vRealize Orchestrator

The IBM Storage Plug-in for VMware vRO allows VMware administrators to include IBM storage discovery and provisioning in their vRO-automation workflows. The plug-in package can be downloaded from IBM Spectrum Connect, and then deployed on the vRO server. The deployment includes the matching of a unique token key that is set on both servers. Through vRO Client, dedicated IBM Storage control elements become available, which allows the issuing of workflows with storage volumes that are attached to the vRO server. Rather than issuing volume operations manually and being limited to one manual operation at a time, VMware administrators can plan and automate storage operations in their virtualized cloud environments.

4.7.1 Configuring IBM Spectrum Connect for VMware vRealize Orchestrator

The IBM Storage Plug-in for the VMware vRO is used for discovery of the IBM storage resources and provisioning automation workflows in the vRO.

To access the vRO management options, go to the Interfaces window of the IBM Spectrum Connect server and add the vRO server interface, as shown in Figure 4-62.

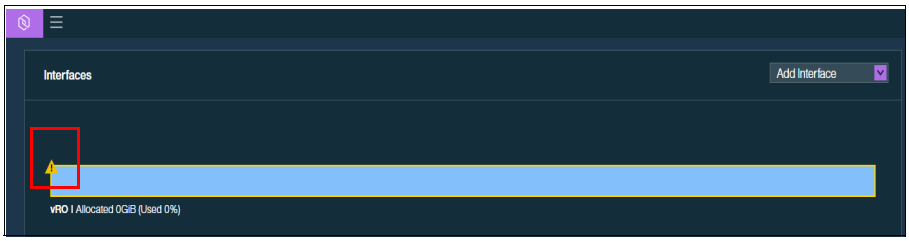


Figure 4-62 Interfaces window

The yellow frame and the exclamation mark (Figure 4-62) indicate that Storage Services are not yet delegated to the interface. You can then manage the integration with vRO as described in the following sections.

Downloading and installing the plug-in package for vRO

To enable the IBM Storage workflows in the vRO, you must first download the IBM Storage plug-in package from IBM Spectrum Connect and install it on the vRO server.

To download and install the IBM Storage plug-in package, complete the following steps:

1. On the Interfaces window, right-click the vRO server, and then select **Modify**.
2. On the bottom of the vRO Settings dialog, click **Download plug-in package** to save the package to your local computer (Figure 4-63).

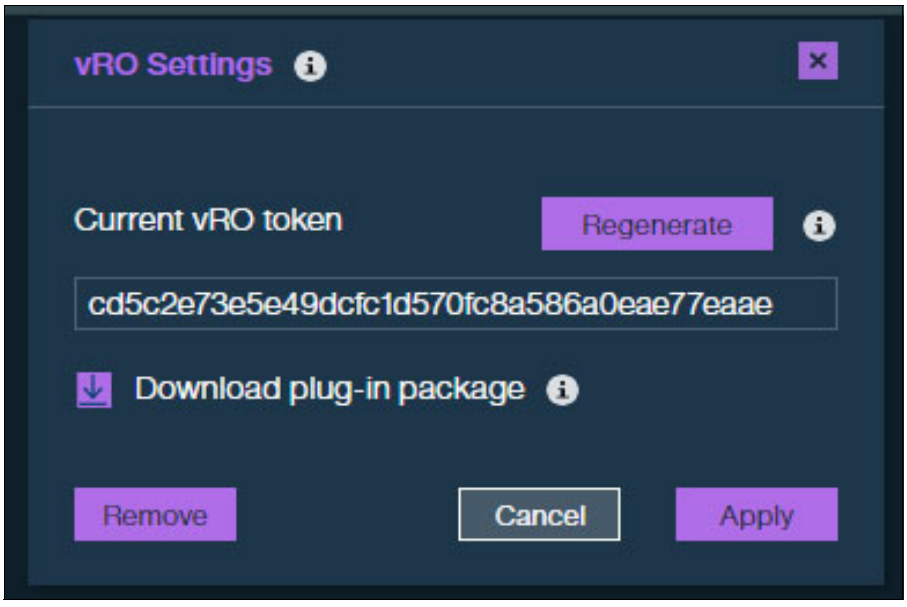


Figure 4-63 Downloading the plug-in package

Alternatively, you can download the package from Downloading and installing the plug-in package for vRO.

3. Copy the *current vRO token key* from the Current vRO token input box. The current vRO token key will be used in step 11 on page 95.
4. In the vSphere Client, select the **Configure** tab.
5. Click **Manage Plug-Ins** in the Plug-Ins category. Select **Browse** → **Install**.
6. Locate and choose the downloaded plug-in file.
7. Accept the license agreement and click **INSTALL** (Figure 4-64).

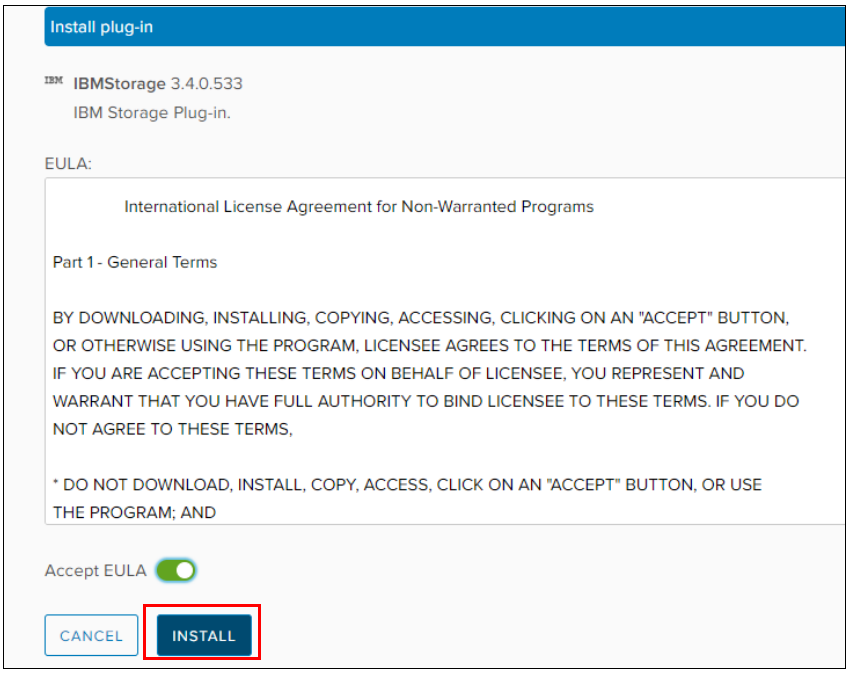


Figure 4-64 Installing the plug-in

The “Plug-in ‘IBMStorage’ (3.x.x build xxx) is installed” message is displayed (Figure 4-65).

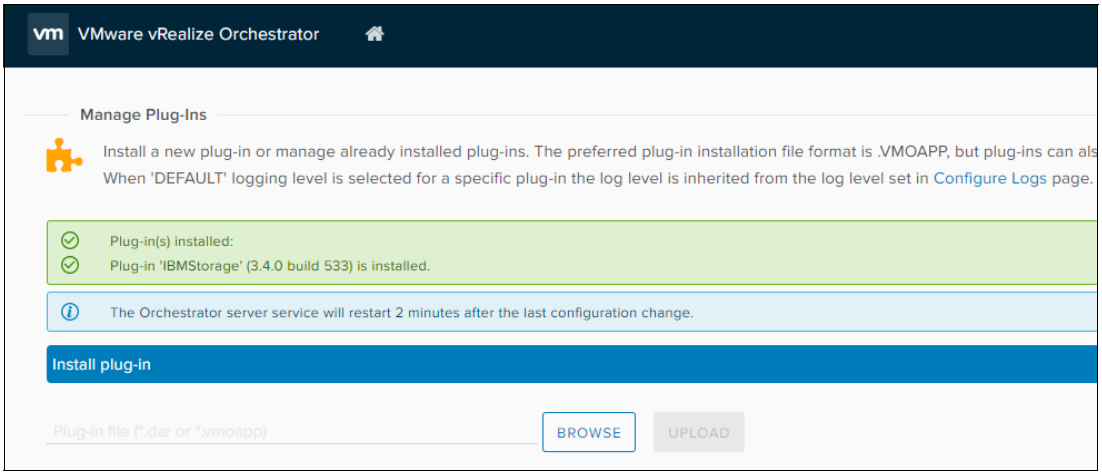


Figure 4-65 Confirmation message

Installation is completed and the IBM Storage plug-in is displayed in the list of vRO plug-ins.

8. Start the VRO Client and go to the **Workflows** tab.

9. On the **Workflows** tab, add a search filter for “IBM” to list the new workflows available using IBM Spectrum Connect (Figure 4-66).

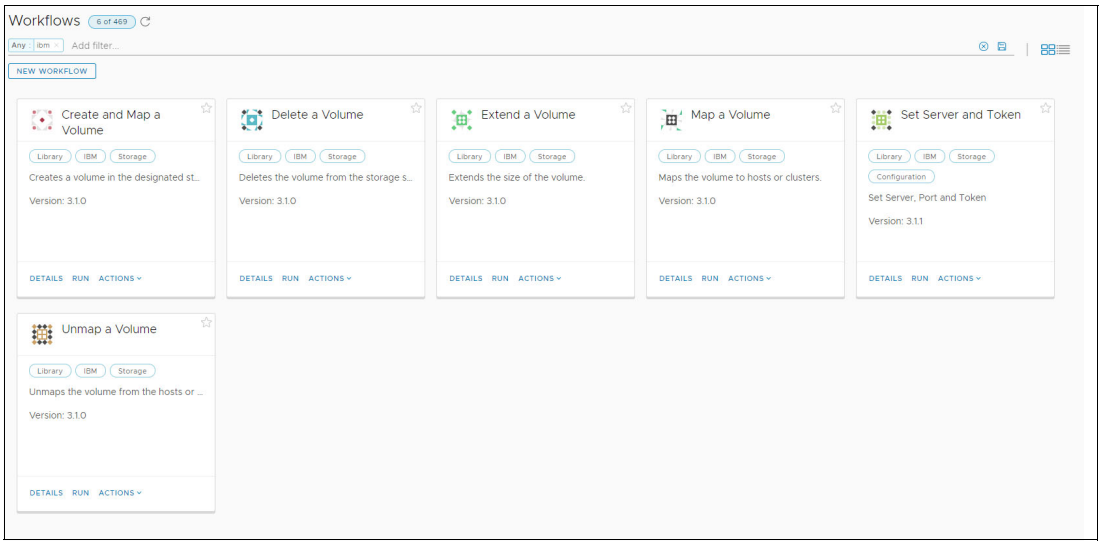


Figure 4-66 Listing the new workflows

10. Locate the Set Server and Token workflow and click **Run**.
11. Enter the following information in the correct fields in Figure 4-67:
    - The server field: Enter the FQDN.
    - The port field: Enter the port of the IBM Spectrum Connect server.
    - The token field: Paste the token from step 3 on page 94, and click **Run**.

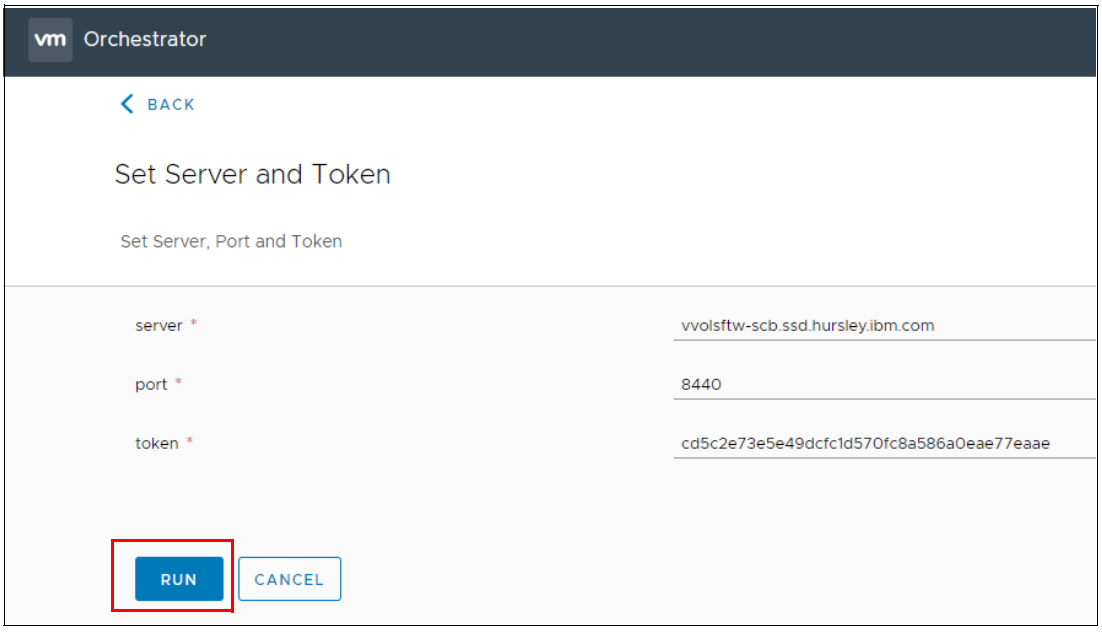


Figure 4-67 Set Server and Token screen

Tip: If you experience issues running the Set Server and Token workflow, retry the procedure with a web browser in Incognito or Private Browsing modes.

The workflow starts and a completion status is returned (Figure 4-68).

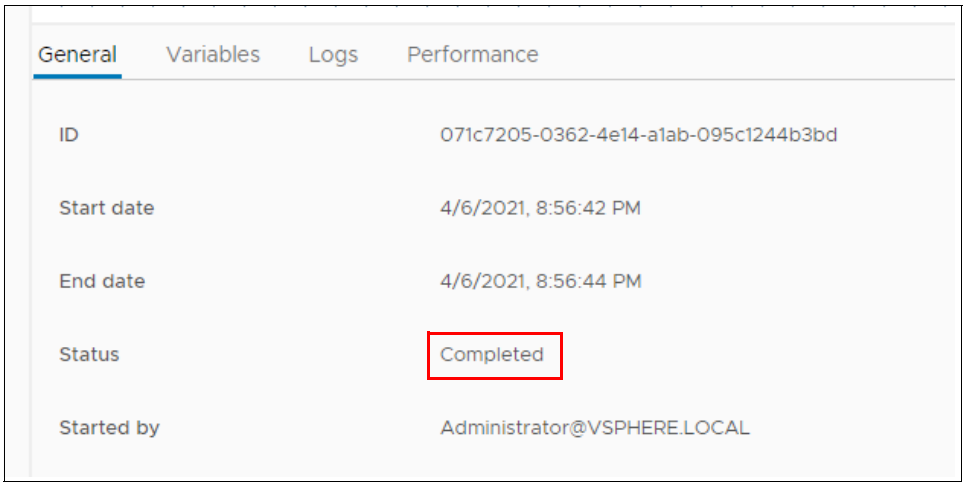


Figure 4-68 Workflow completed

The initial configuration of the vRO plug-in is now complete.

4.7.2 Using vRealize Automation and VMware vRealize Orchestrator

If a Storage Service with an allocated storage resource is not already provisioned, you must now create one and delegate it to the vRO interface (as described in 4.2.1, “Creating Storage Services” on page 52) before any volumes can be created by using vRO workflows.

After a Storage Service is created and allocated, complete the following steps:

1. Run the Create and Map a Volume workflow (Figure 4-69).

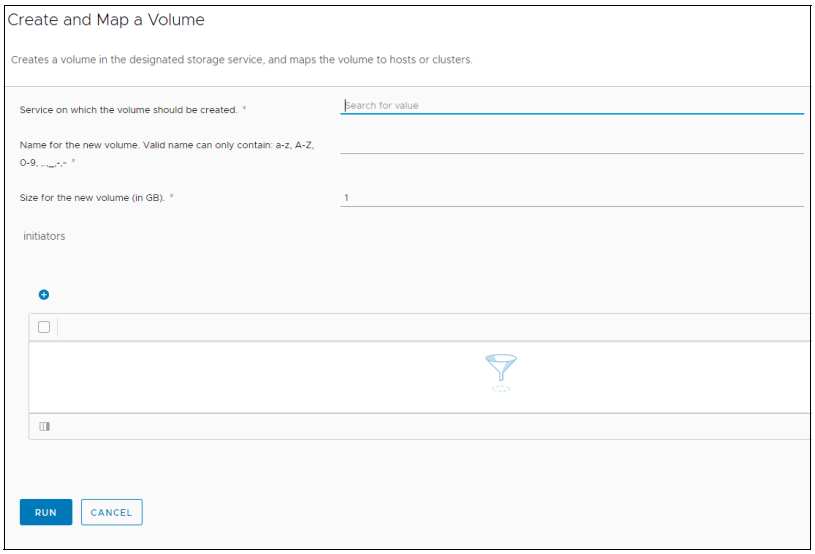


Figure 4-69 Create and Map a Volume workflow

2. Click in the “Service on which the volume should be created” window and search for a Storage Service (Figure 4-70 on page 97).

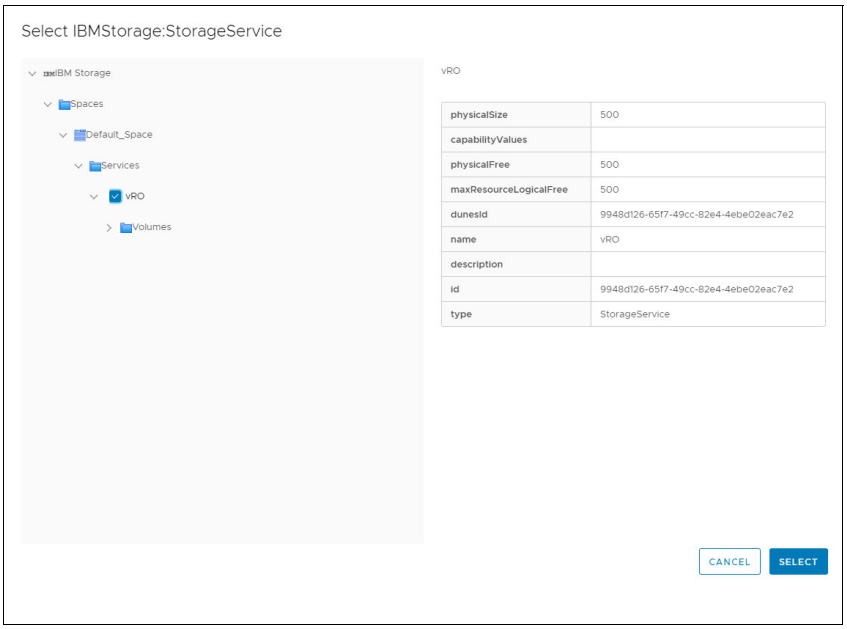


Figure 4-70 Selecting a Storage Service

3. Enter the following information:
   - Name for the new volume: Enter a volume name.
   - Size for the new volume (in GB): Enter the capacity for the new volume.
   - Click the plus icon (**+**) at the bottom of the window and enter the worldwide port name (WWPN) (Fibre Channel (FC)) or iqn (Internet Small Computer System Interface (iSCSI)) for the hosts that you want the volume mapped to and click **Run** (Figure 4-71).

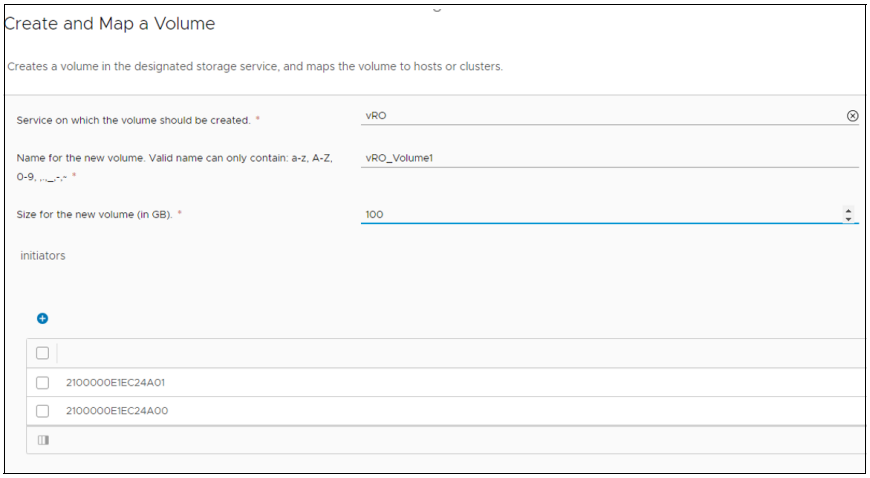


Figure 4-71 Create and Map a Volume

4. After the task is complete, review the Logs workflow (Figure 4-72) and the Audit Log (Figure 4-73) of the Storage System.

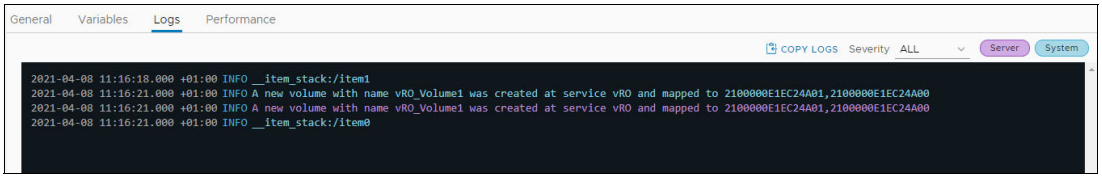


Figure 4-72 Logs workflow

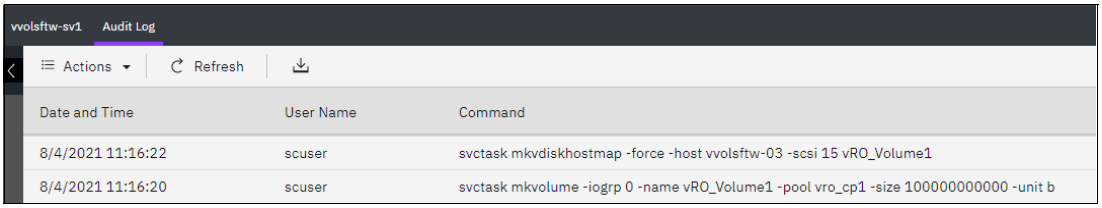


Figure 4-73 Audit Log

4.8 IBM Storage Management Pack for VMware vRealize Operations Manager

The IBM Storage Management Pack for VMware vROps Manager allows vROps Manager users to obtain comprehensive monitoring information about the IBM storage resources that are used in their virtualized environment.

Connectivity to VMware vRealize Operations Manager can be provided either by using IBM Spectrum Connect to act as the middle-ware between vROps and IBM Spectrum Virtualize, or by using the dedicated storage plug-in that is provided by VMware that communicates directly with the storage system.

The VMware plug-in is available through the VMware Marketplace and is published as “vRealize Operations Management Pack for IBM SAN Volume Controller and IBM Storwize 4.0.0”. For more information, see Management Pack for IBM SVC and Storwize.

When using IBM Spectrum Connect for vROps integration, the management pack can be downloaded from the IBM Spectrum Connect GUI and then deployed on the vROps Manager server. After a VMware vROps Manager server is registered on an instance of IBM Spectrum Connect that is configured with storage systems, storage spaces, services, and vRealize servers, the storage-related data is pushed to the vROps Managerserver in 5-minute intervals by default.

The dedicated IBM storage system adapter that is deployed on the vROps Manager server enables monitoring of the supported IBM storage system by using the vROps Manager. This adapter reports the storage-related information, such as monitoring data of all logical and physical elements, covering storage systems, storage domains, storage pools, volumes, hosts, modules, target ports, disks, health status, events, thresholds, and performance. It also provides the dashboards that display detailed status, statistics, metrics, and analytics data alongside hierarchical flowcharts with graphic representation of IBM storage system elements.

Relationships between the IBM storage system elements (storage systems, ports, storage pools, volumes, host, host initiator, modules, domain) and data stores, VMs, and hosts are displayed graphically in a drill-down style. This display provides VMware administrators with a complete and up-to-date picture of their used storage resources.

4.8.1 Configuring IBM Spectrum Connect for VMware vRealize Operations Manager

Before you can use the IBM Storage Management Pack for VMware vROps Manager, you must set a connection to at least one vROps Manager server, and then define which storage systems should be monitored in vROps.

Downloading the vROps management package

IBM Spectrum Connect provides management package in the form of a PAK file, which can be deployed on the vROps Manager.

To download the PAK file from IBM Spectrum Connect, complete the following steps:

1. Go to the Monitoring window of the IBM Spectrum Connect GUI. The Set vROps Server dialog is displayed (Figure 4-74).

Figure 4-74 Set vROps Server dialog

2. On the bottom of the Monitoring window, click **Download PAK file** to save the file to your local computer. Alternatively, you can access the PAK file by using SSH or Secure Copy Protocol (SCP) by copying the package from the following directory on the IBM Spectrum Connect server:

   `/opt/ibm/ibm_spectrum_connect/downloads/static/IBM_Storage_Adapter-3.x.xxxx_signed.pak`

   – *x.x* is the release and mod number.

   – *xxx* is the current build number.

3. Save the file to your computer to later upload it to the vROps Manager.

4.8.2 Installing Management Pack in VMware vRealize Operations Manager

After the management package is downloaded to the computer, it must be deployed on the vROps Manager.

To deploy the management package on the vROps, complete the following steps:

1. Access the vROps Manager administrative web console by using `https://<hostname or IP address of the vROps UI>`.
2. Select **Administration** → **Solutions** → **Repository**.
3. In the Repository window, click **ADD/UPGRADE** to add a management package. The Add Solution dialog is displayed.

4. In the Add Solution dialog, click **Browse** and select the management package that is downloaded from IBM Spectrum Connect (Figure 4-75). Click **UPLOAD** to start deployment.

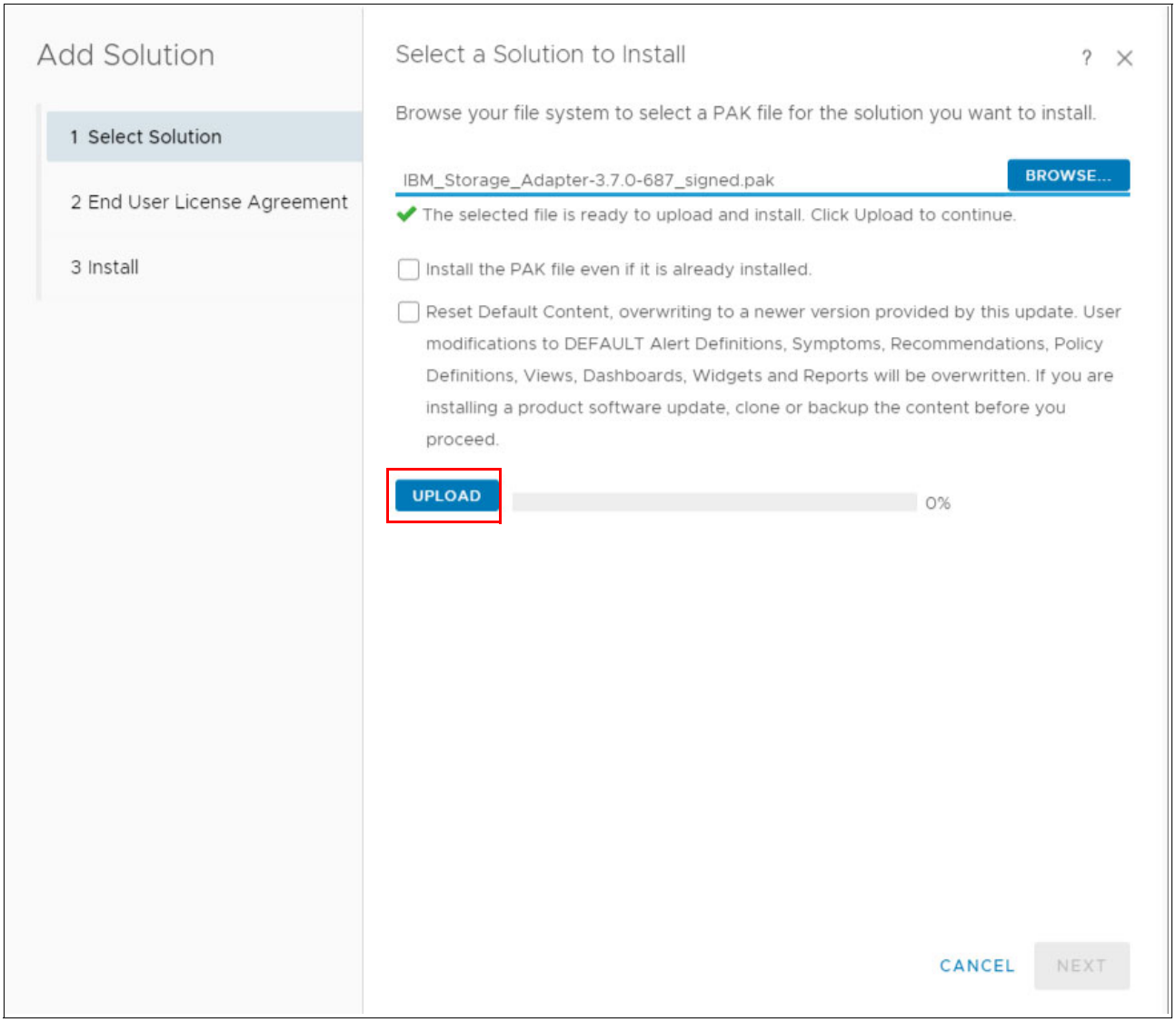


Figure 4-75 Starting a deployment

After the package is uploaded, the package information is displayed (Figure 4-76).

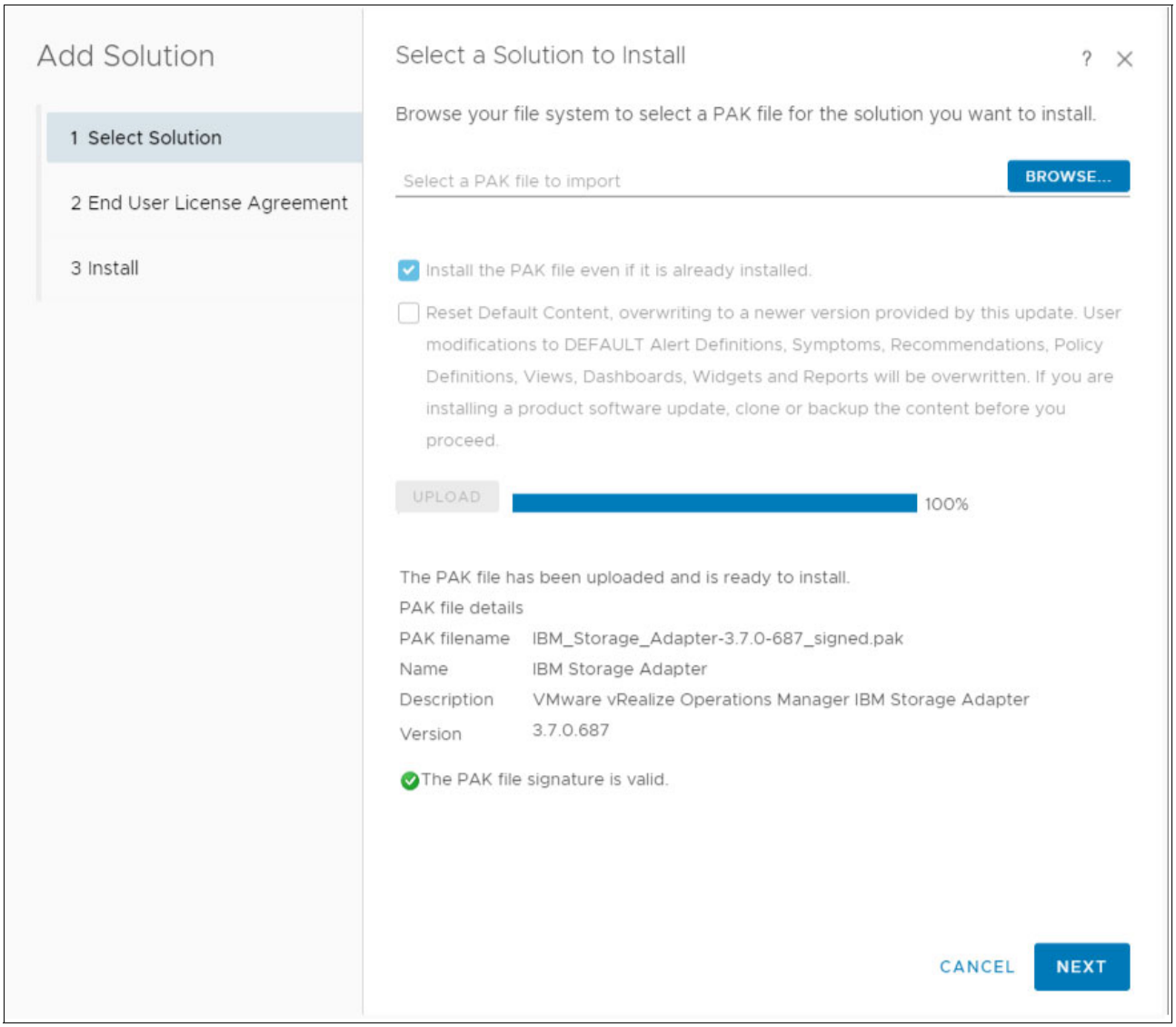


Figure 4-76 Package information

5. Click **NEXT**. The IBM license agreement is displayed.
6. Accept the IBM license agreement and click **NEXT** to continue. The Installation Solution progress is displayed.
7. Click **FINISH** to complete the installation. More configuration of the management package is not required on the vROps. Under certain conditions, the package's status might appear as `Not configured on the vROps`. You can disregard this information.

Connecting the vROps server to IBM Spectrum Connect

After the management package is successfully deployed, you must add the vROps Manager server to IBM Spectrum Connect. Then, the vROps Manager server must be connected to IBM Spectrum Connect.

To add the VROps Manager server to IBM Spectrum Connect, complete the following steps:

1. Go to the Monitoring window of the IBM Spectrum Connect GUI (Figure 4-77).
2. Enter the following information:
   - IP/Hostname: IP address or FQDN of the vROps Manager server
   - Username
   - Password
3. Select the checkbox to confirm you installed the PAK file on the vROps Manager server.

Figure 4-77 Monitoring window

4. Click **Apply** to save the settings.

Controlling storage system monitoring on the vROps server

When a single IBM Spectrum Connect instance is managing multiple IBM Spectrum Virtualize storage systems, you can select which systems will be monitored and which will be ignored.

To enable monitoring, complete the following steps:

1. Go to the Monitoring window of the IBM Spectrum Connect GUI.
2. In the Storage Systems window, right-click a storage system that you intend to monitor, and select **Start vROps monitoring**, or click **Start Monitoring** on the storage system (Figure 4-78). The monitored system color changes to indicate the connection to the vROps server. IBM Spectrum Connect starts pushing the information to vROps Manager by using RESTful API requests.

   To stop monitoring a storage system, click **Stop Monitoring** on the monitored system.

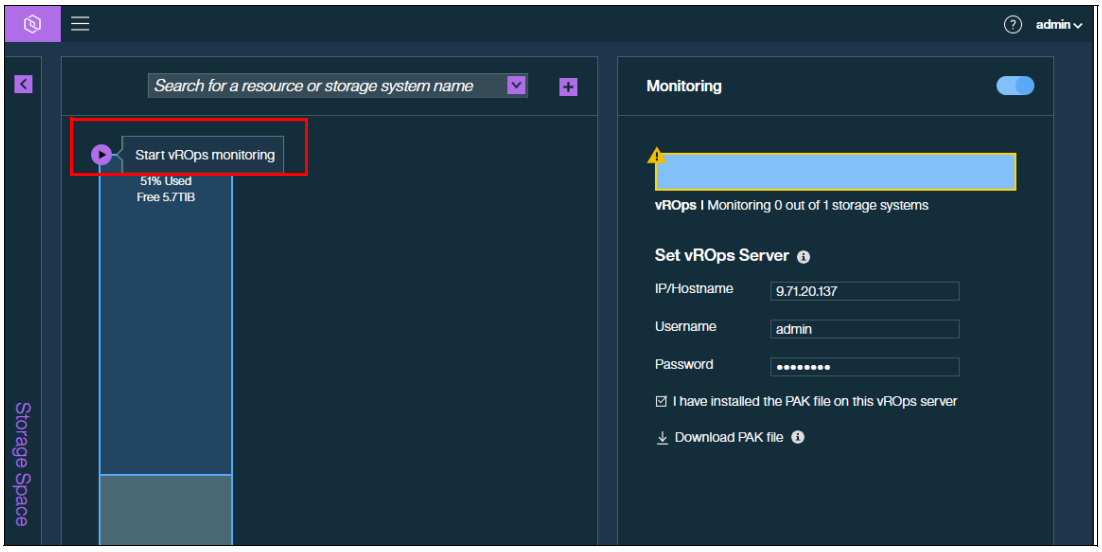


Figure 4-78 Start vROps monitoring

After a vROps server connection is defined and storage systems are associated with the vROps server, detailed monitoring information for these storage systems becomes available in vROps.

5

VMware and IBM Spectrum Virtualize multi-site guidelines

The term *disaster recovery* (DR) is normally used regarding a large, significant, and disruptive event, such as an earthquake or flood. But DR can also be valuable for smaller events, such as power-loss or a localized network failure.

Companies prepare for DR by implementing *business continuity* solutions to maintain and restore operations if a disruption or disaster occurs, but they do not always test those solutions regularly. The ability to recover the systems needs to be tested regularly to make sure that procedures work, rather than waiting until a disruption happens. Flaws might be detected each time you test because perfection is impossible to achieve when the environment changes every day.

Copy services are a collection of functions that provide capabilities for business continuity, disaster recovery, data migration, and data duplication solutions. This chapter provides an overview and the preferred best practices guide for VMware and IBM Spectrum Virtualize for copy services capabilities. These capabilities include FlashCopy, Metro Mirror, Global Mirror, and VMware Site Recovery Manager.

This chapter describes some of the solutions that can help you prepare your environment to recover from a disruption and includes the following sections:

5.1 Copy Services overview

IBM Spectrum Virtualize storage system offers a complete set of copy services functions to VMware that provide capabilities for business continuity, disaster recovery, data movement, and data duplication solutions. The IBM Spectrum Virtualize Family Storage Replication Adapter (SRA) is a software add-on that integrates with the VMware Site Recovery Manager (SRM) solution and enables SRM to perform failovers together with supported IBM FlashSystem Storage. You can make mirror images of part or all of your data between two sites, which is advantageous in DR scenarios with the capabilities of copying data from production environments to another site for resilience.

The following copy services (relationships) are supported by IBM Spectrum Virtualize storage system:

- FlashCopy, for point-in-time copy
- Metro Mirror, for synchronous remote copy
- Global Mirror, for asynchronous remote copy
- Global Mirror with Change Volumes (GMCV), for asynchronous remote copy for a low-bandwidth connection

Replication relationships can be created between a maximum of four independent IBM Spectrum Virtualize storage systems. Partnerships can be a mix of any of the IBM Spectrum Virtualize systems. For example, an IBM FlashSystem storage array replicating to a SAN Volume Controller (SVC) storage system and vice versa. For more information about these services, see Chapter 10, “Advanced Copy Services” in the IBM Redbooks publication titled*Implementing IBM FlashSystem with IBM Spectrum Virtualize V8.4*, SG24-8492.

Note: All these services are supported by VMware SRM when using IBM Spectrum Virtualize Family SRA.

5.1.1 FlashCopy

FlashCopy is known as a *point-in-time copy*. It makes a copy of the blocks from a source volume and duplicates them to the target volumes.

When you initiate a FlashCopy operation, a FlashCopy relationship is created between a source volume and a target volume. A FlashCopy relationship is a mapping of the FlashCopy source volume and a FlashCopy target volume. This mapping allows a point-in-time copy of that source volume to be copied to the associated target volume. If it is a persistent FlashCopy, the FlashCopy relationship exists between this volume pair from the time that you initiate a FlashCopy operation until the storage unit copies all data from the source volume to the target volume, or you delete the FlashCopy relationship.

5.1.2 Metro Mirror

Metro Mirror is a type of remote copying that creates a synchronous copy of data from a primary volume to a secondary volume that is read-only. A secondary volume can be located either on the same system or on another system. The maximum distance that is allowed between systems in Metro Mirror relationships is 300 km. When a host issues a **`write`** command to a volume, the data is replicated to the remote cluster before the host is notified that the I/O completed.

Tip: Metro Mirror can impact write latency. For best performance, use shorter distances and create Metro Mirror relationships only between systems with similar performance.

5.1.3 Global Mirror

The Global Mirror function provides an asynchronous copy process. When a host writes to the primary volume, confirmation of I/O completion is received before the write operation completes for the copy on the secondary volume. The maximum acceptable distance between systems in Global Mirror relationships is 25.000 km or 250 ms latency.

Global Mirror change volumes

Global Mirror *change volumes* are copies of data from a primary volume or secondary volume that are used in Global Mirror relationships. Using change volumes lowers bandwidth requirements by addressing only the average throughput, not the peak.

5.1.4 Remote copy consistency groups

You can group Metro Mirror or Global Mirror relationships into a *consistency group* so that they can be updated at the same time. A command is then simultaneously applied to all of the relationships in the consistency group.

5.1.5 VMware Site Recovery Manager

VMware SRM is well known in the virtualization world for providing simple, affordable, and reliable business continuity and disaster recovery management.

Using SRM with IBM Spectrum Virtualize storage system can help you protect your virtual environment.

SRM automates the failover processes and the ability to test failover processes or DR without having a negative impact on the live environment, which helps you meet your recovery time objectives (RTOs).

VMware SRM supports two forms of replication:

- Array-based replication (ABR), where the storage system manages the virtual machine (VM) replication with the following attributes:
  - Compatible storage is required, such as systems powered by IBM Spectrum Virtualize.
  - Storage arrays are configured with a SRA.
- Host-based replication, which is known as vSphere Replication (VR), where the Elastic Sky X integrated (ESXi) manages the VM replication with the following attributes:
  - Does not depend on storage array compatibility.
  - Increased network efficiency by replicating only the most recent data in the changed disk areas.
  - Minimum RPO = 5 minutes.

Figure 5-1 shows an overview of the VMware SRM.

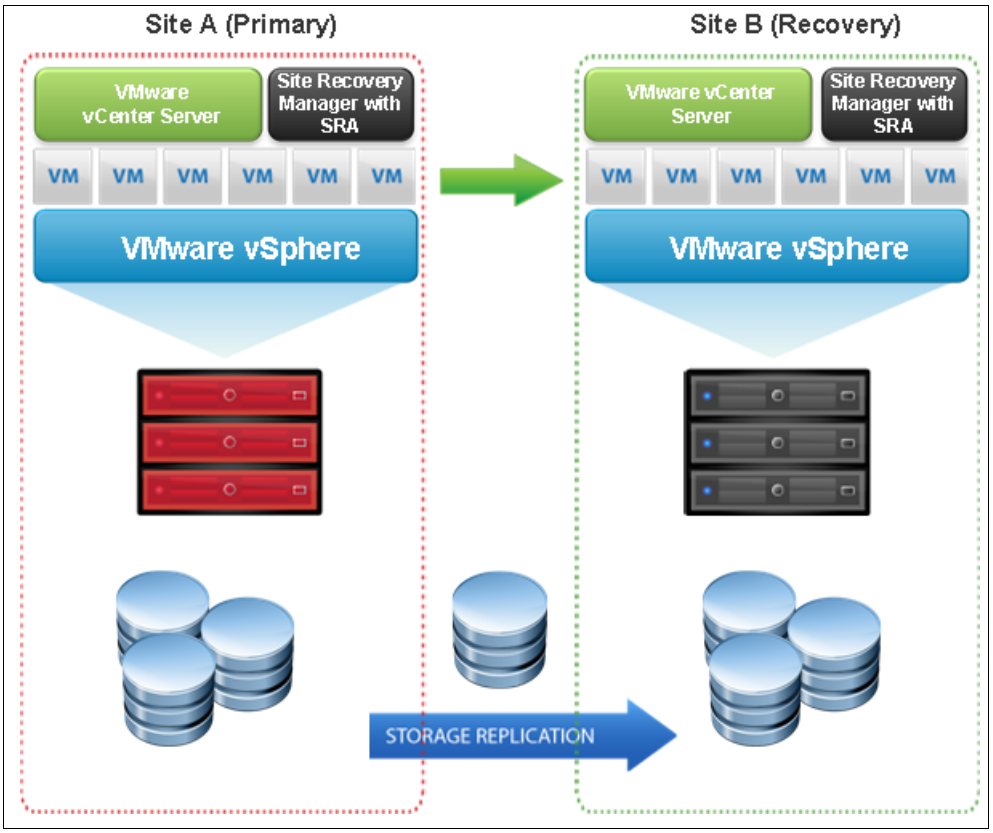


Figure 5-1 VMware SRM

VMware SRM requires one vCenter server in each site with the respective licenses. Also, if you are using SRM with IBM FlashSystem storage, you are required to use an *IBM Spectrum Virtualize SRA*, which is described in 5.1.6, “Storage Replication Adapter” on page 108.

For more information about SRM, see the VMware Site Recovery Manager Documentation.

5.1.6 Storage Replication Adapter

SRA is a storage vendor plug-in that is developed by IBM. SRA is required for the correct functioning of SRM when it is used with IBM Spectrum Virtualize storage systems.

The adapter is used to enable the management of Advanced Copy Services (ACS) on IBM FlashSystem Storage, such as Metro Mirror and Global Mirror (including changed volumes).

The combination of SRM and SRA enables the automated failover of VMs from one location to another, connected by either Metro Mirror or Global Mirror technology.

By using the IBM Spectrum Virtualize Family Storage Replication Adapter, VMware administrators can automate the failover of an IBM FlashSystem 9200 at the primary SRM site to a compatible system, such as another IBM FlashSystem 9200, 7200, or IBM SAN Volume Controller at a recovery (secondary) SRM site.

In a failover, the ESXi servers at the secondary SRM site mount the replicated data stores on the mirrored volumes of the auxiliary storage system. When the primary site is back online, perform fallback from the recovery site to the primary site by clicking **Reprotect** in the SRM.

For more information, see the IBM Spectrum Virtualize Family Storage Replication Adapter documentation.

5.2 Storage Replication Adapter with VMware Site Recovery Manager

Figure 5-2 shows how an IBM Spectrum Virtualize storage system is integrated in a typical VMware SRM DR solution.

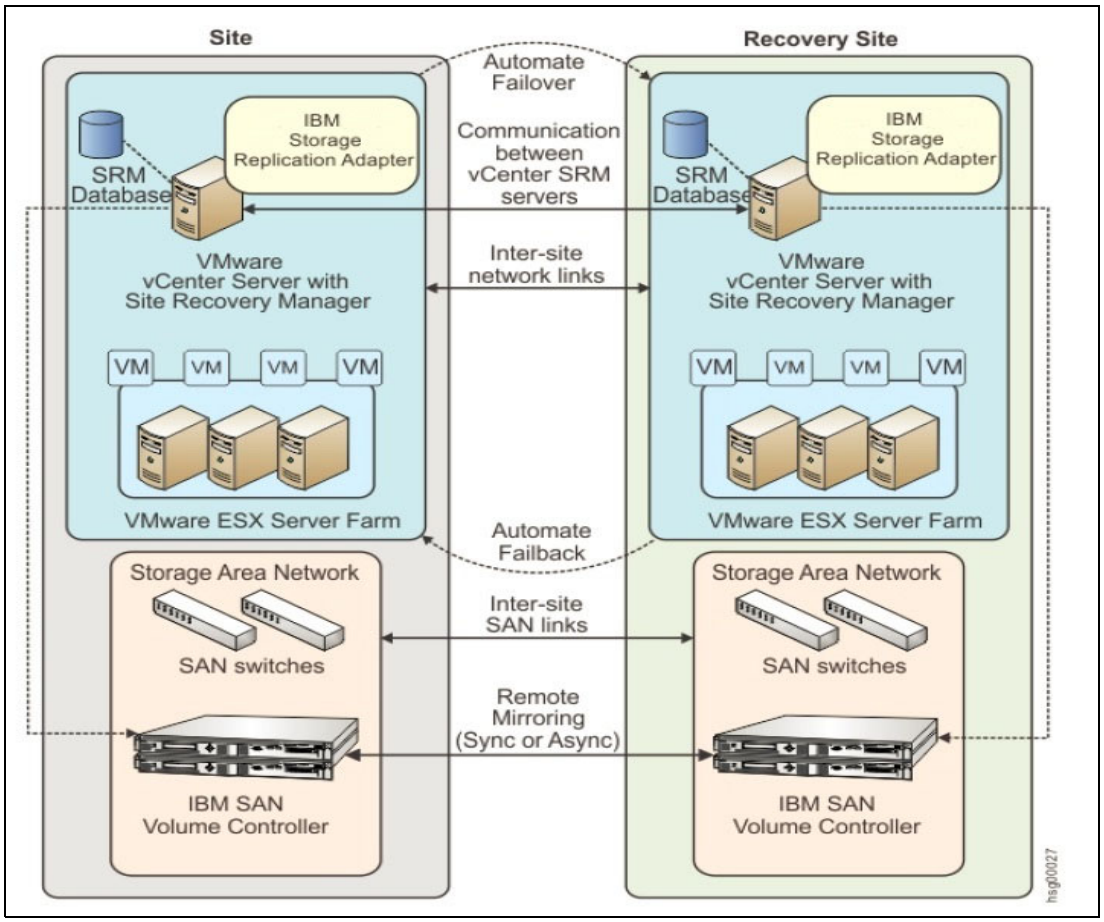


Figure 5-2 SRA and VMware SRM with IBM Spectrum Virtualize integrated solution

SRA configuration might vary depending on the specific site configuration. Consider the following preparatory steps and configuration when using SRA with SRM.

5.2.1 Storage replication adapter planning

This section describes storage replication adapter planning.

Preparing the SRA environment

To prepare the SRA environment, complete the following steps:

1. Ensure that the supported storage systems firmware version is used.
2. Provision appropriate-sized target volumes on the storage system at the recovery (secondary) site. Create Metro or Global-Mirror relationships between the source and target volumes and add the relationships to consistency groups, as needed.
3. Create a dedicated user on both source and target storage systems with the right privileges for the SRA:
   - For SRA non-preconfigured settings, a user with *Administrator* or higher privilege is needed.
   - For SRA preconfigured settings, a user with *CopyOperator* or higher privilege is needed.
4. Use the same username and password on both the protected (primary) and the recovery site.

Verifying the mirroring configuration

Consider the following points for verifying the mirroring configuration:

- Ensure that all VMware Elastic Sky X (ESX) hosts, IBM Spectrum Virtualize storage systems, and volumes at both sites are properly connected to their remote counterparts and configured for site mirroring.
- Establish mirror-connectivity between the local IBM Spectrum Virtualize storage system at the protected (primary) site and the IBM Spectrum Virtualize storage system at the recovery (secondary) site. For IBM Stretched Cluster when using SVC, stretched volumes are created, and both copies of a stretched volume are online. For IBM HyperSwap, HyperSwap volumes are created, and both the primary volume and secondary volume of a HyperSwap volume are online.
- Use a unique name for each IBM Spectrum Virtualize storage system at both the protected and the recovery sites.
- Make sure that the storage pools that contain the replicated volumes at both sites have sufficient available capacity for creating the snapshots of all replicated volumes concurrently.
- For non pre-configured environments, an extra space for Test Failover and Failover is necessary. Ensure that enough space is available in the pool.
- Ensure that protected volumes are mapped to the protected VMware ESX hosts:
  - For Stretched Cluster, the stretched volumes are mapped to both the protected and recovery VMware ESX hosts.
  - For IBM HyperSwap, the primary volume of a HyperSwap is mapped to both the protected and recovery VMware ESX hosts.
- Ensure that remote copy relationships exist for all volumes:
  - For IBM Stretched Cluster, stretched volumes are created, and both copies of a stretched volume are online.
  - For IBM HyperSwap, HyperSwap volumes are created, and both the primary volume and secondary volume of a HyperSwap volume are online.

- Ensure that for non-preconfigured environments, the recovery volumes remain unmapped.
- Make sure that the recovery VMware ESX or ESXi hosts are defined as hosts at the recovery site and report as online.

Verifying the VMware Site Recovery Manager installation

Before you embark with the installation of IBM Spectrum Virtualize storage system SRA container, verify that the VMware SRM is already installed and accessible at both protected (primary) site and the recovery (secondary) site by following these steps:

1. Log in to the VMware vSphere Web Client (Figure 5-3).

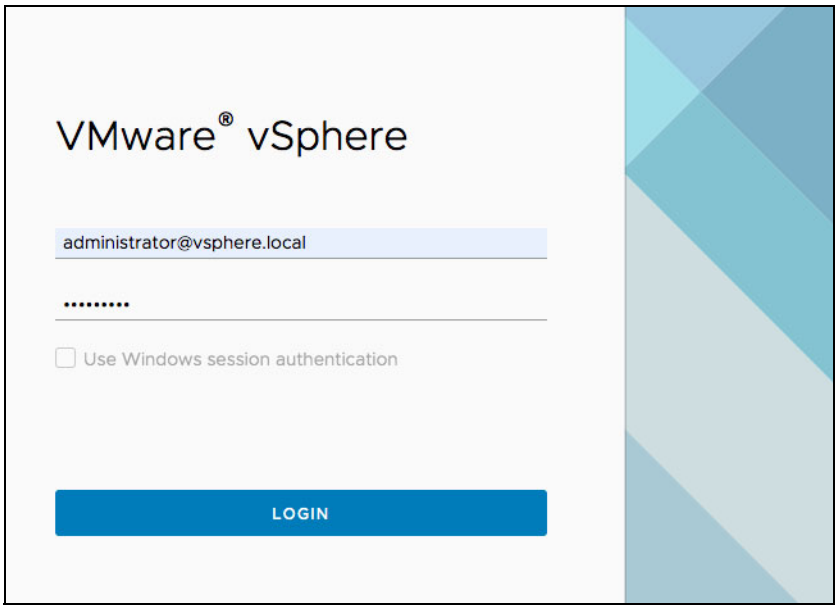


Figure 5-3 VMware vSphere Web Client login

2. Go to the vSphere Client window and verify that the Site Recovery icon is displayed (Figure 5-4).

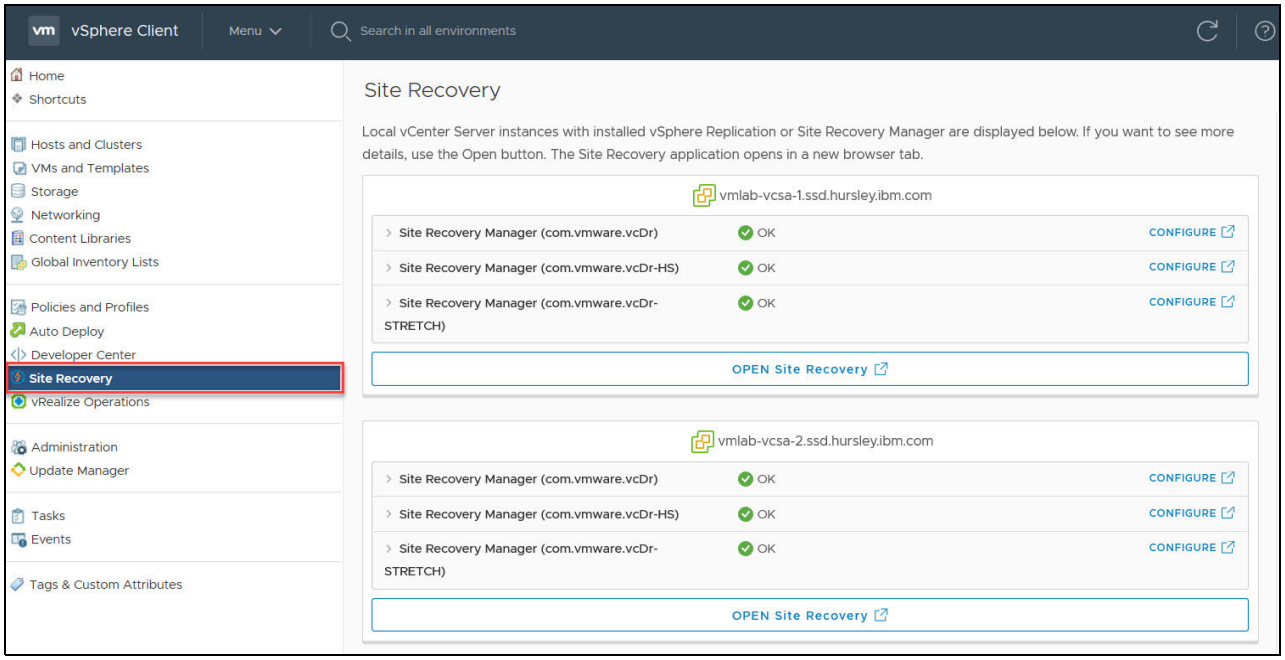


Figure 5-4 VMware vSphere Web Client home page

5.2.2 Storage Replication Adapter for VMware installation

For more information about how to download IBM Spectrum Virtualize Family Storage Replication Adapter for VMware, see Downloading the SRA.

As a best practice, stop your currently installed version of the SRA container before running a different version. Ensure that you satisfy all of the prerequisites that are listed in before you run the SRA container. Follow the below steps to run the IBM Spectrum Virtualize storage system SRA container on the SRM server.

Note: The IBM Spectrum Virtualize Family SRA installation creates a shortcut that is named IBM Spectrum Virtualize Family SRA Configuration `Utility.exe` on the desktop. The configuration utility must be run on both the protected (primary) site and the recovery (secondary) site SRM host.

1. Log in to the VMware SRM Appliance Management interface as admin, as shown in Figure 5-5 on page 113.

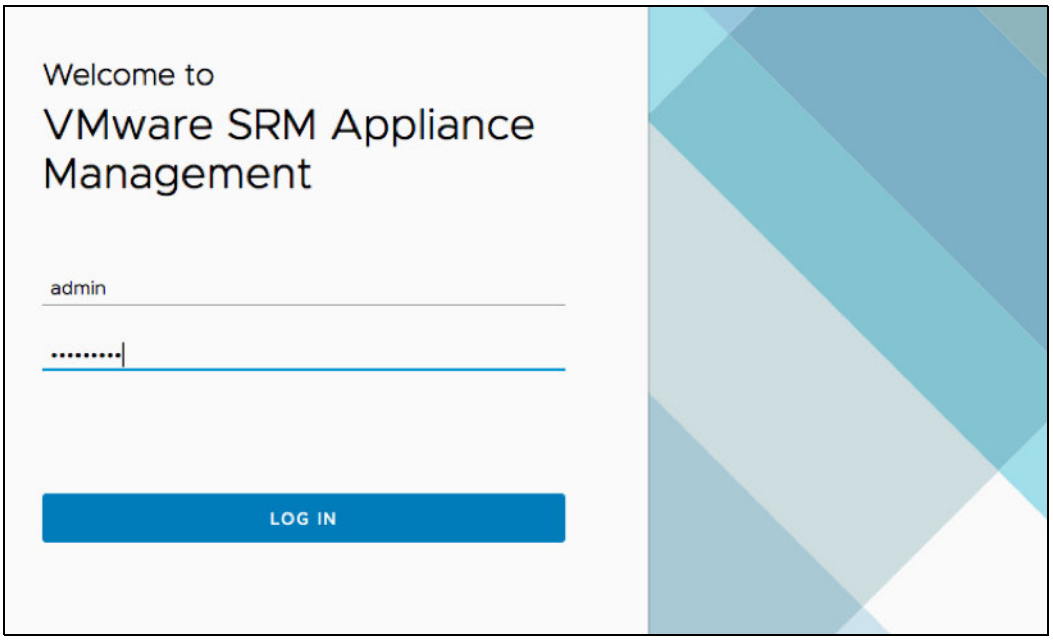


Figure 5-5 SRM Appliance Management interface login

2. In the SRM Appliance Management interface, select **Storage Replication Adapters** → **NEW ADAPTER (**Figure 5-6).

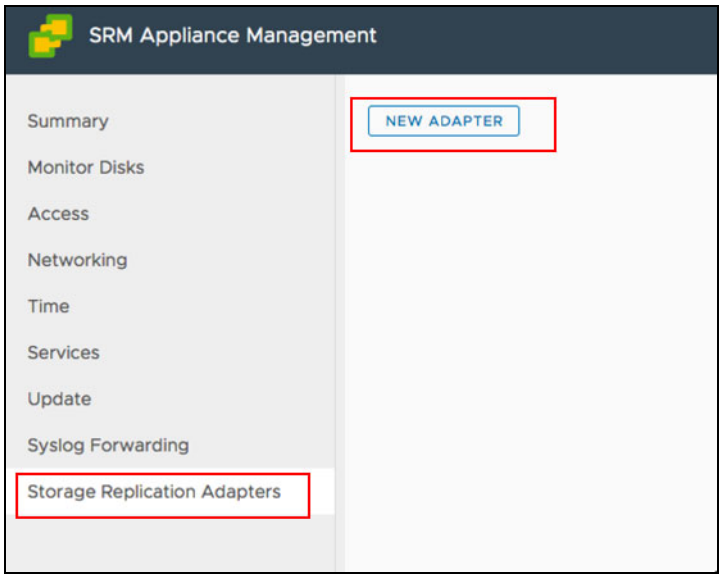


Figure 5-6 SRM Appliance Management interface

3. Click **Upload**. Go to the directory where you saved the SRA file, and double-click it. The SRA upload process begins.

4. When the upload process finishes, click **Close**. The SRA card is displayed on the Storage Replication Adapters view (Figure 5-7).

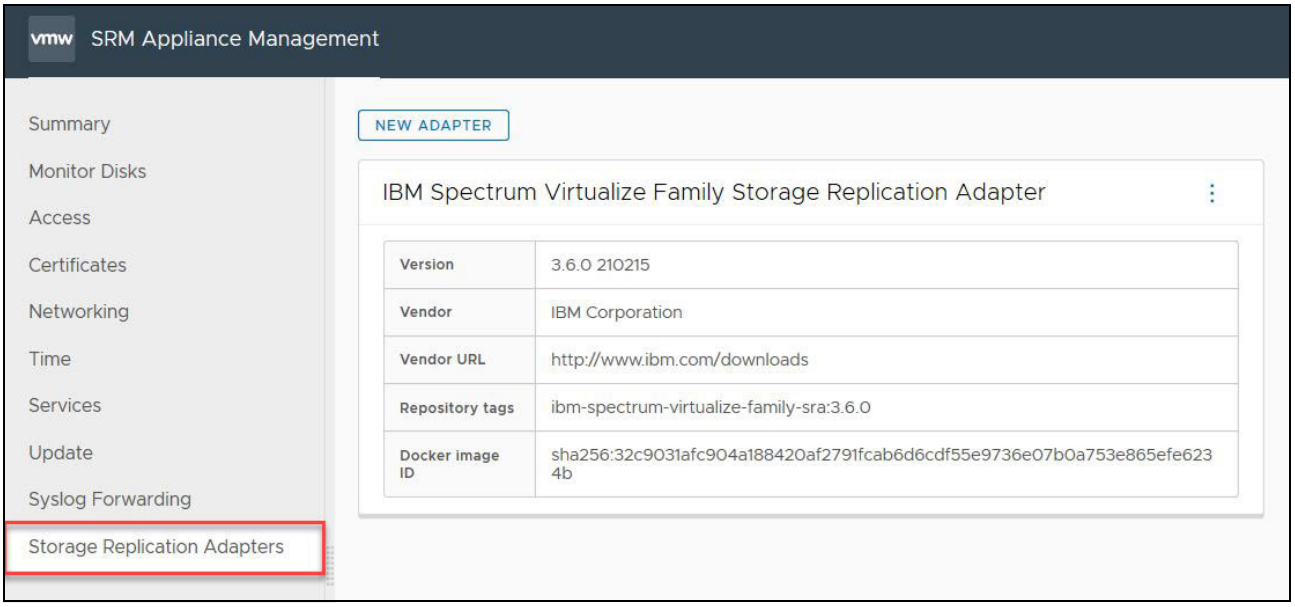


Figure 5-7 Storage Replication Adapters view

5. Log in to the vSphere Web Client.
6. Select **Site Recovery** → **Open Site Recovery**, select a site pair, and click **View Details**.
7. On the Site Pair tab, select **Configure** → **Array Based Replication** → **Storage Replication Adapters** → **RESCAN ADAPTERS**. After the rescan is complete, the Status field value is updated to OK (Figure 5-8 on page 115).

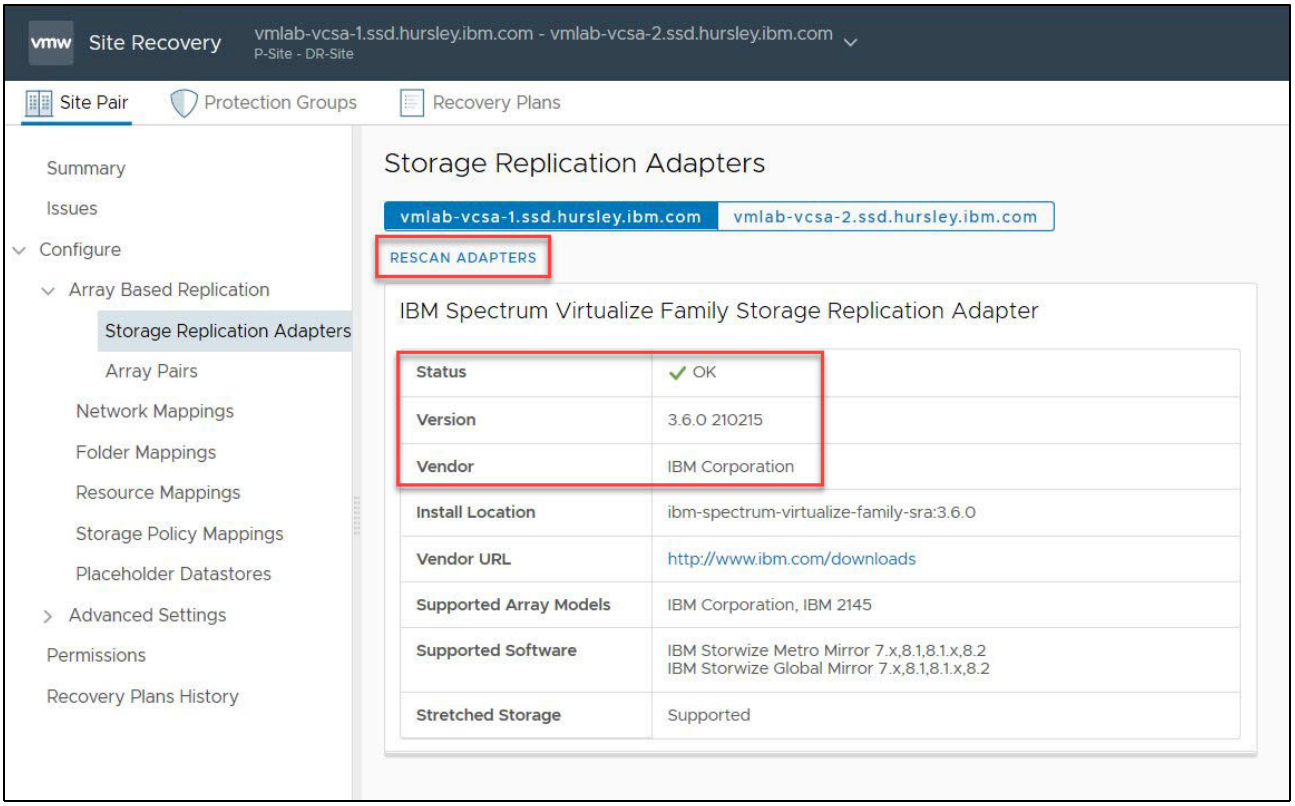


Figure 5-8 Site Recovery view

5.2.3 Storage Replication Adapter configuration and usage guide

When the storage replication adapter installation is complete, see Configuration.

Account for the following practices when working with SRA and VMware ARM:

- Create Metro or Global Mirror relationships between the source and target VDisks and add them to consistency groups, as explained in “Preparing the SRA environment” on page 110.
- Before you use the SRA, make sure that the Metro or Global Mirror relationships and consistency groups are in a consistent synchronized state.
- For Stretched Cluster, make sure that the two copies of a stretched volume are at different sites and that both copies are online.
- For IBM HyperSwap, make sure that the primary volume and secondary volume of a HyperSwap volume are online.
- Consider the following tips when you add an Array-air to the VMware SRM are as follows:
  - Enter the same Common Information Model Object Manager (CIMOM) address and Common Information Model (CIM) port as the primary IBM Spectrum Virtualize Cluster, if you are using IBM Stretched Cluster or IBM HyperSwap.
  - If $M{:}N$ topology is being used, enter any one of the IBM Spectrum Virtualize Clusters (N) on the remote site for the CIM address of the remote IBM Spectrum Virtualize Cluster. The term $M{:}N$ refers to the number of IBM Spectrum Virtualize Clusters on the local site (M) and remote site (N). Calculations are not derived from this term. M:N is used only to denote the replication process.

- IBM Spectrum Virtualize multiple-port configuration is supported by SRA when 1:1 IBM Spectrum Virtualize topology is used. With IBM Spectrum Virtualize M:N topology and multiple-port configuration, replication might not work because SRA communicates with IBM Spectrum Virtualize PORT 1.

For more information, see Adding an array pair to the VMware Site Recovery Manager.

- All volumes that participate in SRM and belong to the same remote copy consistency group are shown under a single local consistency group. To avoid data inconsistencies when adding replicated VDisks to the same VM or data store, all VDisks used by a single VM or application must be added to the same consistency group.

  If you plan to use VMware SRM to manage replicated volumes, use the Name Filter by using prefixes for the volume name. The volume names can be different for each site, but prefixes must be paired with the remote site array manager. For example, if the local site volume name is Win2019, and on the remote site it is mapped to Rec_Win2019, then you could enter the prefix `Pri` in the Name Filter field for the local site and the prefix `Rec` on the remote site. To use the Name Filter for the consistency group, enter the same names and prefixes at both the local and remote sites. For more information, see Filtering volumes and consistency groups by name.

- Consider the following items for managing data stores and consistency groups:
  - The data stores of one VM should be in the same consistency group.
  - The data store of the VM and the raw disk in the VM should be in the same consistency group.
  - You must have administrator privileges to install the SRM.
  - Set the appropriate timeout and rescan values in SRM for the recovery of many VMs.

5.3 IBM HyperSwap with VMware vSphere Metro Storage Cluster

The IBM Spectrum Virtualize storage system supports multiple VMware vSphere stretched-storage cluster solutions with HyperSwap to provide the following benefits:

- Highly available active-active vSphere data stores
- Workload mobility
- Cross-site automated load-balancing
- Enhanced downtime avoidance
- Disaster avoidance

In this document, the focus is on solutions that rely both on VMware vSphere Metro Storage Cluster (vMSC) and VMware SRM in relation to IBM Spectrum Virtualize.

5.3.1 IBM HyperSwap

IBM Spectrum Virtualize HyperSwap is a dual-site solution that provides continuous availability of data during planned and unplanned outages. If storage at either site goes offline, HyperSwap automatically fails over storage access to the system at the surviving site.

When you configure a system with a HyperSwap topology, the system is split between two sites for data recovery, migration, or high availability use cases. When a HyperSwap topology is configured, each node or enclosure, external storage system, and host in the system configuration must be assigned to one of the sites in the topology. Both node canisters of an I/O group must be at the same site. This site must be the same site of any external storage systems that provide the managed disks to that I/O group. When managed disks are added to storage pools, their site attributes must match. This requirement ensures that each copy in a HyperSwap volume is fully independent and spans multiple failure domains (Figure 5-9).

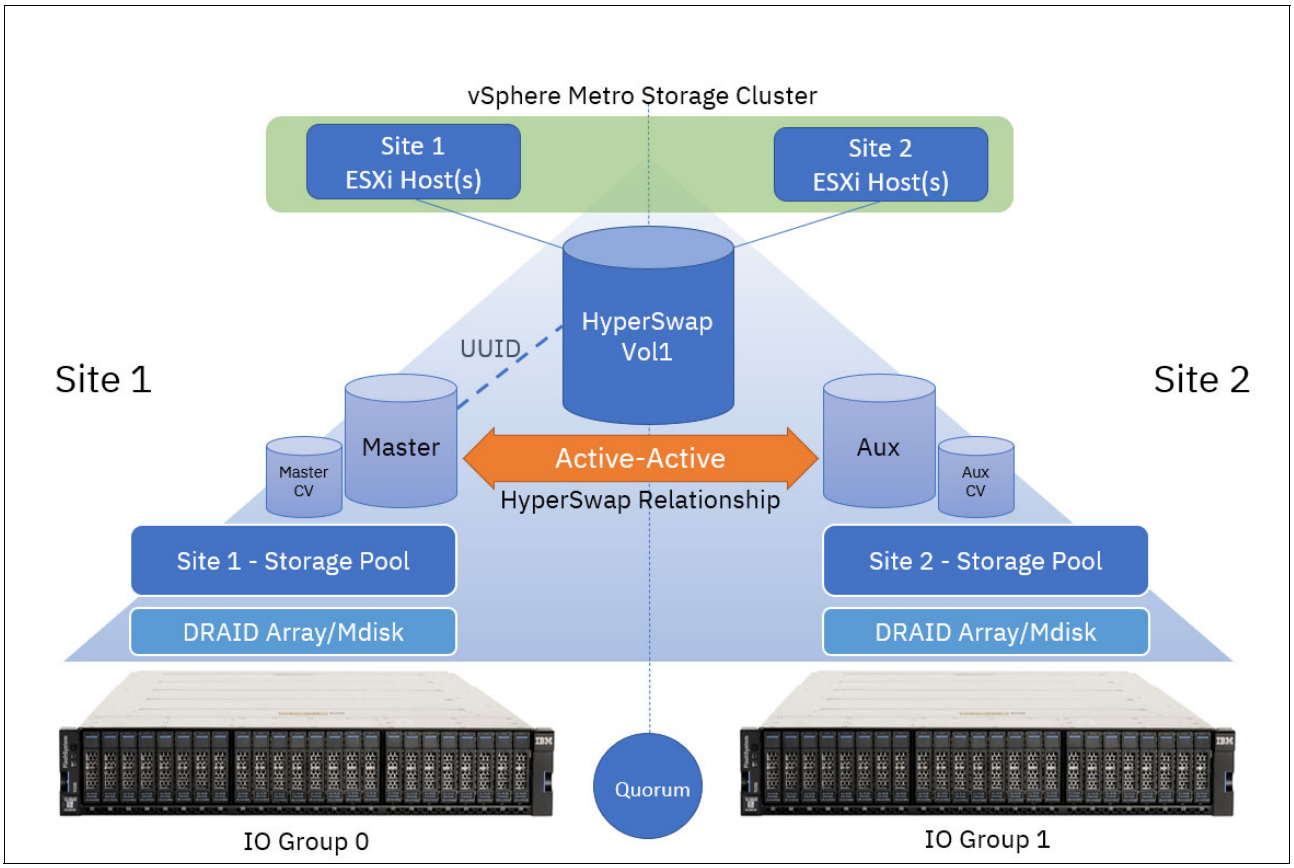


Figure 5-9 IBM HyperSwap

HyperSwap Volume is a group of volumes and remote copy relationship all working together to provide the active-active solution, and ensure that data is synchronized between sites.

A single HyperSwap Volume consists of the following items:

- A Master Volume and a Master Change Volume (CV) in one system site
- An Auxiliary Volume and an Auxiliary CV in the other system site

An active-active HyperSwap relationship exists between the Master and Auxiliary volumes to facilitate the data synchronization and replication between sites.

However, when you create a HyperSwap volume, the necessary components are created automatically, and the HyperSwap Volume can be managed as a single object.

Like a traditional Metro Mirror relationship, the active-active relationship attempts to keep the Master Volume and Auxiliary Volume synchronized while also servicing application I/O requests. The relationship uses the CVs as journaling volumes during the resynchronization process (Figure 5-10).

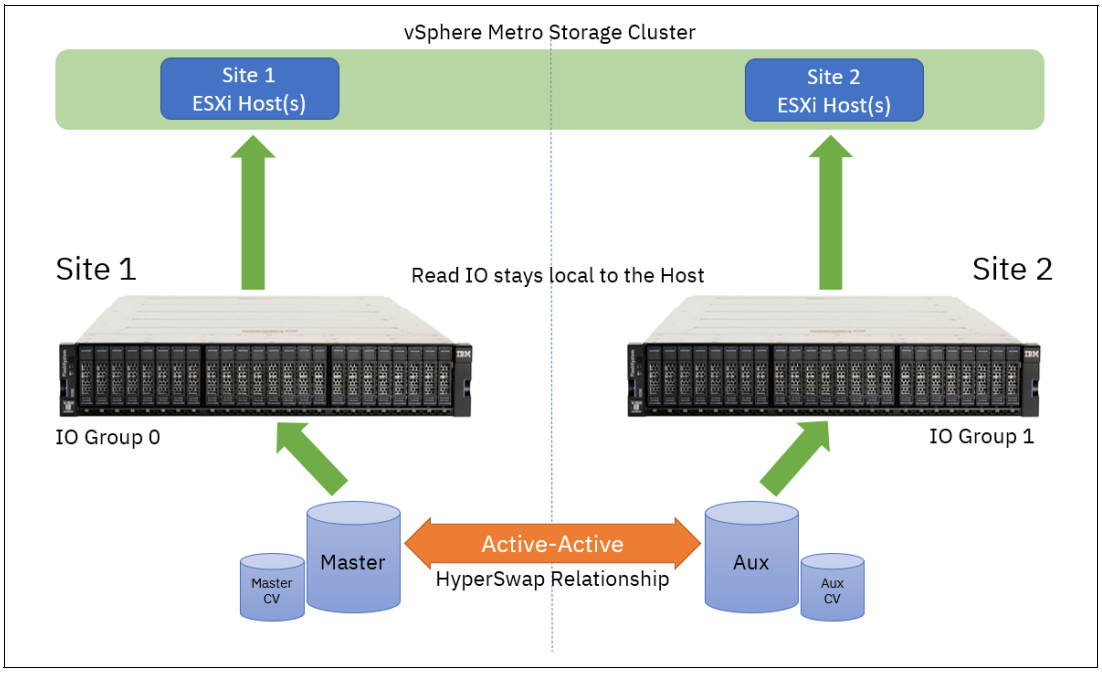


Figure 5-10 Read operations from hosts on either site are serviced by the local I/O group

The HyperSwap Volume always uses the unique identifier (UID) of the Master Volume. The HyperSwap Volume is assigned to the host by mapping only the Master Volume even though access to the Auxiliary Volume is ensured by the HyperSwap function. For each HyperSwap volume, hosts across both sites see a single volume that is presented from the storage system with the UID of the Master Volume.

To preserve application consistency when spanning multiple volumes, Consistency Groups can be created to keep a group of HyperSwap volumes consistent.

Cluster considerations

Consider the following tips when you work with HyperSwap and VMware vSphere Metro Storage Cluster (vMSC):

- One IBM Spectrum Virtualize-based storage system, which consists of at least two I/O groups. Each I/O group is at a different site. Both nodes of an I/O group are at the same site.
- HyperSwap-protected hosts on IBM Spectrum Virtualize must be connected to both storage nodes by using Internet Small Computer System Interface (iSCSI) or Fibre Channel.
- In addition to the two sites that are defined as failure domain, a third site is needed to house a quorum disk or IP quorum application.
- More system resources are used to support a fully independent cache on each site. This allows full performance even if one site is lost.

HyperSwap relationships

One site is considered as the Primary for each HyperSwap Volume or Consistency Group. This site is dynamically chosen according to the site that writes more data (more than 75% of write I/Os) to the volume or consistency group over a 20-minute period.

This role can change as follows:

- After a period of 20 minutes, if an I/O majority is detected in nodes on the non-Primary site
- Immediately, if a Primary-site outage occurs

Note: Low write-throughput rates do not trigger a direction switch to protect against unnecessary direction changes when experiencing a trivial workload.

While the I/O group on each site processes all reads from the hosts on that local site, any write requests must be replicated across the inter-site link, which incurs added latency. Writes to the primary site experience a latency of 1x the round-trip time (RTT) between the sites, however due to the initial forwarding process, writes to the non-primary site will experience 2x the round-trip time. This additional performance overhead should be considered when you provision storage for latency-sensitive applications.

These relationships automatically run and switch direction according to which copy or copies are online and up-to-date.

Relationships can be grouped into consistency groups, in the same way as other types of remote-copy relationships. The consistency groups fail over consistently as a group based on the state of all copies in the group. An image that can be used for disaster recovery is maintained at each site.

HyperSwap I/O flow

This section contains examples of the HyperSwap I/O flow. In the examples below, Site 1 is considered as Primary.

Read operations from ESXi Hosts at either site

Figure 5-11 illustrates the flow of read I/O requests from hosts at either site in the VMware vSphere Metro Storage Cluster.

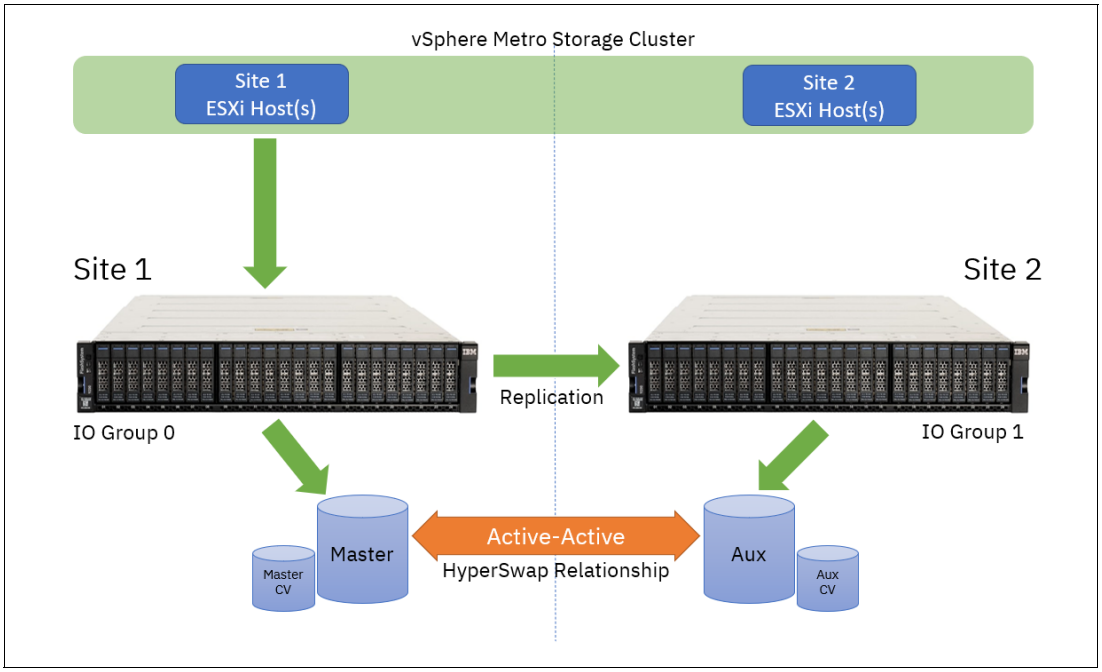


Figure 5-11 Write operations from hosts on either site are serviced by the local I/O group

Read I/O is facilitated by whichever I/O group is local to the requesting host, which prevents the I/O from having to transfer the long-distance link and incurring unnecessary latency.

Write operations from ESXi hosts at local site

When hosts write on the Primary site, and the host, node, controller, or managed disk (MDisk) site awareness is correctly configured, the write I/Os go straight to the Primary site volume and I/O Group. The write I/Os are then replicated to the Secondary site volume and I/O Group (Figure 5-12 on page 121).

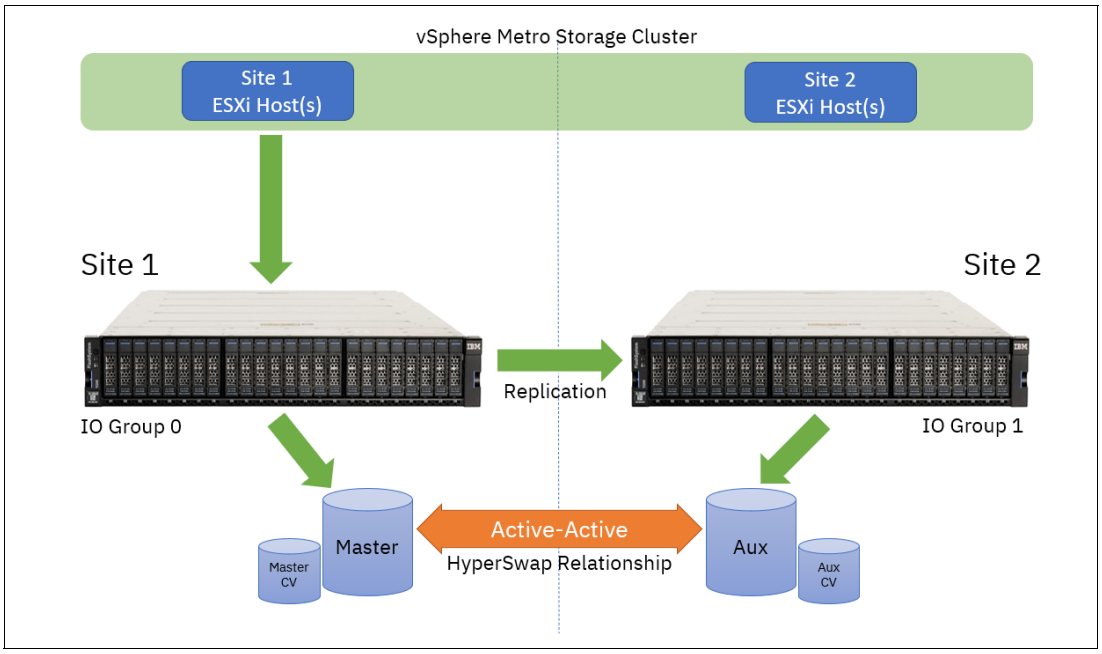


Figure 5-12 Write operations from hosts on primary are replicated

Write operations from ESXi hosts at remote site

In this scenario, a write to I/O Group 1 needs to be applied to both copies, but the replication code cannot handle that task on I/O Group 0 (because I/O Group 0 currently holds the Primary copy). The write data is initially transferred from the host into a data buffer on a node in I/O Group 1. The node in I/O Group 1 sends the write, both metadata and customer data, to a node in I/O Group 0 (Figure 5-13).

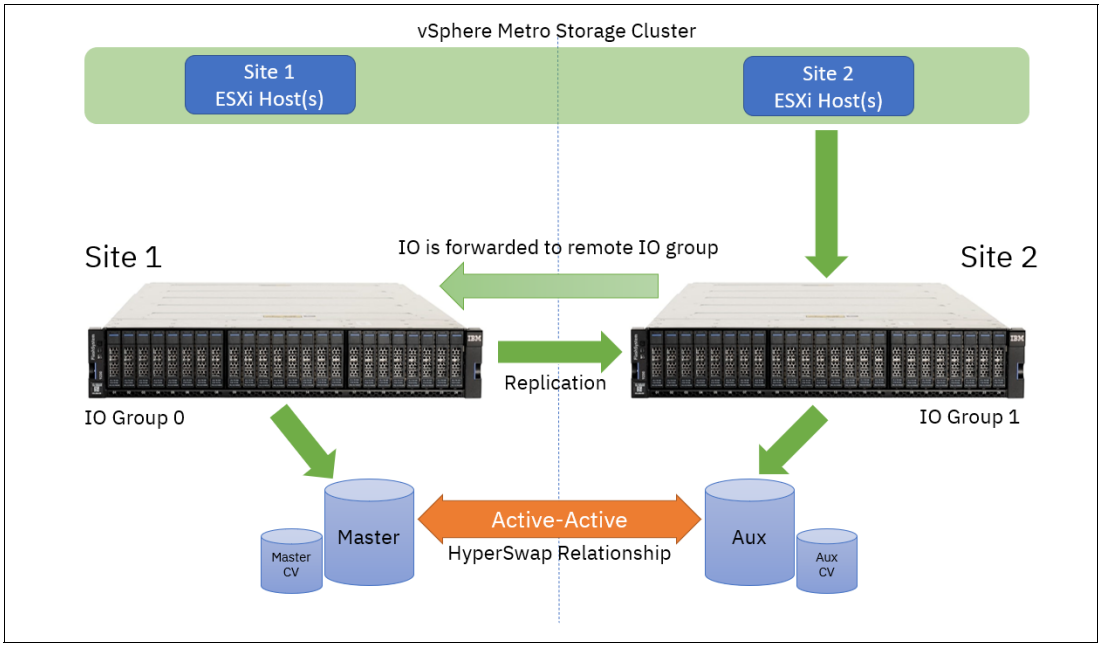


Figure 5-13 Write operations from hosts on Secondary are forwarded and replicated

On the node in I/O Group 0, the write is handled as though it were written directly to that I/O Group by a host. The replication code applies the write to the I/O Group 0 cache, and replicates it to I/O Group 1 to apply to the cache there, which means that writes to the secondary site have increased latency and use more bandwidth between the sites. However, sustained writes mainly to the secondary site over a 20-minute period will switch the direction of the HyperSwap relationship, which eliminates this impact.

Note: Whenever the direction of a HyperSwap relationship changes, there is a brief pause to all I/O requests to that volume. In most situations, this pause is less than 1 second. Where possible, consider how application workload to a single HyperSwap volume (or HyperSwap consistency group) across sites can reduce the likelihood of repeated direction changes.

5.3.2 VMware vSphere Metro Storage Cluster

VMware vMSC is a unique storage-related feature and configuration that combines replication with array-based clustering that allows a single cluster to operate across geographically separate data centers. This capability allows two separated data centers to operate as a single cluster that provides significant benefits when maintaining data availability during both planned and unplanned downtimes.

IBM Spectrum Virtualize facilitates vMSC with the ability to create single storage cluster that spans both sites, such that a data store must be accessible in both locations. In other words, the data store must be able to read and be written to simultaneously from both sites by using the HyperSwap feature of IBM Spectrum Virtualize. In a failure, the vSphere hosts are able to continue read and write access to the data store from either location seamlessly and with no impact on ongoing storage operations.

Uniform versus Non-Uniform vMSC configurations

There are two ways in which a vMSC can be configured. The following terms refer to the different configurations of host connectivity across sites:

- *Non-Uniform host access configuration* is where an ESXi host has storage connectivity to only the storage system local to that site.
- *Uniform host access configuration* is where an ESXi host has storage connectivity to both local and remote storage systems.

For every volume presented from the IBM FlashSystem, a preferred node is automatically elected. To evenly distribute the workload across the IBM FlashSystem upon volume creation, the preferred node usually alternates between each node in the I/O group.

When you map a volume to a host object and rescan the host bus adapter (HBA) on the ESXi host, ESXi automatically identifies the available paths to both nodes in the I/O group, as follows:

- The paths to the preferred node for each volume are identified as the “Active/Optimized paths”.
- The paths to the non-preferred node are identified as “Active/Non-Optimized paths”.

By default, ESXi uses the Round-Robin path selection policy (PSP) to distribute I/O as follows:

- Over any available “Active/Optimized paths” to the preferred node.
- Only fail over to “Active/Non-Optimized paths” if available paths to the preferred node do not exist.

Non-Uniform configuration

In Non-Uniform vMSC implementations, ESXi hosts use Small Computer System Interface (SCSI) Asymmetric Logical Unit Access (ALUA) states to identify Active/Optimized paths to the preferred node in the local I/O group and Active/Non-Optimized paths to the partner node. The host has no visibility of the storage system at the remote site (Figure 5-14).

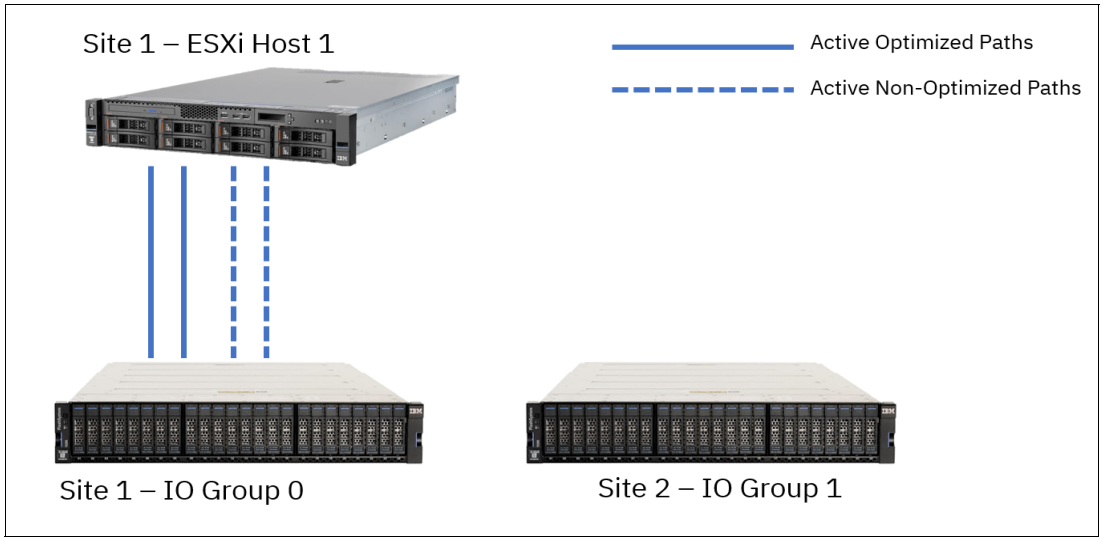


Figure 5-14 Non-Uniform host access vMSC

With Non-Uniform vMSC environments, if a storage failure occurs at the local site, the ESXi hosts lose access to the storage because paths are not available to the storage system at the remote site. However, this architecture might be useful when you run clustered applications like Microsoft SQL or Microsoft Exchange with servers that are located at each site. It might be preferable to have a clustered application fail over so that an application can continue to run with locally available storage.

Uniform configuration

In Uniform vMSC implementation, ESXi hosts also uses SCSI ALUA states to identify Active/Optimized paths to the preferred node and Active/Non-Optimized paths to the partner node in the local I/O group. Extra paths to the remote I/O group are automatically detected as “Active/Non-Optimized”. ESXi uses the Optimized paths to the local preferred node where possible, and only fail over to the Non-Optimized paths if there are no available paths to the preferred node (Figure 5-15).

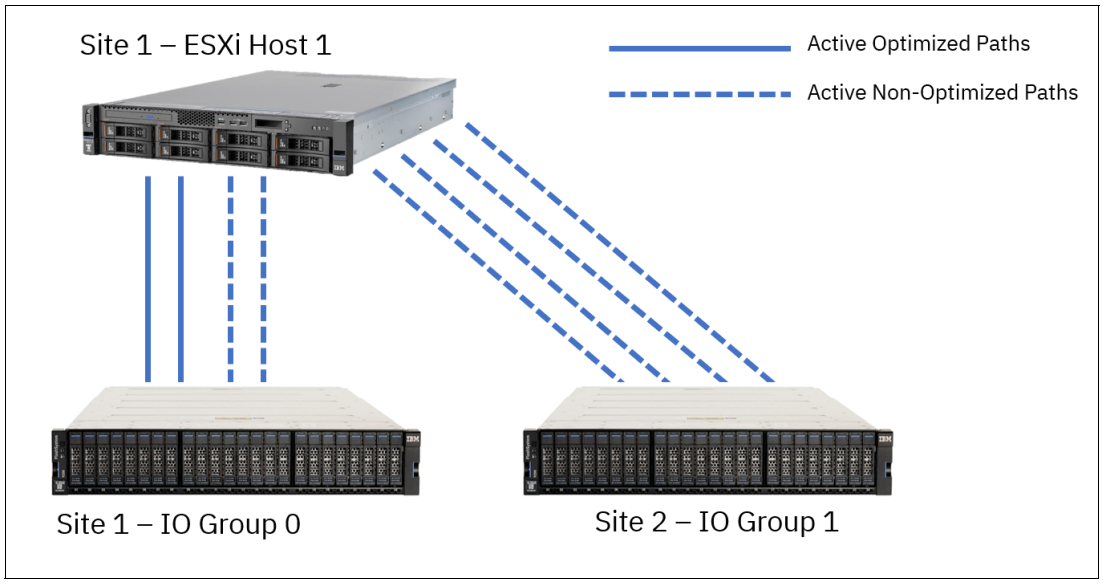


Figure 5-15 Uniform host access vMSC

5.3.3 IBM HyperSwap with VMware vSphere Metro Storage

Given the performance characteristics of a HyperSwap topology, it might be beneficial to review how HyperSwap volumes can be used by ESXi hosts to ensure optimal performance, data continuity, and reliability.

Virtual Machine File System data store provisioning

As shown in Figure 5-16 on page 125, the following three different data store architectures can be considered when provisioning and consuming Virtual Machine File System (VMFS) data stores in a HyperSwap vMSC environment:

- Traditional Disaster Recovery
- Mixed access
- Alternating access

Each of these models applies at a single data store level. Therefore, it is possible to incorporate all three methods in an IBM FlashSystem HyperSwap architecture.

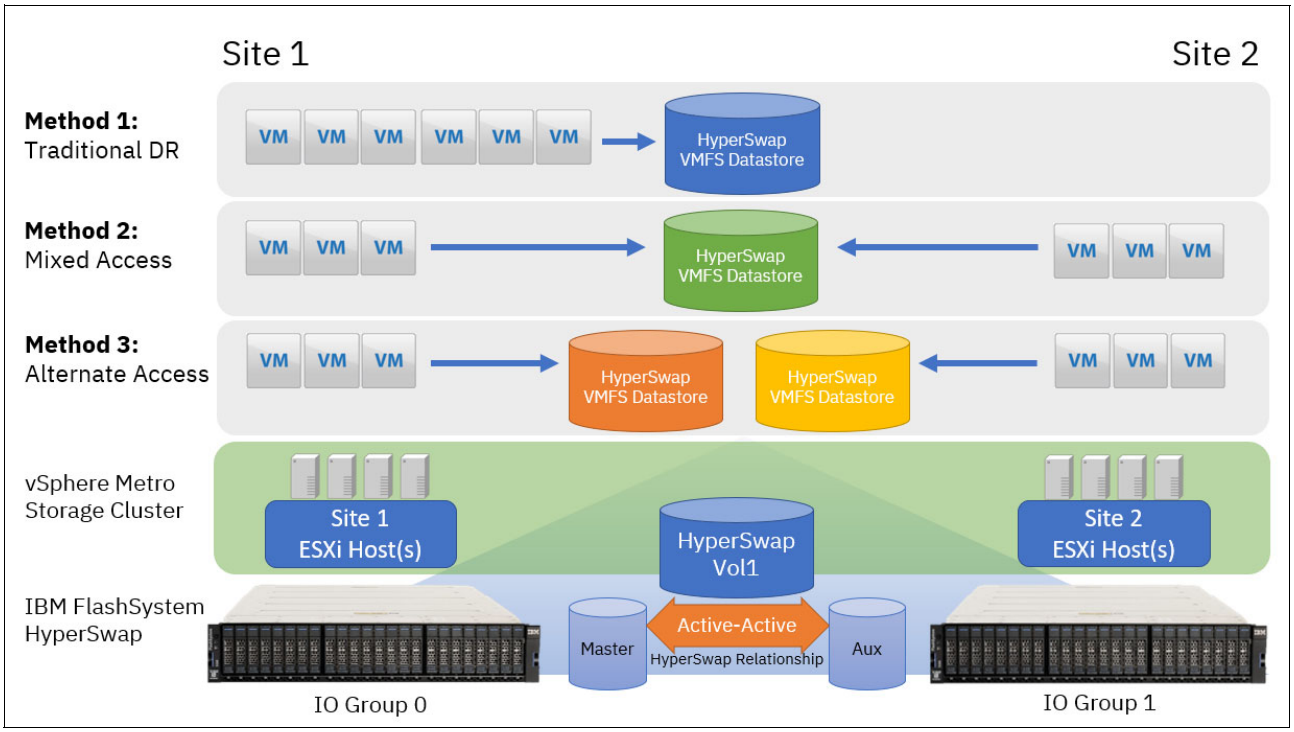


Figure 5-16 Three data store architectures

Method 1: Traditional Disaster Recovery

This model can be conceptualized as a traditional DR configuration. In normal operating conditions, all VMs are running from a single site, and only failover to ESXi hosts at the remote site in a system or storage outage at Site 1. In this scenario, the direction of the HyperSwap relationship for the VMFS data store is static because I/O is not running at Site 2. In a site outage, all the VMs are powered up on the ESXi hosts at Site 2 without requiring intervention on the storage system to enable access to the volumes.

The negative aspect of this configuration is that there can be many ESXi servers and storage resource at Site 2 that are idle.

Method 2: Mixed Access

In this scenario, the number of VMs for a data store is split between both sites. ESXi servers at both sites are used, which maximizes compute resource at both sites. However, as discussed in “HyperSwap I/O flow” on page 119, any writes being performed at the non-primary site incur more latency when traversing the inter-site link. Potentially half of the vSphere infrastructure can then be exposed to extra, unnecessary latency overhead.

In addition, if the workload is unregulated, the direction of the HyperSwap relationship can be swapped repeatedly, which generates more I/O pauses for each change of direction.

Method 3: Alternative Access

This scenario requires a minor additional management overhead to provision storage and maintain the vSphere environment. However, this scenario likely enables optimal storage performance and maximum realization of available compute resource.

Consider creating multiple VMFS data stores where the primary associated site alternates between Site 1 and Site 2, and keep VM-storage resources local to hosts that are running the VMs.

An example configuration is as follows:

- When creating data stores, use odd-numbered data stores to have a site preference of Site 1, and even-numbered data stores designated to Site 2.
- Migrate the compute resource for half of the VMs over to ESXi hosts on Site 2 to create an even distribution of workload between the two sites.
- For VMs running on ESXi hosts at Site 1, ensure the VMDKs are provisioned on odd-numbered data stores.

  For VMs running on ESXi hosts at Site 2, ensure the VMDKs are provisioned on the even-numbered data stores.
- Create Host/VM Groups and Host/VM Rules in vSphere to ensure that Distributed Resource Scheduler (DRS) can function correctly to redistribute VMs across hosts within a site if required, but still enable failover in an outage.

As detailed in “Write operations from ESXi hosts at remote site” on page 121, Alternative Access Method ensures that instead of any write I/Os at either site having to be forwarded over the inter-site link before being replicated to the remote site, they will be serviced by the I/O group at the site where the I/O originated. This method reduces the overall latency and increases the performance and throughput.

The intention is for a given HyperSwap volume or consistency group to keep VM I/O workloads local to the host running the VM, which minimizes the workloads being driven from a host at the non-primary site.

In a site outage at either site, vSphere high availability (HA) automatically recover the VMs on the surviving site.

For more information about DRS Host/VM groups and rules, see Create a Host DRS Group.

VMware Distributed Resource Scheduler

VMware DRS capabilities bring efficiency in the management of workloads through grouping VMware ESXi hosts into resource clusters to separate computing requests to different sites or failure domain. Employing VMware DRS in an active-active storage solution (such as HyperSwap) provides highly available resources to your workloads.

In an ideal HyperSwap environment, you do not want VMs to move to the other site. Rather, you want VMs to move to the other site only in a site failure, or intentionally to balance workloads and achieve optimal performance.

ESX hostnames

Create a logical naming convention that allows you to quickly identify which site a host is in. For example, the site can be included in the chosen naming convention or you can choose a numbering system that reflects the location (for example odd hosts are in site one). The naming convention makes the designing and the day-to-to running of your system easier.

Data locality and host affinity rules

Ideally, hosts should access data from their local storage array to improve response time. To ensure that this situation is the case, use VMware affinity rules to define the preferred site for VMs to run from a local logical unit number (LUN).

Logically name the LUNs with their home sites

This task is not a *must* and some might argue that they want the flexibility to move LUNs between data centers, but it makes it easier for business as usual (BAU) staff to track which are local data stores.

VMware vSphere host multipathing ensures that VMs that are running continue to operate during various failure scenarios. Table 5-1 outlines the tested and supported failure scenarios when using SVC or IBM FlashSystem Family HyperSwap function, and VMware vSphere Metro Storage Cluster (vMSC).

Table 5-1 IBM Spectrum Virtualize HyperSwap and VMware vSphere Metro Storage Cluster supported failure scenarios

Failure scenario	HyperSwap behavior	VMware HA impact
Path failure: SVC or IBM FlashSystem Family Back-End (BE) Port	Single path failure between SVC or IBM FlashSystem Family control enclosure and flash enclosure. No impact on HyperSwap.	No impact.
Path failure: SVC or IBM FlashSystem Family Front-End (FE) Port	Single path failure between SVC or IBM FlashSystem Family control enclosure and vSphere host. vSphere host uses alternative paths.	No impact.
BE flash enclosure failure at Site-1	SVC or IBM FlashSystem Family continues to operate from the volume copy at Site-2. When the flash enclosure at Site-1 is available, HyperSwap synchronizes the copies.	No impact.
BE flash enclosure failure at Site-2	Same behavior as failure at Site-1.	No impact.
SVC or IBM FlashSystem Family control enclosure failure	SVC or IBM FlashSystem Family continues to provide access to all volumes through the other control enclosures.	No impact.
Complete Site-1 failure (The failure includes all vSphere hosts and SVC or IBM FlashSystem Family controllers at Site-1)	SVC or IBM FlashSystem Family continues to provide access to all volumes through the control enclosures at Site 2. When the control enclosures at Site-1 are restored, the volume copies are synchronized.	VMs running on vSphere hosts at the failed site are impacted. VMware HA automatically restarts them on vSphere hosts at Site-2.
Complete site 2 failure	Same behavior as a failure of Site-1.	Same behavior as a failure of Site-1.
Multiple vSphere host failures Power Off	No impact.	VMware HA automatically restarts the VMs on available ESXi hosts in the VMware HA cluster.
Multiple vSphere host failures, network disconnect	No impact.	VMware HA continues to use the data store heartbeat to exchange cluster heartbeats. No impact.
SVC or IBM FlashSystem Family inter-site link failure, vSphere cluster management network failure	SVC or IBM FlashSystem Family active quorum is used to prevent a split-brain scenario by coordinating one I/O group to remain servicing I/O to the volumes. The other I/O group goes offline.	vSphere hosts continue to access volumes through the remaining I/O group. No impact.
Active SVC or IBM FlashSystem Family quorum disk failure	No impact to volume access. A secondary quorum disk is assigned upon failure of the active quorum.	No impact.

Failure scenario	HyperSwap behavior	VMware HA impact
vSphere host isolation	No impact.	HA event dependent upon isolation response rules configured for the vSphere cluster. VMs can be left running, or rules can dictate for VMs to shut down and restart on other hosts in the cluster.
vCenter server failure	No impact.	No impact to running VMs or VMware HA. VMware DRS function is affected until vCenter access is restored.

6

Embedded VASA Provider for Virtual Volumes

This chapter describes the implementation of the Embedded VASA Provider feature and includes the following sections:

6.1 Overview

VMware vSphere Virtual Volumes require the VASA application programming interfaces (APIs) to function, which are facilitated by a VASA Provider (also known as a storage provider). Historically, IBM storage systems that are powered by IBM Spectrum Virtualize required a separate, external application to fulfill the VASA provider role, that is, IBM Spectrum Connect. This application was installed in a Linux environment and required TCP/IP connectivity between the VMware vSphere environment and the management interface of the IBM Spectrum Virtualize storage system.

However, this external IBM Spectrum Connect component introduces an additional administrative burden in VMware vSphere Virtual Volume (vVol) environments because it requires the following items:

- Dedicated virtual machines (VMs).
- Installation of a supported Linux distribution to host the application.
- Installation of several prerequisite packages and services.
- On-going support to secure, update, and back up the VM or application.
- A separate management interface.
- Additional complexity in configuring and maintaining the environment.

Starting with IBM Spectrum Virtualize firmware 8.5.1.0 or later, the VASA Provider function has been incorporated natively in to the configuration node of the cluster to simplify the overall architecture of a vVol environment. This feature is referred to as the *Embedded VASA Provider*.

6.1.1 Supported platforms for the Embedded VASA Provider

Of all the hardware platforms that support the 8.5.1.0 firmware, Table 6-1 shows the ones that support the Embedded VASA Provider feature of IBM Spectrum Virtualize.

Table 6-1 Supported platforms for the Embedded VASA Provider

Platform name	Supports Embedded VASA Provider
IBM FlashSystem 5015	No
IBM FlashSystem 5035	No
IBM FlashSystem 5100	Yes
IBM FlashSystem 5200	Yes (minimum of 64 GB of memory required)
IBM Storwize V7000	Yes (Gen3 only)
IBM FlashSystem 7200	Yes
IBM FlashSystem 7300	Yes
IBM FlashSystem 9110	Yes
IBM FlashSystem 9150	Yes
IBM FlashSystem 9200	Yes
IBM FlashSystem 9500	Yes

Platform name	Supports Embedded VASA Provider
IBM SAN Volume Controller - 2145-SV2	Yes
IBM SAN Volume Controller - 2145-SV3	Yes
IBM SAN Volume Controller - 2145-SA2	Yes

6.1.2 Feature comparison between the Embedded VASA Provider and IBM Spectrum Connect

In the initial release of the Embedded VASA Provider, there are several limitations that might restrict functions when compared to existing vVol support that uses IBM Spectrum Connect. Evaluate the requirements of your environment before selecting a VASA Provider.

Table 6-2 shows the feature comparison.

Table 6-2 Feature comparison between the Embedded VASA Provider and IBM Spectrum Connect

Item	IBM Spectrum Connect	Embedded VASA Provider
Enhanced Stretched Cluster	Yes	No
vVol mirroring	Yes	No
Multiple vCenter connectivity	Yes, with multiple IBM Spectrum Connect instances	No
Multiple vVol data stores	Yes	Yes (command-line interface (CLI) configuration only)

6.2 System prerequisites

In this section, we describe the system prerequisites for implementing the Embedded VASA Provider feature.

6.2.1 Preparing IBM Spectrum Virtualize for vVol

Before vVols can be enabled in the GUI, there are several prerequisites that must be completed. When you select **Settings** → **vVols GUI** and the window opens, you see that there are many checks that are being evaluated.

The Enable vVols toggle switch is disabled until the following tasks are completed, as shown in Figure 6-1.

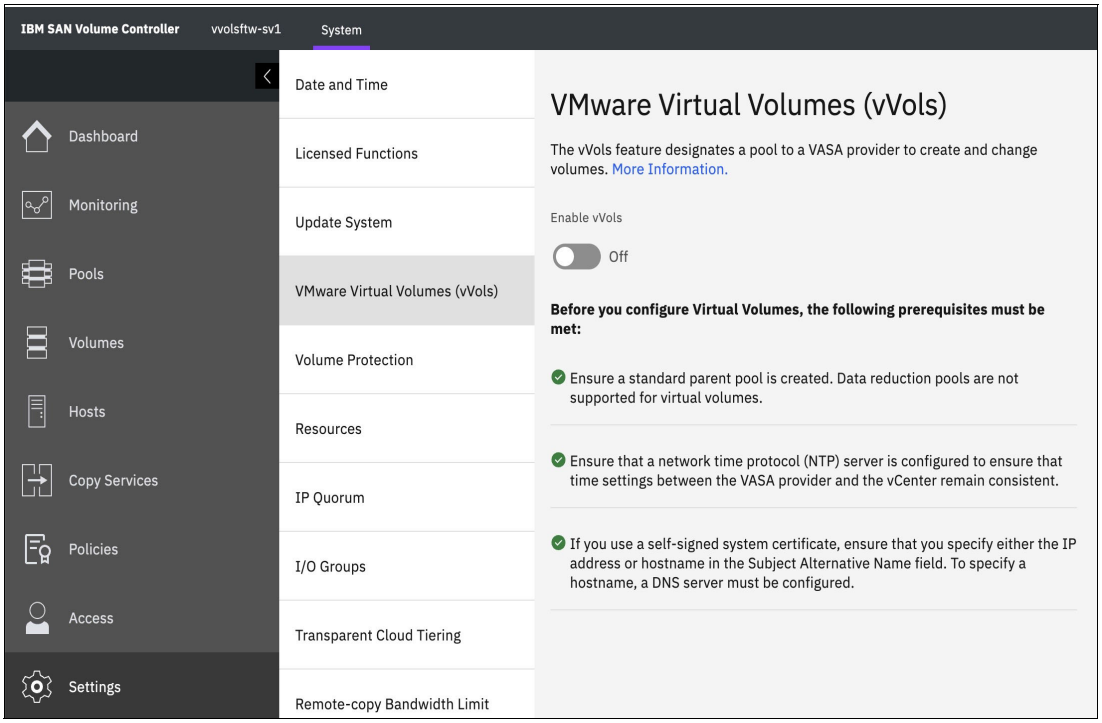


Figure 6-1 vVols prerequisites

- The system must have a standard pool with storage capacity that is allocated.

Note: Data reduction pools (DRPs) are not supported for either the metadata volume disk (VDisk) or individual vVols.

- The system must be configured with a Network Time Protocol (NTP) server to ensure that the date and time are consistent with the VMware infrastructure.
- The system must be configured with a certificate with a defined Subject Alternative Name value.

Note: Hosts that require access to the vVol data store must be configured with the vVol host type.

6.2.2 Configuring the NTP server

To configure the NTP server on the system, complete the following steps:

1. Go to the **Settings** → **System** window in the GUI, and select **Date and time**, as shown in Figure 6-2.

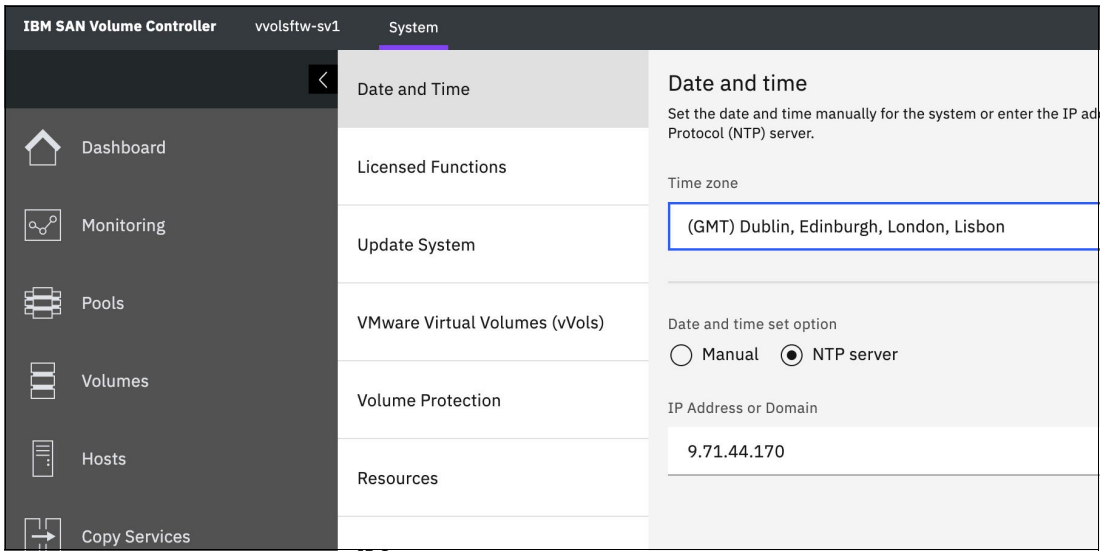


Figure 6-2 NTP time zone

2. Select the time zone.
3. Specify NTP server and enter the IP address or fully qualified domain name (FQDN) for the NTP server within your environment.

Note: If you use an FQDN or DNS name for the NTP server, you must ensure that a DNS server is configured in the system. To configure DNS servers, select **Settings** → **Network** and select **DNS**.

4. Click **Save** to complete the change.

6.2.3 Configuring a storage system certificate

Rather than using simple username and password credentials, the Embedded VASA Provider uses SSL certificates for secured communication between vSphere and the IBM Spectrum Virtualize storage system.

When you use a self-signed certificate, you must update the Subject Alternative Name field in the certificate before registering the Embedded VASA Provider within vCenter. When you use a signed certificate, this value is likely defined.

To configure a storage system certificate, complete the following steps:

1. Confirm whether this value is defined on the system certificate by connecting to the web user interface for the storage system and inspecting the certificate information in the browser window. Review the certificate information in the browser window, as shown in Figure 6-3.

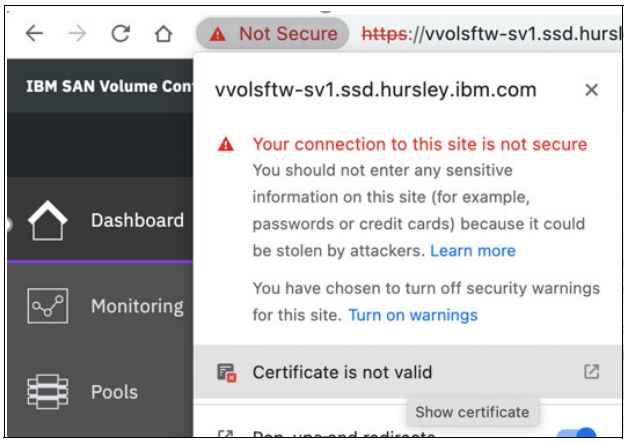


Figure 6-3 Reviewing the certificate information

2. Expand the details and review the Subject Alternative Name value, as shown in Figure 6-4.

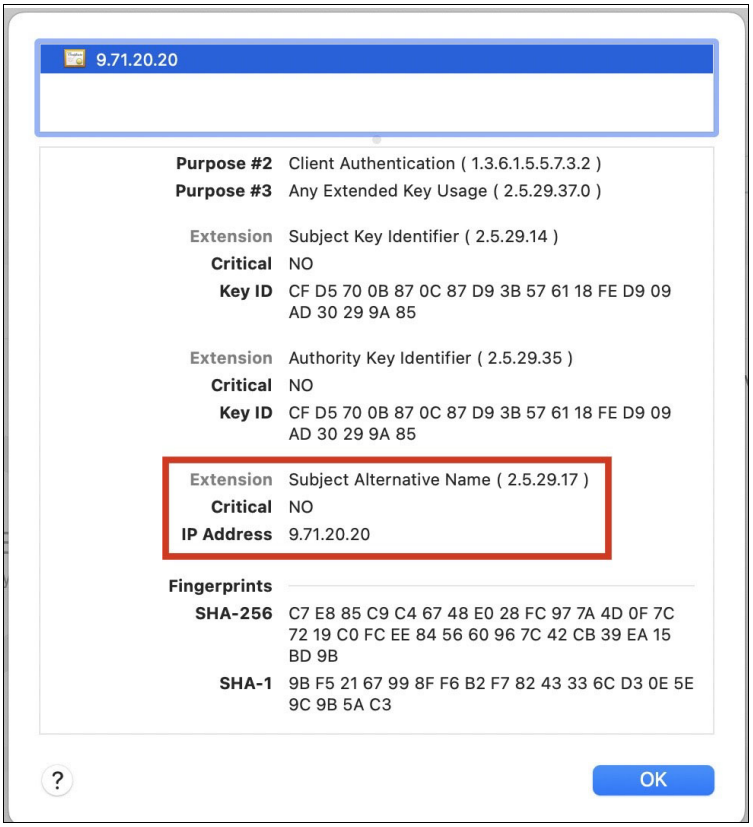


Figure 6-4 Reviewing the Subject Alternative Name value

3. Alternatively, run the **lssystemcert** command, which shows the following output:

```
IBM_2145:vvolsftw-sv1:superuser>lssystemcert | grep -A 1
"Subject Alternative Name"
            X509v3 Subject Alternative Name:
                IP Address:9.71.20.20
```

4. If no Subject Alternative Name field is defined, update the self-signed certificate. To do this task, select **Settings** → **Security** and select **Secure Communications**, as shown in Figure 6-5.

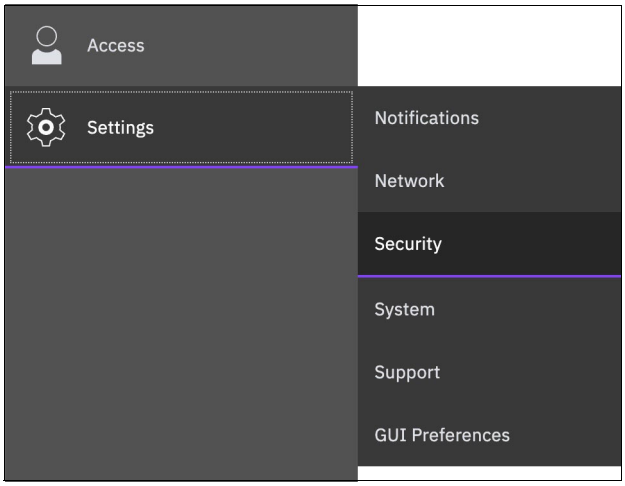


Figure 6-5 Secure Communications

5. Click **Update Certificate**, as shown in Figure 6-6.

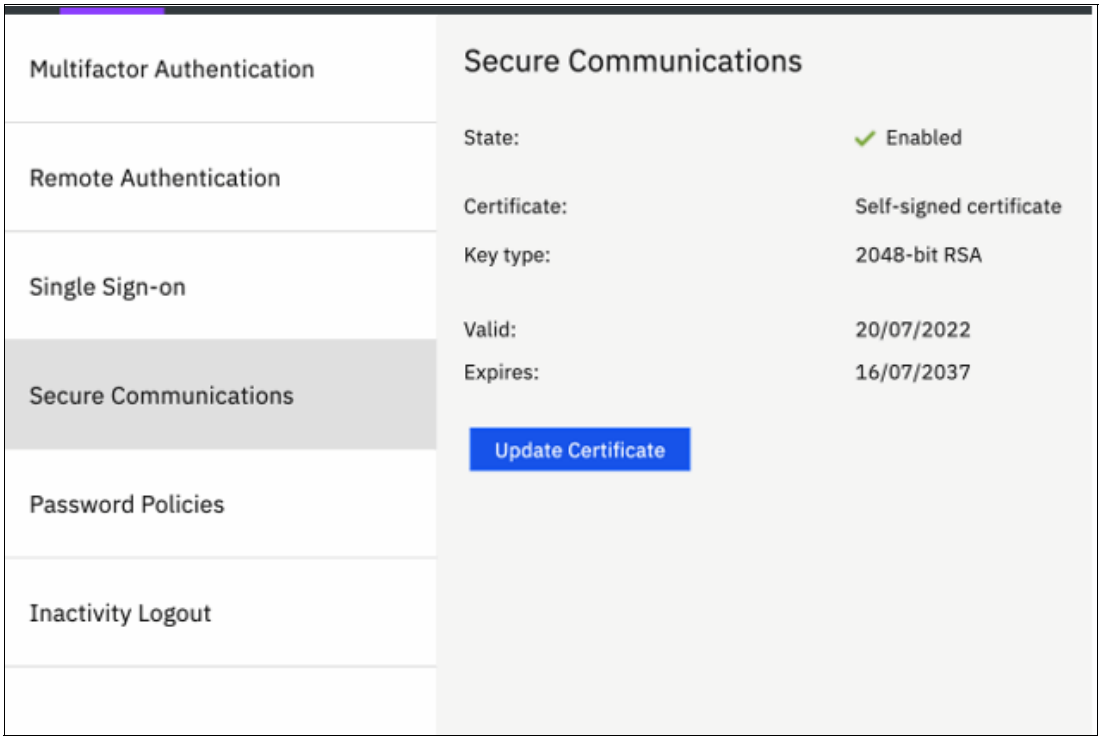


Figure 6-6 Update Certificate

6. Complete the certificate notification and ensure that a Subject Alternative Name value is defined. This value can either be an IP address, DNS name, or FQDN, but the specified Subject Alternative Name extension must resolve to the same host as the VASA provider's advertised IP address, hostname, or FQDN, as shown in Figure 6-7.

Note: If you use an FQDN or DNS name, you must ensure that a DNS server is configured in the system. To configure DNS servers, select **Settings** → **Network and select DNS**.

Figure 6-7 Update Certificate

During this step, the cluster IP is unavailable for a few minutes while the new security settings are applied. After a few minutes, you might need to refresh your browser window, and then you are prompted to accept the new self-signed certificate.

6.2.4 Preparing Elastic Sky X integrated hosts for vVol connectivity

Any Elastic Sky X integrated (ESXi) hosts that require access to a vVol data store must be defined as a vVol host type in the storage system.

Note: As a best practice, create a single host object in IBM Spectrum Virtualize to represent each physical ESXi server in the configuration. When you use clustered host environments, for example, when multiple ESXi hosts are part of a vSphere cluster, you should use IBM Spectrum Virtualize Host Cluster objects.

Complete the following steps:

1. To create a Host Cluster, select **Hosts** → **Host Clusters** in the management GUI, and select **Create Host Cluster**, as shown in Figure 6-8.

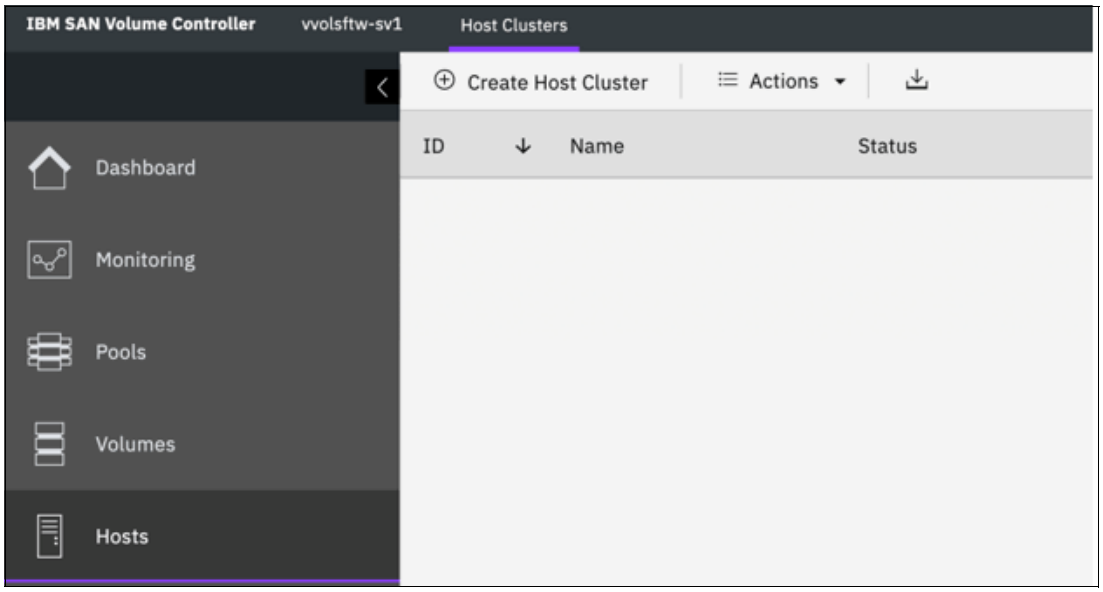


Figure 6-8 Create Host Cluster

2. Specify a name for the host cluster object and click **Next**. To simplify troubleshooting, consider using the same name for the Host Cluster object as is defined on the vSphere Cluster within vCenter, as shown in Figure 6-9.

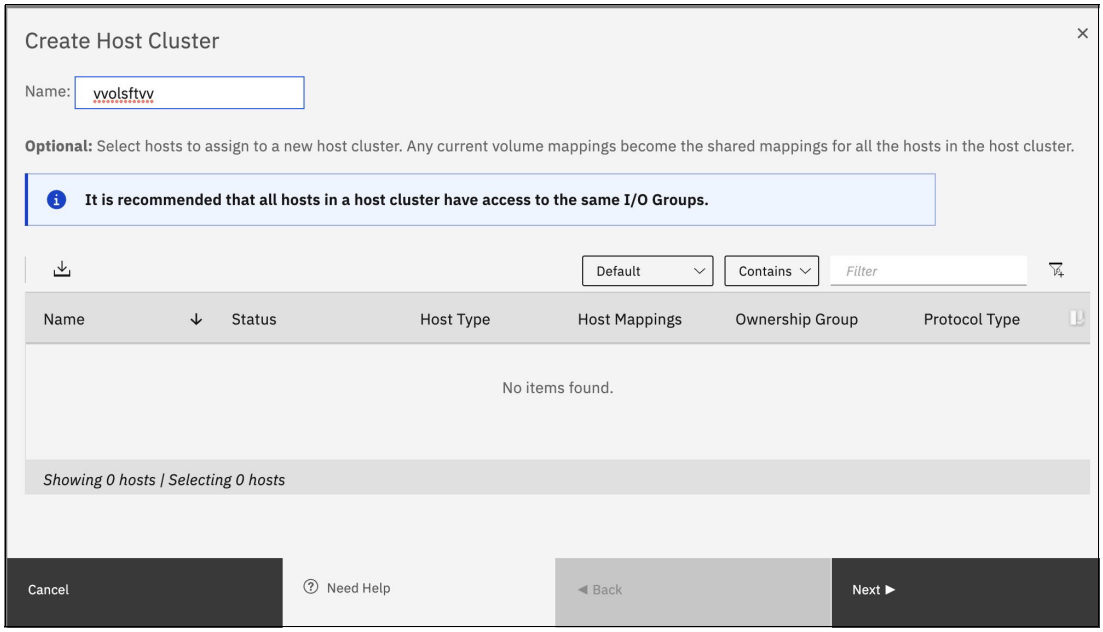


Figure 6-9 Create Host Cluster

3. Review the summary window, and click **Make Host Cluster**, as shown in Figure 6-10.

Figure 6-10 Make Host Cluster

4. When creating a host object, ensure that it is defined with the Host Type of vVol. To do this task, access the Hosts view in the GUI by selecting **Hosts** → **Hosts**, and clicking **Add Host**, as shown in Figure 6-11.

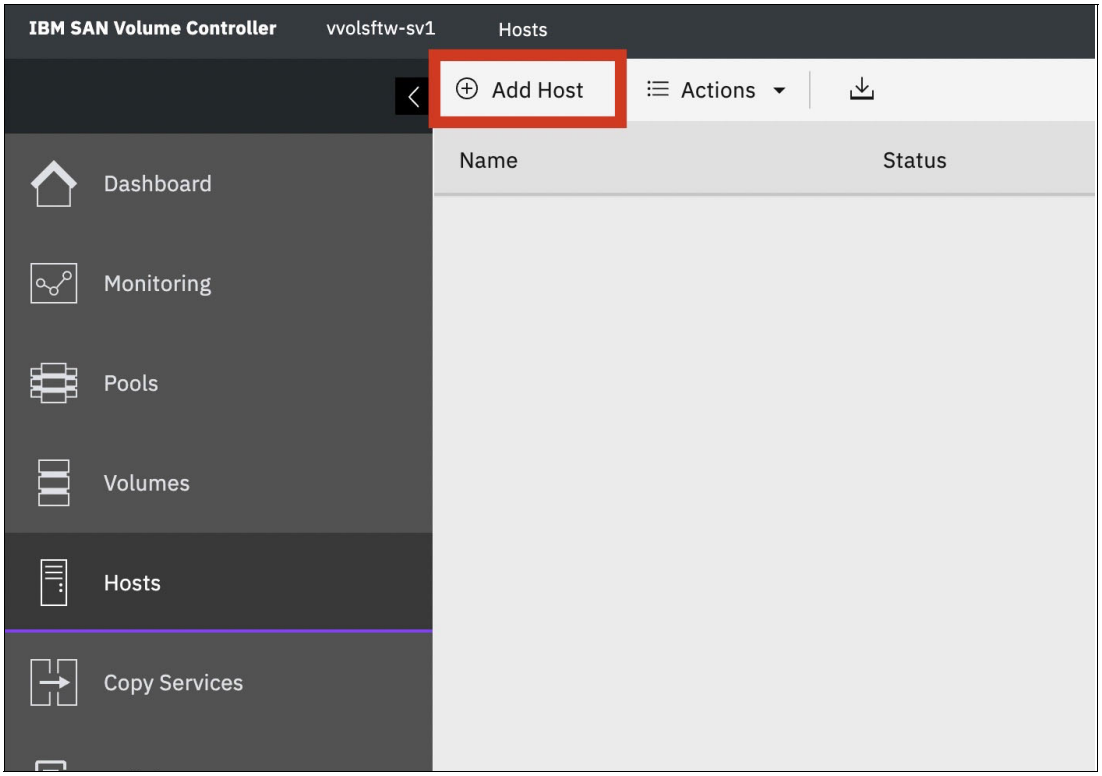


Figure 6-11 Add Host

5. Enter a descriptive name, select the **Host Port** definitions, and define the Host Type as **vVol**. Consider naming the host object in IBM Spectrum Virtualize with the same name as the one that the ESXi host uses in vCenter, as shown in Figure 6-12.

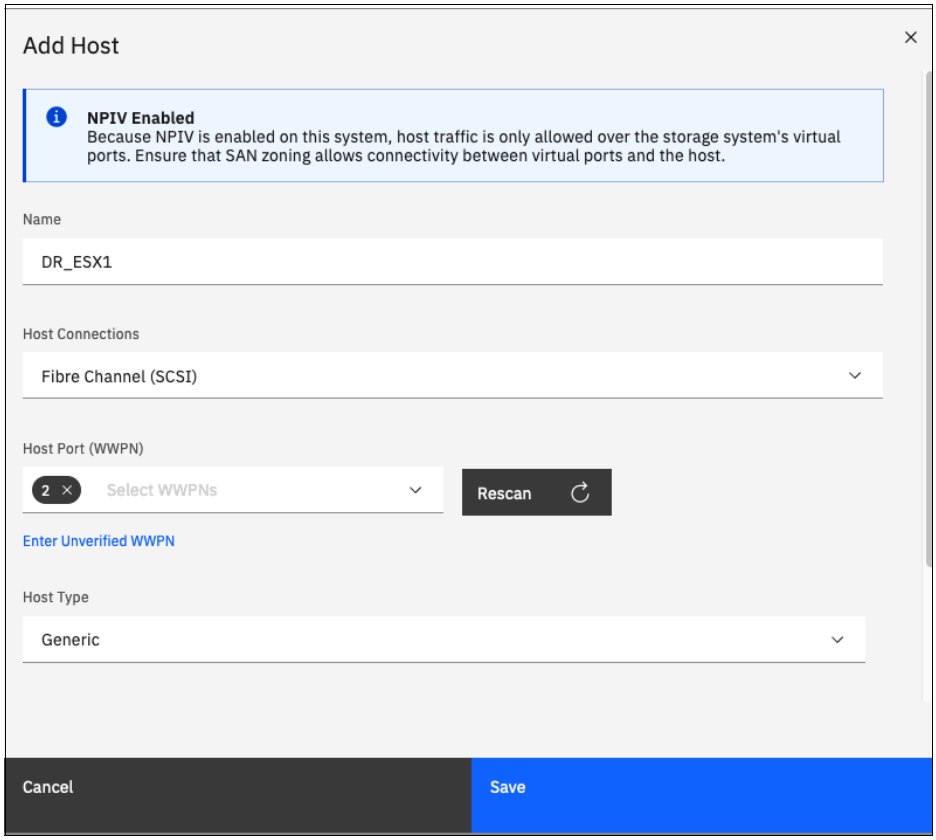


Figure 6-12 Entering the details

6. If the host object will be a member of a Host Cluster, expand the advanced view and select the Host Cluster from the list, as shown in Figure 6-13 on page 141.

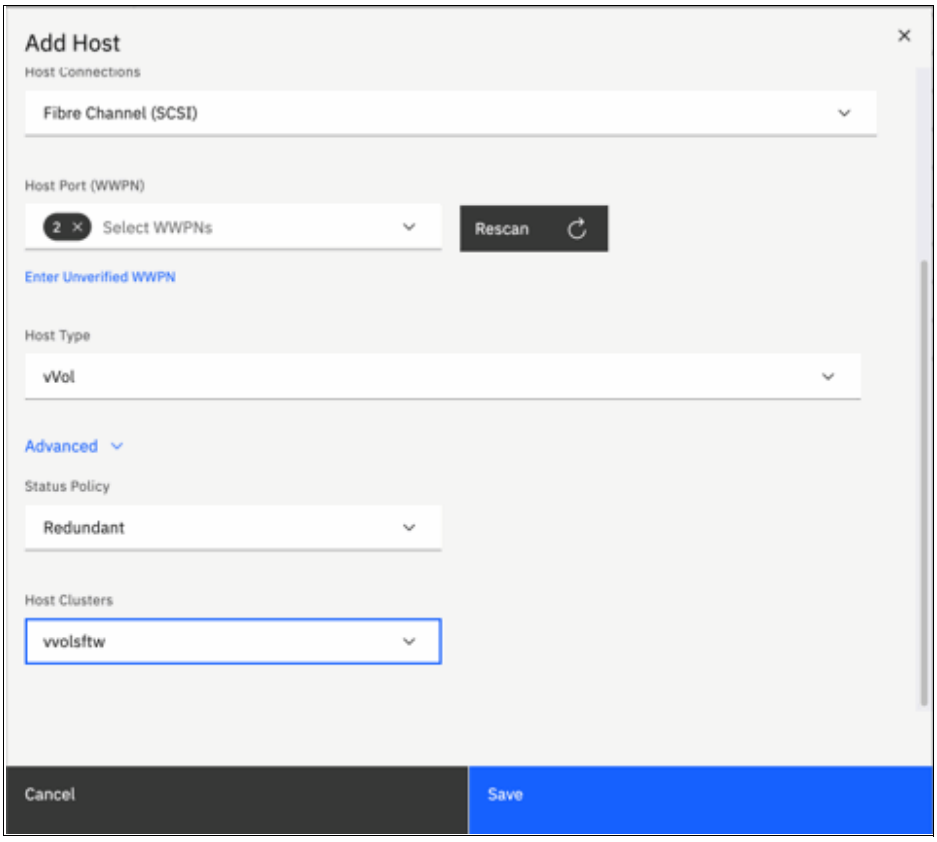


Figure 6-13 Selecting the Host Cluster

Note: When creating the host object by using the CLI, use the host type **`adminlun`**:

```
IBM_2145:vvolsftw-sv1:superuser>mkhost -fcwwpn
2100000E1EC249F8:2100000E1EC249F9 -name vvolsftw-02 -hostcluster vvolsftw -type
adminlun
```

7. Repeat the process for each additional host that you want to create.

8. Verify that all hosts are correctly defined as vVol host types by selecting **Hosts** → **Hosts** in the storage system GUI, as shown in Figure 6-14.

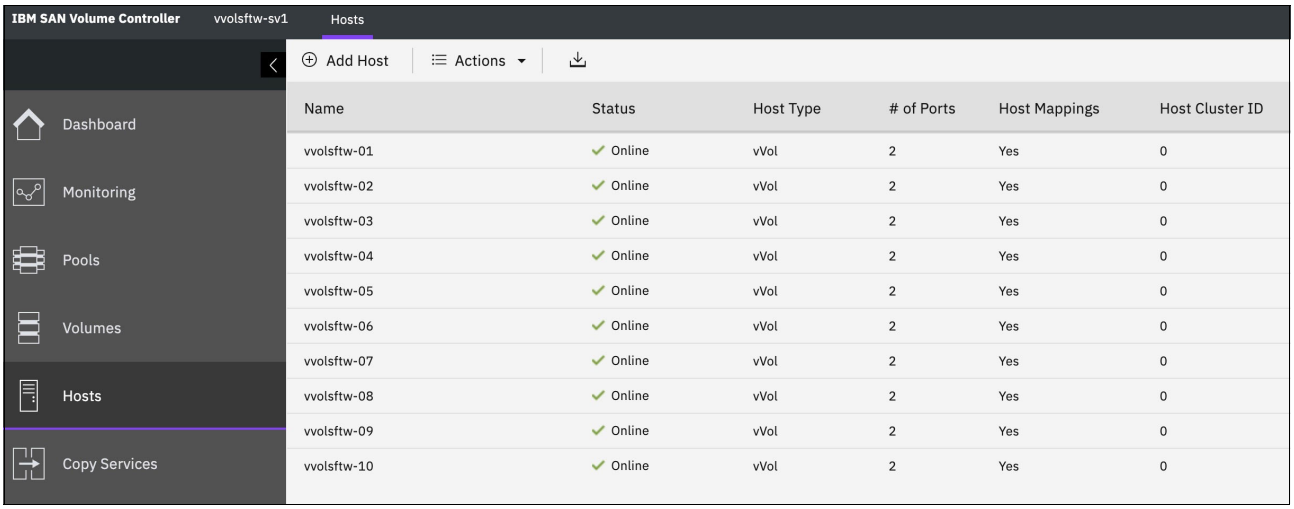


Figure 6-14 Hosts

9. You can ensure consistency across all members of the host cluster by defining the host type at the host cluster level. To do this task, select **Hosts** → **Host Clusters**. Right-click the host cluster and select **Modify Host Types**, as shown in Figure 6-15.

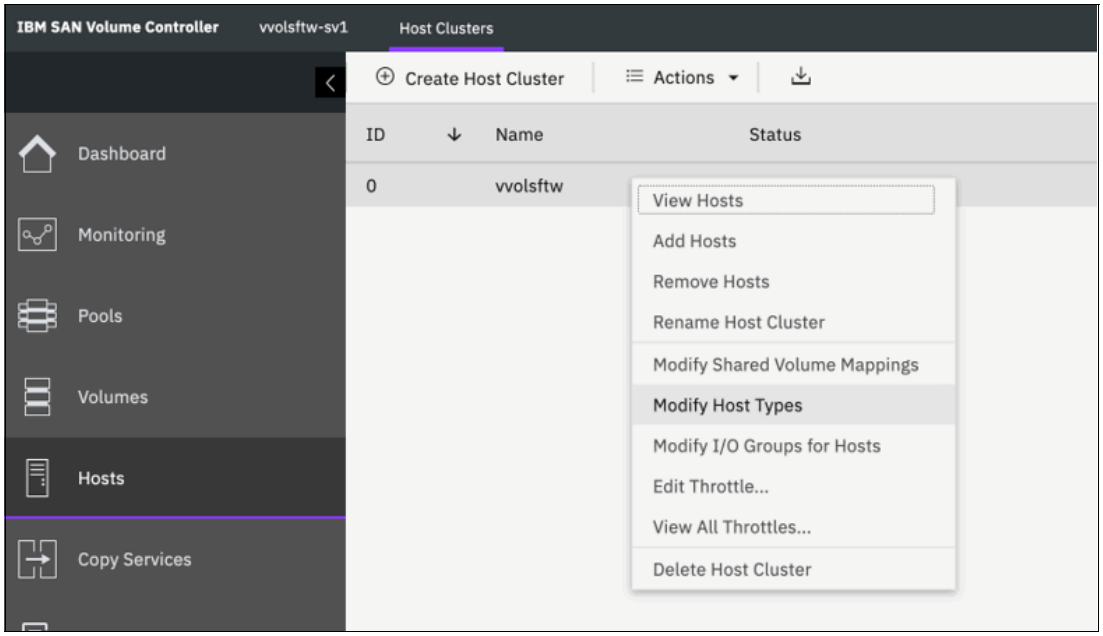


Figure 6-15 Selecting Modify Host Types

10. Select the vVol host type and click **Modify**, as shown in Figure 6-16 on page 143.

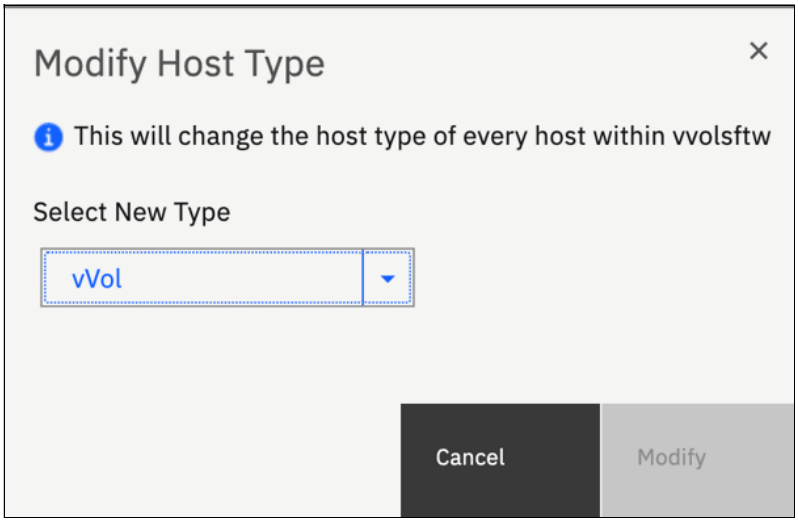


Figure 6-16 Clicking Modify

By configuring the vVol host type on the host or host cluster object, the system automatically presents the Protocol Endpoints to the ESXi hosts.

11. Before continuing with the Embedded VASA Provider configuration, verify that all hosts in the vSphere cluster correctly detected the Protocol Endpoints from the storage system. To do this task, rescan the storage adapters on each ESXi host and verify that there is a Small Computer System Interface (SCSI) device with SCSI ID 768 and 769, as shown in Figure 6-17.

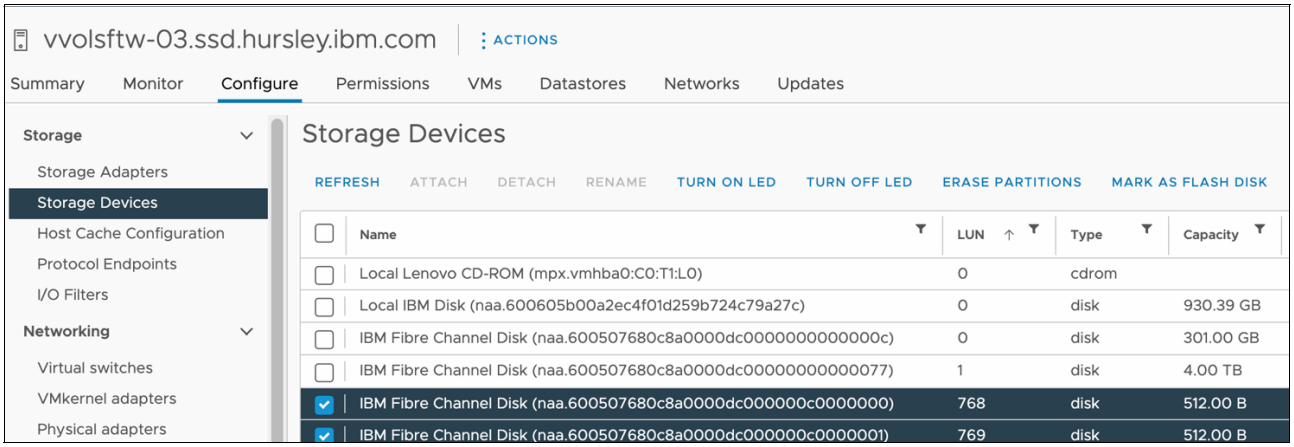


Figure 6-17 Rescanning the storage adapters

Note: A Protocol Endpoint is presented from each node in the IBM Spectrum Virtualize cluster. Ensure that all ESXi hosts correctly identified all Protocol Endpoints (PEs) before continuing.

6.3 Enabling vVols by using Embedded VASA Provider

After the three system prerequisites are met, the Enable vVol toggle becomes available, as shown in Figure 6-18.

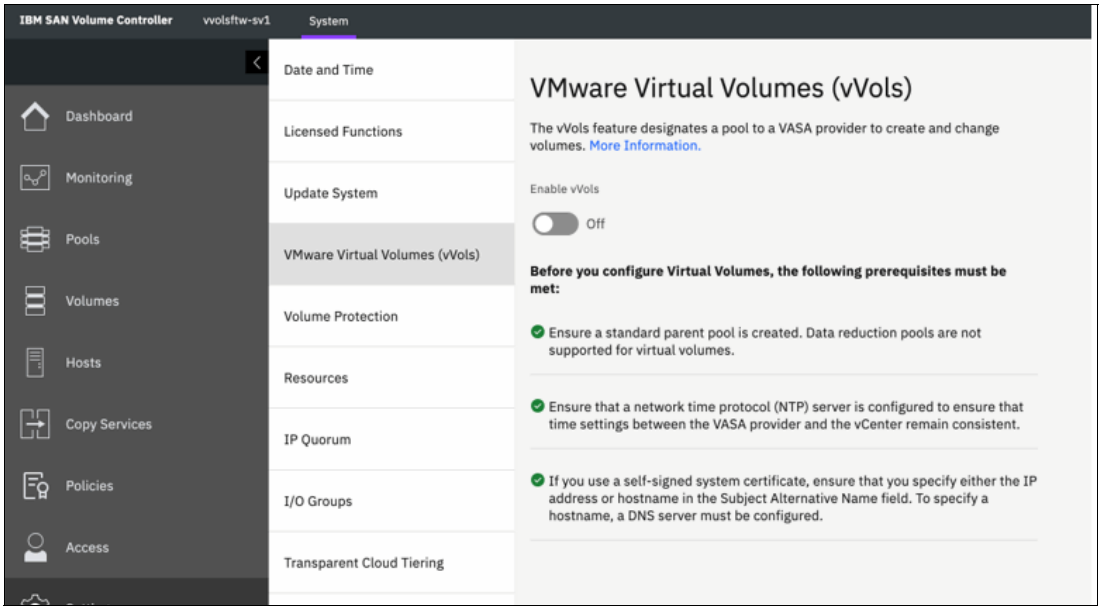


Figure 6-18 Enable vVol toggle becomes available

Click **Enable vVol** to start the process.

After all the values are defined, the GUI creates all the necessary objects within IBM Spectrum Virtualize to facilitate vVol support, as shown in Figure 6-19 on page 145.

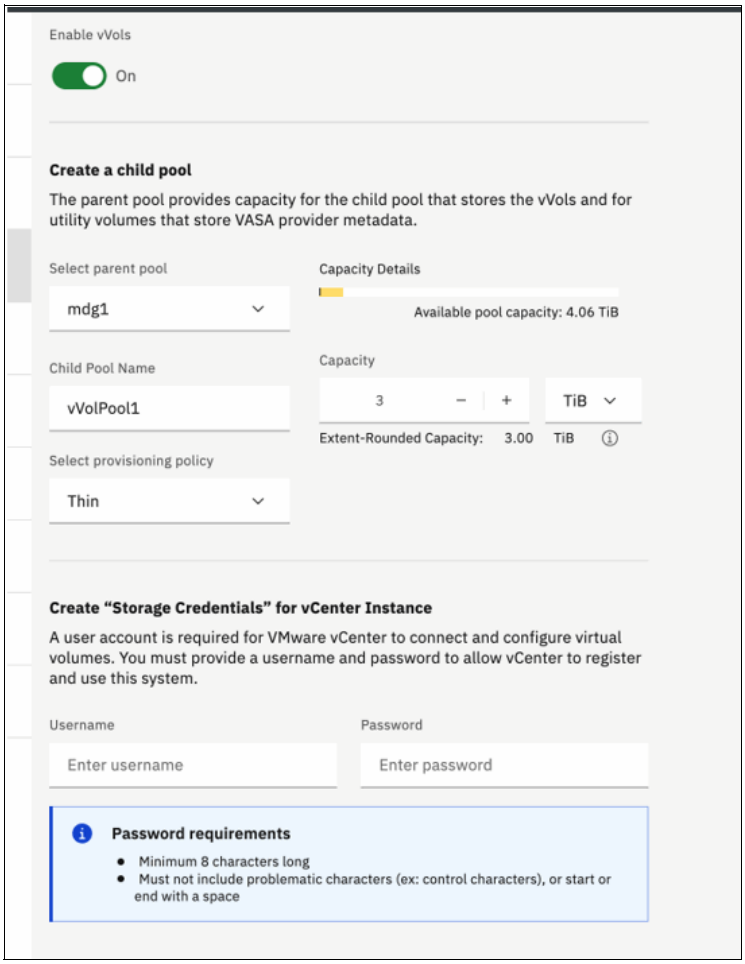


Figure 6-19 Required objects for vVol support are created

6.3.1 Parent pool

You are required to select a parent pool in which to store both the metadata VDisk and the associated (first) child pool. The child pool is presented to vSphere as a vVol-enabled storage container.

A *metadata volume* (utility volume) is created as a thin-provisioned volume with a capacity of 2 TB. This metadata volume is mounted internally on the configuration node and used to store associated metadata for the vVol configuration. This metadata can include metadata for storage containers, virtual volumes, and other objects that are required to facilitate the vVol infrastructure.

Note: Even though it has been allocated a maximum capacity of 2 TB, this volume is only intended to store system metadata, so it is likely that the used capacity will never grow beyond a few gigabytes in size.

6.3.2 Child pool (vVol-enabled Storage Container)

You are prompted to enter a name and capacity for a new child pool. This pool is presented to the vSphere environment as a vVol-enabled storage container, so the specified capacity of the pool dictates the size of the vVol data store within vSphere. The capacity can be increased or decreased later, so there is flexibility for expansion and scale as the infrastructure matures.

Although the initial vVol configuration allows only for the creation of a single child pool, more vVol-enabled child pools can be created later through the storage system CLI if increased capacity or different tiers of storage are required.

6.3.3 Provisioning policy

This provisioning policy dictates how vVols are created within the IBM Spectrum Virtualize storage system. Each vVol child pool is associated with a specific provisioning policy, which means that where possible all vVols that are created in a vVol data store are provisioned in the same way.

Note: Swap vVols always are created as fully allocated volumes within IBM Spectrum Virtualize regardless of the specified provisioning policy.

There are two available provisioning policies to select:

- Standard: The Standard provisioning policy uses fully allocated volumes. All vVols that are created in pools with this policy are created as fully allocated volumes within IBM Spectrum Virtualize.
- Thin-provisioning: The Thin-provisioning policy uses space-efficient, thin-provisioned volumes. All vVols are created as space-efficient, thin-provisioned volumes within IBM Spectrum Virtualize.

6.3.4 Storage credentials

To register the VASA Provider within vCenter, you must enter the following information:

- Name
- URL
- Username
- Password

The storage credentials that are defined in this window are required when registering the Storage Provider within vSphere, and they are initially used to authenticate the vSphere environment against the IBM Spectrum Virtualize storage system.

The system automatically creates a user group that is assigned with a specific role within IBM Spectrum Virtualize. Then, the user account that is specified here is created as a member of this group and granted specific access rights to allow manipulation of vVol objects within the storage system.

However, upon successful registration of the storage provider within vCenter, the password is removed from the user account within IBM Spectrum Virtualize, and instead the vSphere certificate is used to authenticate the user account.

Note: The password that is defined in the window is used once, and it is required only in the initial Storage Provider registration process.

After a successful registration, if it is necessary to reregister the storage provider in vCenter for whatever reason, the defined user account must be reconfigured with a new password and the certificate authentication must be removed. To do this task, connect to the storage system CLI and run the following commands:

1. To list all users that are configured on the system, run the following command:

   ```
   lsuser
   ```

2. Identify the user account that requires reconfiguration, and run the following commands:

   ```
   chuser -nocert <user_id or name>
   chuser -password <new password> <user_id or name>
   ```

Now, the Storage Provider can be reregistered in vCenter by using the new password.

6.3.5 Registering the Storage Provider in vSphere

After you finish your work in the vVol configuration window, you see an option that is called “Copy the following URL:”, as shown in Figure 6-20. This string is the URL that is required when registering the storage provider within vCenter, and it should have the following format:

```
https:// <FQDN/IP address> + :8440 + /services/vasa
```

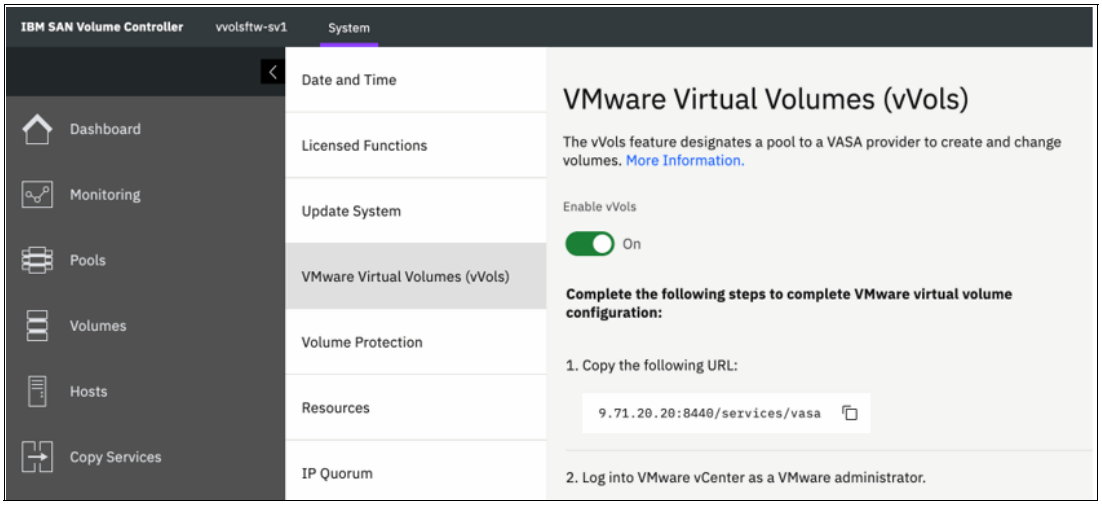


Figure 6-20 Copy the following URL: option

To register the Storage Provider in vSphere, complete the following steps:

1. Click the copy icon for **Copy the following URL** to copy the string to your clipboard.
2. Open the vCenter web interface and find the vCenter server in the inventory tree. Select **Configure** → **Storage Providers**, and then click **Add**, as shown in Figure 6-21.

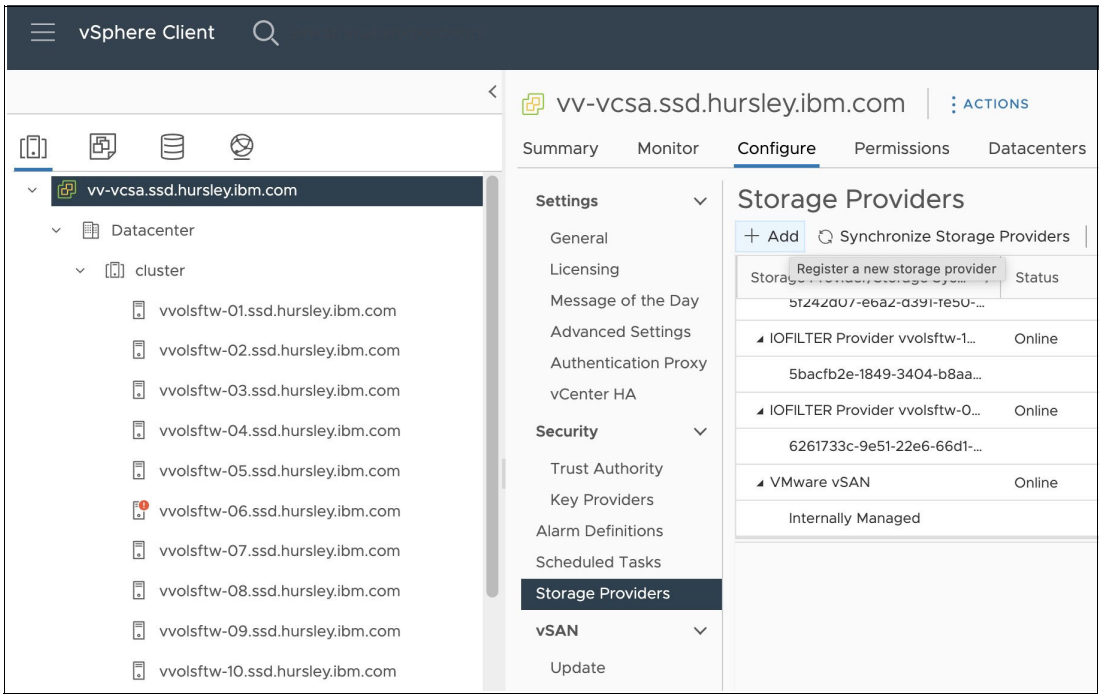


Figure 6-21 Selecting Storage Providers

3. Enter an identifiable name, and paste the URL into the URL field. Add the user credentials that were defined earlier and click **OK**, as shown in Figure 6-22.

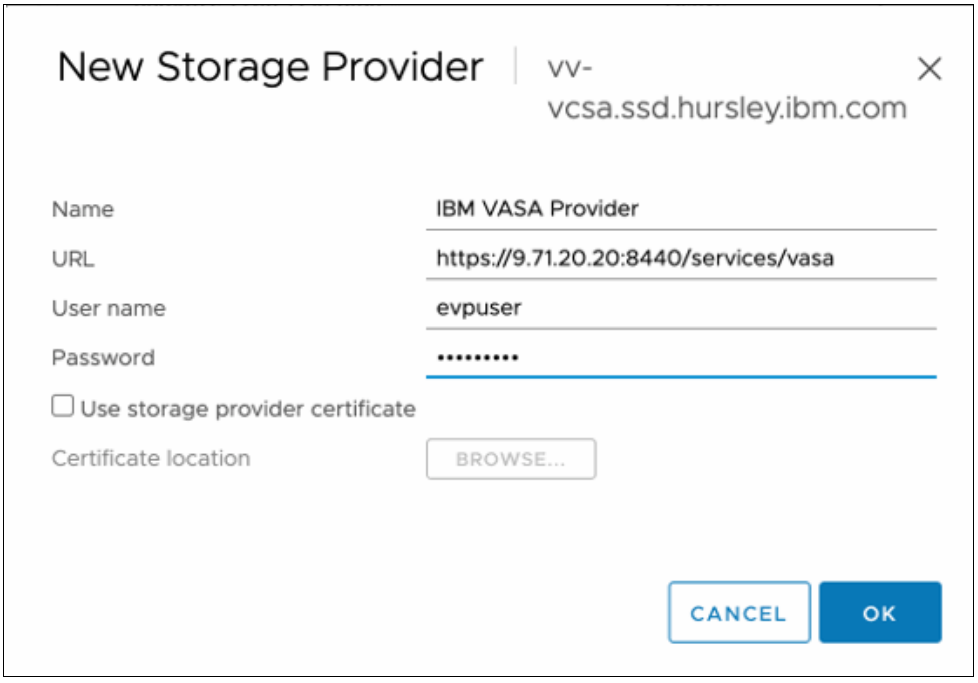


Figure 6-22 New Storage Provider

4. You might see an error or warning saying that the operation failed, but this message is related only to the initial certificate thumbprint warning, so it can be ignored (Figure 6-23).

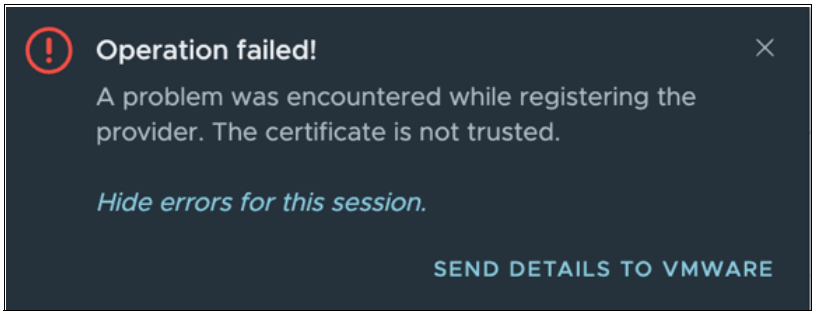


Figure 6-23 Operation failed message

5. Verify that the newly added Storage Provider is showing online and active in the Storage Providers list (Figure 6-24).

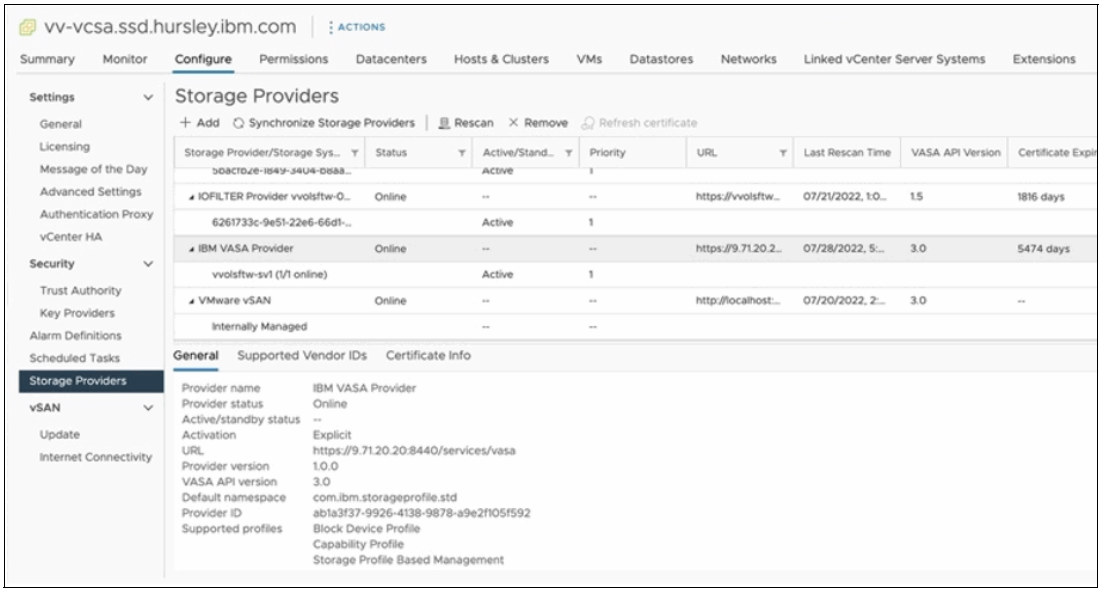


Figure 6-24 Newly added Storage Provider is showing online and active

6.3.6 Creating the vVol data store

Review the vCenter inventory and identify the cluster or host that you want to mount on the vVol data store by completing the following steps:

1. Right-click the cluster and select **Storage** → **New Datastore**, as shown in Figure 6-25.

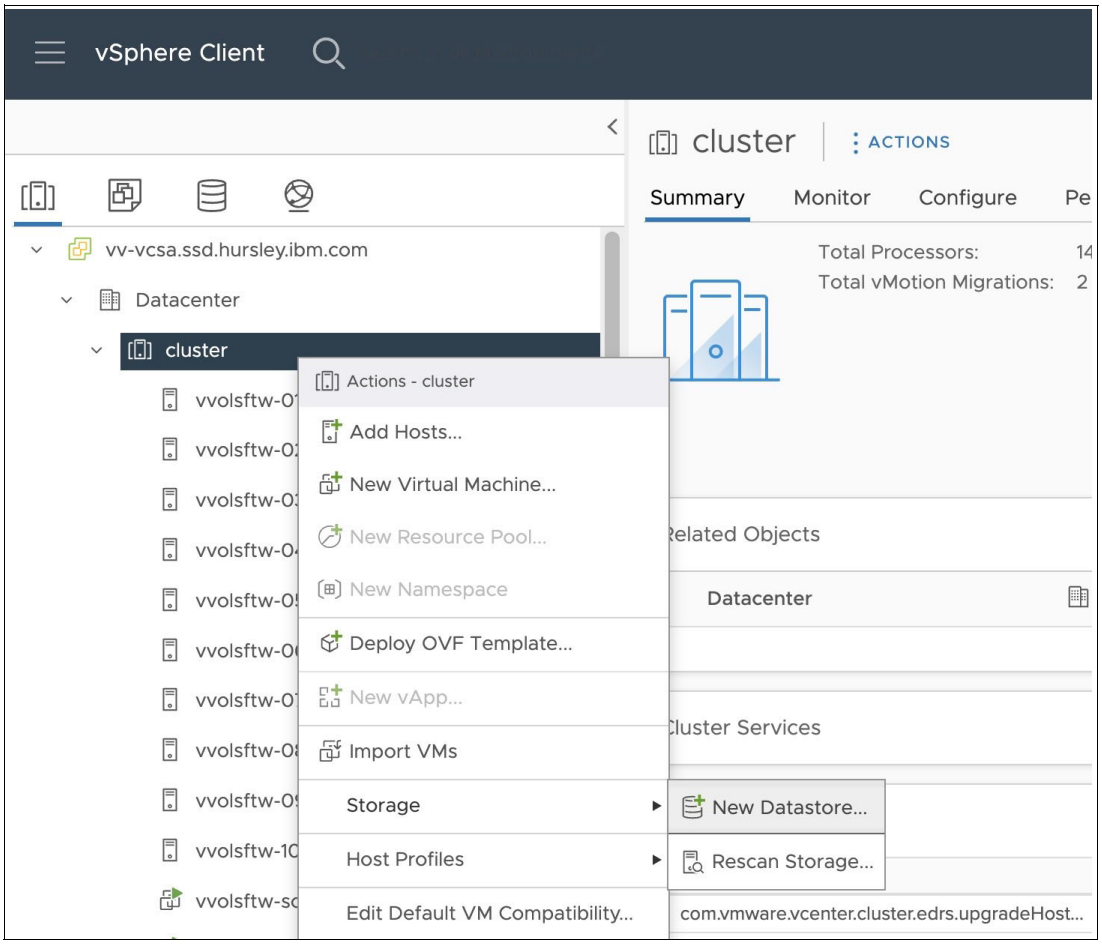


Figure 6-25 Selecting New Datastore

2. Select **vVol** and select **NEXT** (Figure 6-26).

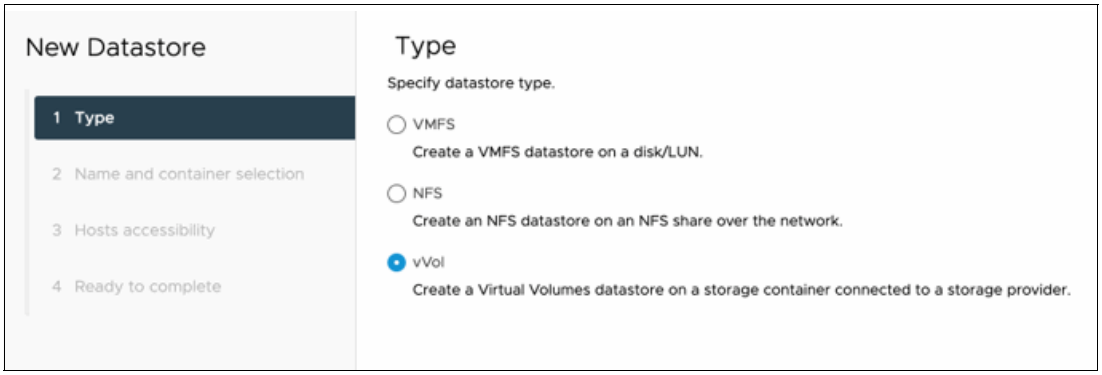


Figure 6-26 Selecting vVol

3. Select the backing storage container in the list and define the name of the new vVol data store. Click **NEXT** (Figure 6-27).

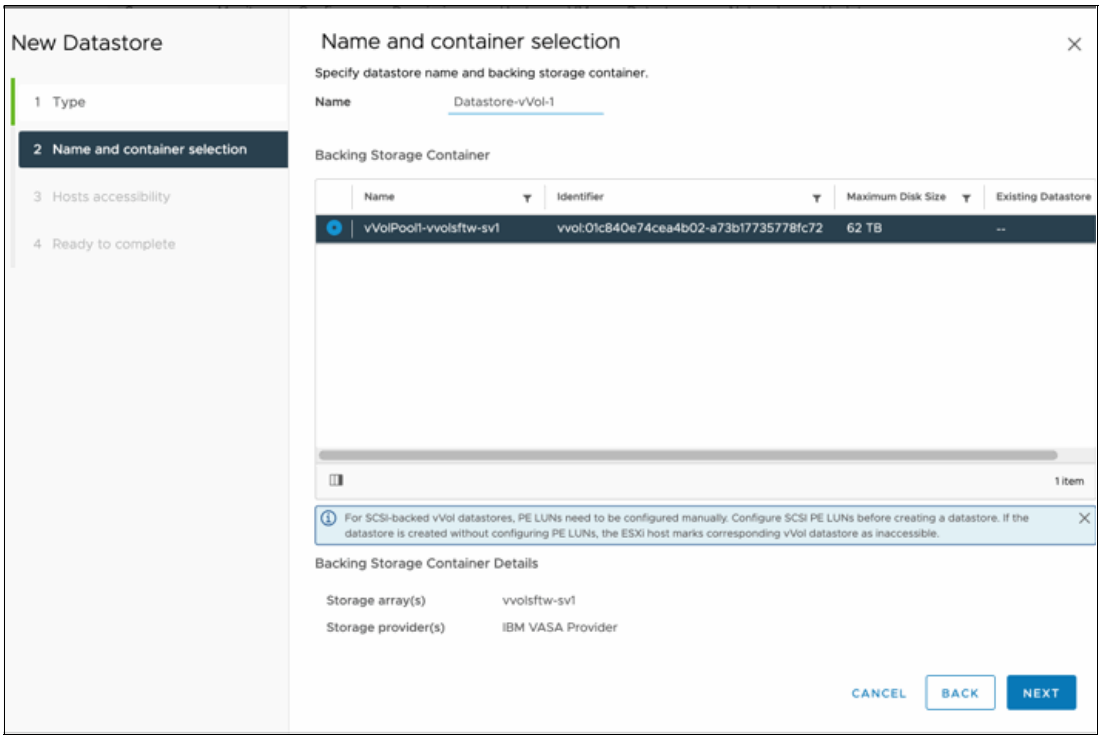


Figure 6-27 Defining the name of the new vVol data store

4. Select the hosts that will access the vVol data store and click **NEXT** (Figure 6-28).

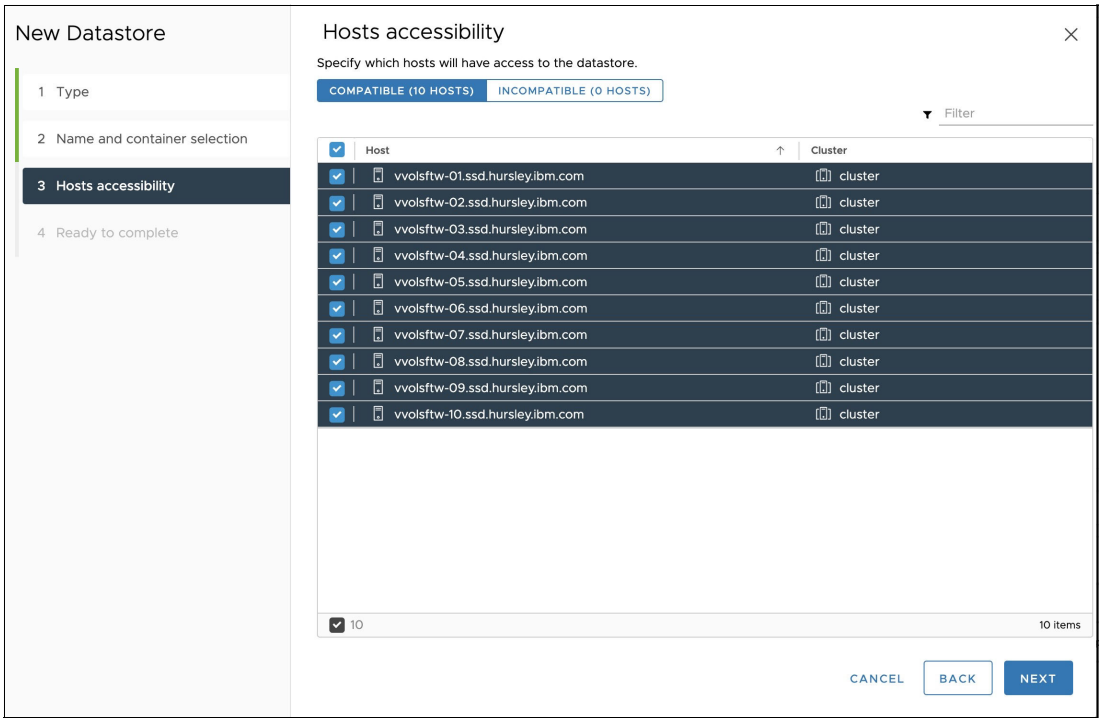


Figure 6-28 Selecting the hosts

5. Review the summary window and click **FINISH**, as shown in Figure 6-29.

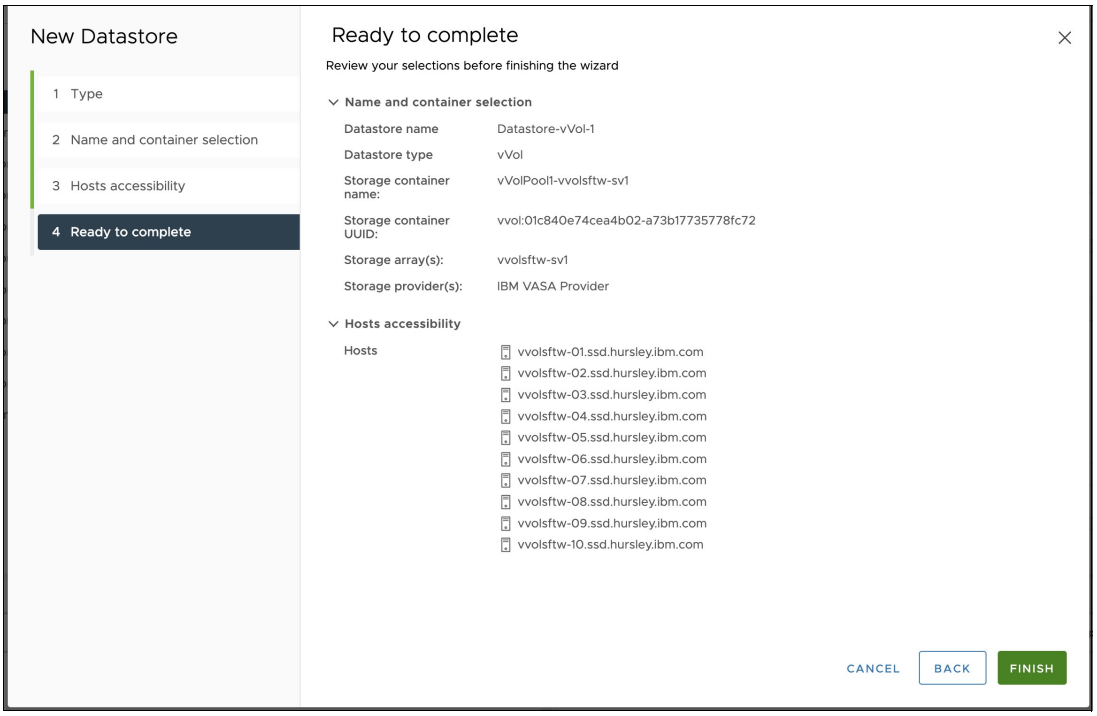


Figure 6-29 Summary window

6. Review the **Datastores** tab and ensure that the capacity and accessibility are correctly reported, as shown in Figure 6-30.

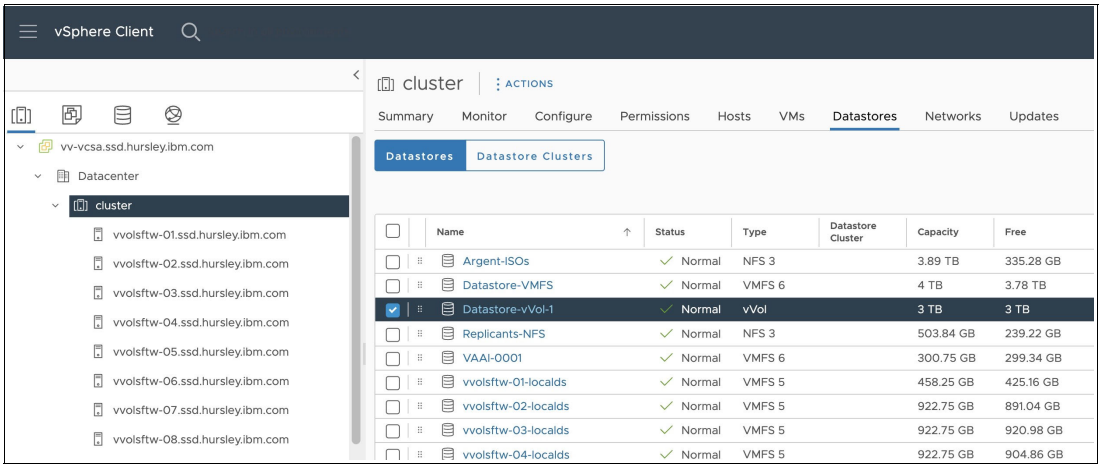


Figure 6-30 Reviewing the Datastores tab

6.3.7 Provisioning more vVol data stores

In this section, we cover the steps for provisioning more vVol data stores.

Identifying the parent pool

To create more child pools, use Secure Shell (SSH) to connect to the CLI of the IBM Spectrum Virtualize management interface. Familiarize yourself with the available parent pools by running the following command:

```
lsmdiskgrp -filtervalue type=parent
```

Identify the target parent pool in which to create the child pool and note the **mdiskgrp** ID or name.

Identifying the ownership group

Identify the ownership group that is assigned to the VASA provider by running the command in Example 6-1.

Example 6-1 The lsownershipgroup command

```
IBM_2145:vvolsftw-sv1:superuser>lsownershipgroup
id name
0  VASA
```

By default, the name that is associated with the VASA ownership group is `VASA`.

Identifying the provisioning policy

When vVol was enabled in IBM Spectrum Virtualize, the requested provisioning policy was created along with the other required objects. If both provisioning policies are not listed, you might need to create them manually.

Identify the provisioning policy that is required for the new vVol child pool by running the **lsprovisioningpolicy** command, as shown in Example 6-2.

Example 6-2 The lsprovisioningpolicy command

```
IBM_2145:vvolsftw-sv1:superuser>lsprovisioningpolicy
id name     capacity_saving deduplicated in_use
0  Thin     thin            no           yes
1  Standard none            no           no
```

Creating the vVol Storage Container

To create a vVol-enabled child pool by using the values that are identified in the earlier sections, run the command that is shown in Example 6-3.

Example 6-3 The mkmdiskgrp command

```
svctask mkmdiskgrp -name <name> -owner vvol_child_pool -ownershipgroup <VASA
ownershipgrp_id or name> -parentmdiskgrp <mdiskgrp id or name> -provisioningpolicy
<Thin or Standard> -size <capacity> -unit tb
```

6.4 Migrating from existing IBM Spectrum Connect vVol configurations

In this section, we cover the process of migrating from existing IBM Spectrum Connect vVol configurations.

6.4.1 Supported migration path

In the initial release of the Embedded VASA Provider, it is not possible to use both external and embedded VASA providers from the same vCenter to the same storage system. Therefore, there is no direct method of migrating between vVol data stores that are presented through IBM Spectrum Connect and vVol data stores that are provided by the Embedded VASA Provider.

To perform a migration between vVol solutions, you must complete the following steps:

1. Provision Virtual Machine File System (VMFS) data stores with sufficient capacity to store all VMs that are on the vVol storage.
2. Use Storage vMotion to migrate existing VMs or templates from vVol storage to VMFS data stores.
3. Remove the vVol IBM Spectrum Connect configuration from vCenter.
4. Disable or decommission the vVol function on the storage system.
5. Enable vVol by using the Embedded VASA Provider and create vVol data stores.
6. Use Storage vMotion to migrate VMs and templates from VMFS storage to the new vVol data stores.
7. Remove the temporary VMFS data stores if they are no longer needed.

6.4.2 VM migrations by using Storage vMotion

Temporary storage is required to store the VMs during the decommissioning of IBM Spectrum Connect. If required, create volumes of suitable capacity to store all the VMs or templates while the IBM Spectrum Connect vVol storage is being decommissioned.

After the new volumes are created and mapped to the appropriate hosts or host cluster in the storage system, create the VMFS data stores.

Identify the VMs that are on the vVol data stores that are presented by IBM Spectrum Connect and perform a storage migration to move them to the new VMFS data stores. Depending on the number of VMs, consider using, for example, PowerCLI to automate bulk VM migrations.

Note: VM templates might exist on the vVol data stores that require conversion into a VM before they can be migrated. After the migration completes, they can be converted back into a VM template.

6.4.3 Removing the vVol IBM Spectrum Connect configuration from vCenter

In this section, we describe removing the vVol IBM Spectrum Connect configuration from vCenter.

Removing or unmounting the vVol data stores

After all VMs and templates are migrated away from the IBM Spectrum Connect vVol data stores, it is safe to unmount and remove them from vCenter. To do this task, complete the following steps:

1. Select the vVol data store to be removed and click **ACTIONS**, as shown in Figure 6-31.

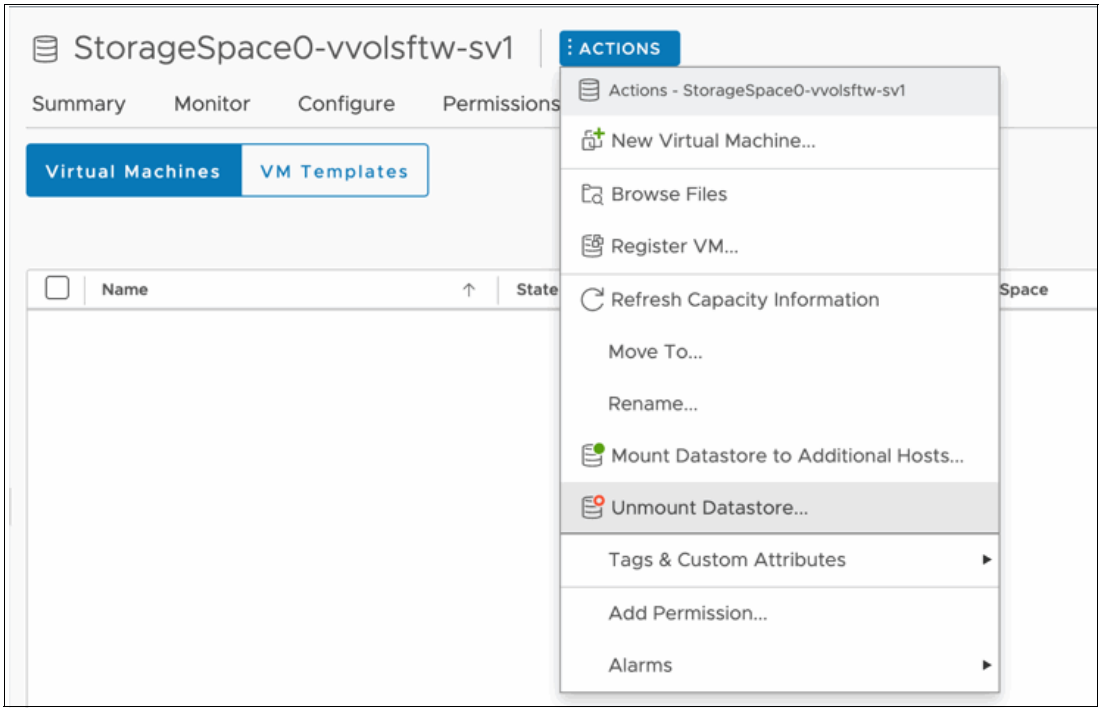


Figure 6-31 Selecting Unmount Datastore

2. Select **Unmount Datastore** and select all connected hosts to unmount the data store from all hosts, as shown in Figure 6-32. Click **OK**.

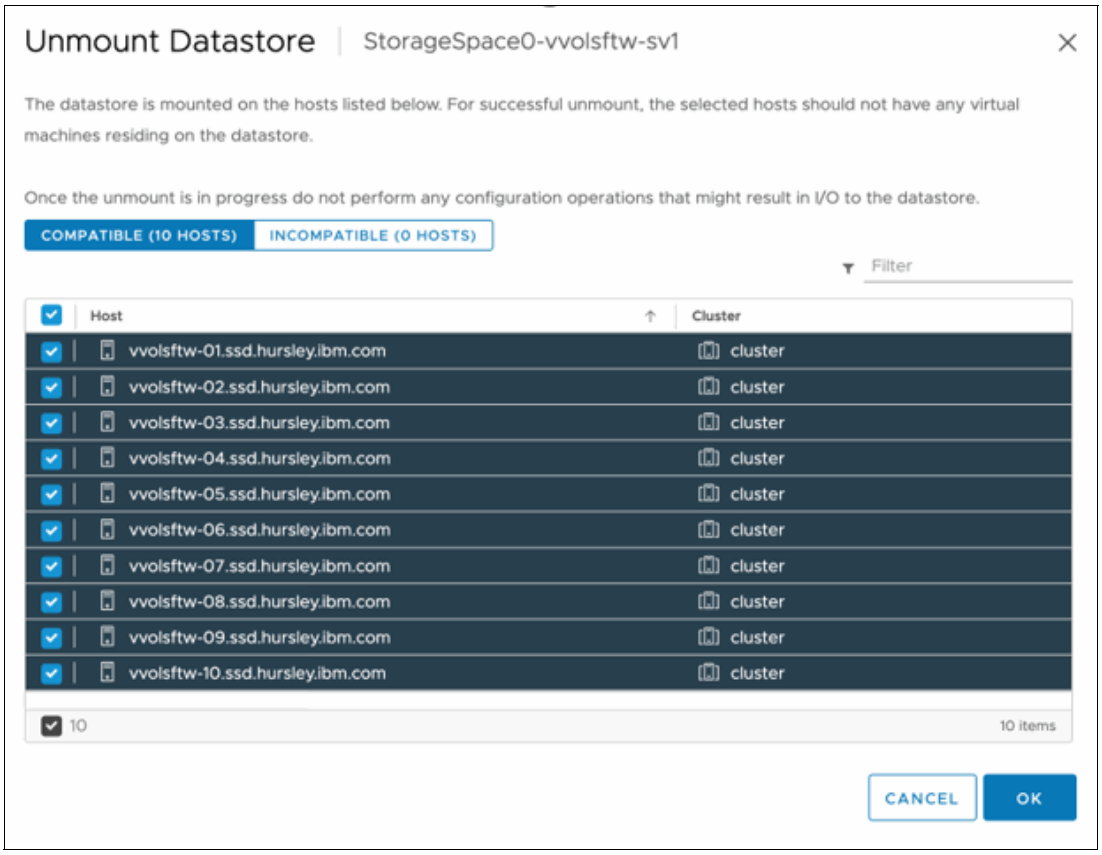


Figure 6-32 Unmount Datastore option

3. After the data store is unmounted from all hosts, it automatically is removed from vCenter.
4. Repeat these steps for all the vVol data stores that are presented by IBM Spectrum Connect.

Removing Storage Policies

Check for any configured Storage Policies that correspond to IBM Spectrum Connect. To do this task, complete the following steps:

1. Go to the Policies and Profiles view within vCenter, as shown in Figure 6-33 on page 157.

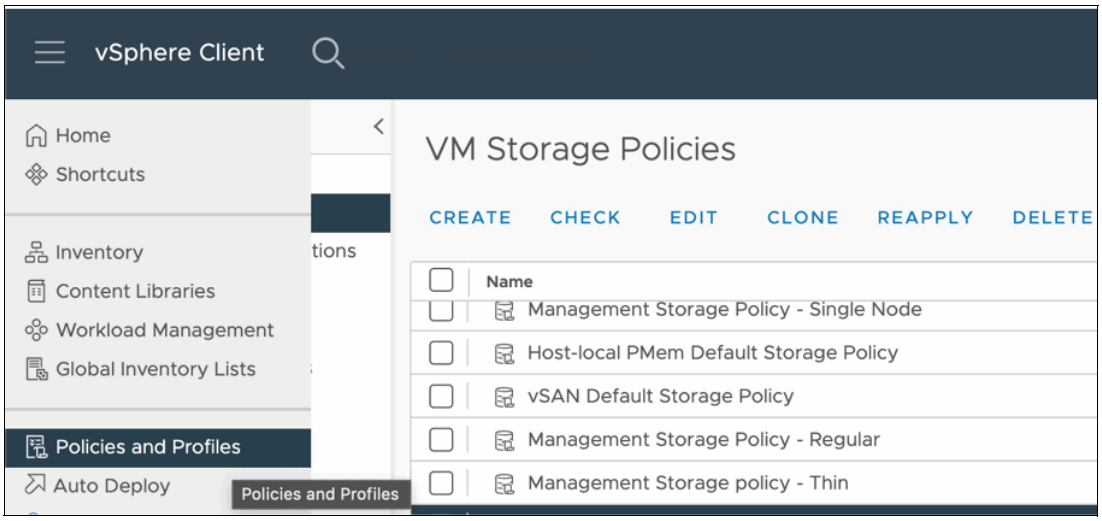


Figure 6-33 Policies and Profiles view

2. Select **VM Storage Policies** and identify any policies that were created by using IBM Spectrum Connect.
3. Select the policy to remove, and click **DELETE**, as shown in Figure 6-34.

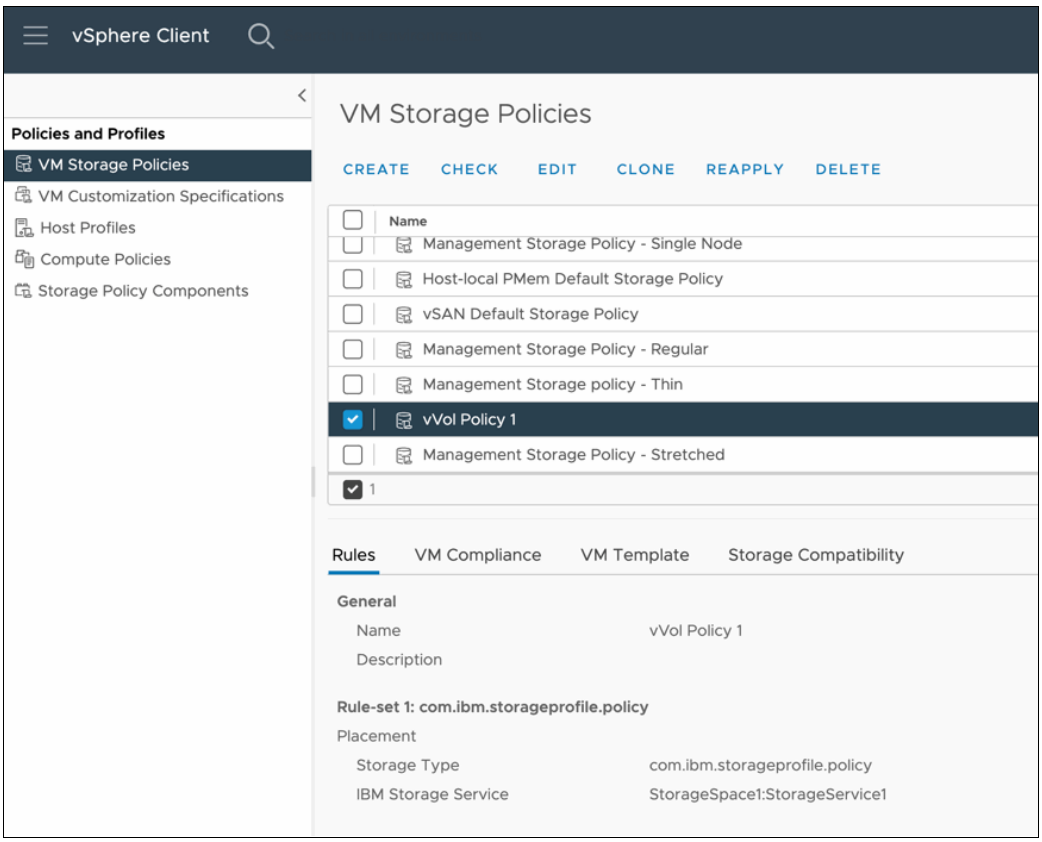


Figure 6-34 Selecting the policy to remove

4. Repeat these steps for any remaining policies that are associated with IBM Spectrum Connect.

Removing IBM Spectrum Connect Storage Provider

After all the vVol data stores are unmounted and removed, it is safe to remove the Storage Provider from within vCenter. To do this task, complete the following steps:

1. Find the vCenter entry in the inventory tree and click the **Configure** tab.
2. Select **Storage Providers**.
3. Identify the IBM Spectrum Connect Storage Provider in the list and select **REMOVE** (Figure 6-35).

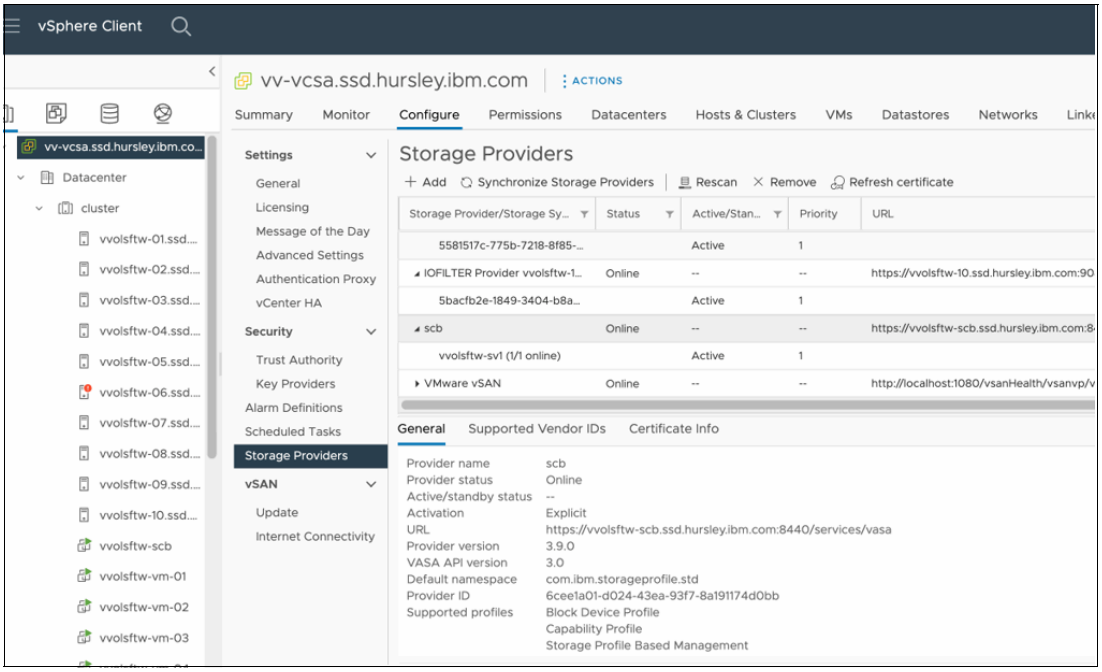


Figure 6-35 Removing the IBM Spectrum Connect Storage Provider

6.5 Decommissioning IBM Spectrum Connect

IBM Spectrum Connect offers multiple integrations into different VMware products. Before continuing, review the integration interfaces that are configured, and be conscious of how they are being used in your environment.

6.5.1 Identifying and removing the vVol child pools for IBM Spectrum Connect

Identify the child pools that were allocated to any vVol Storage Spaces within the IBM Spectrum Connect GUI and delete them, as shown in Figure 6-36 on page 159.

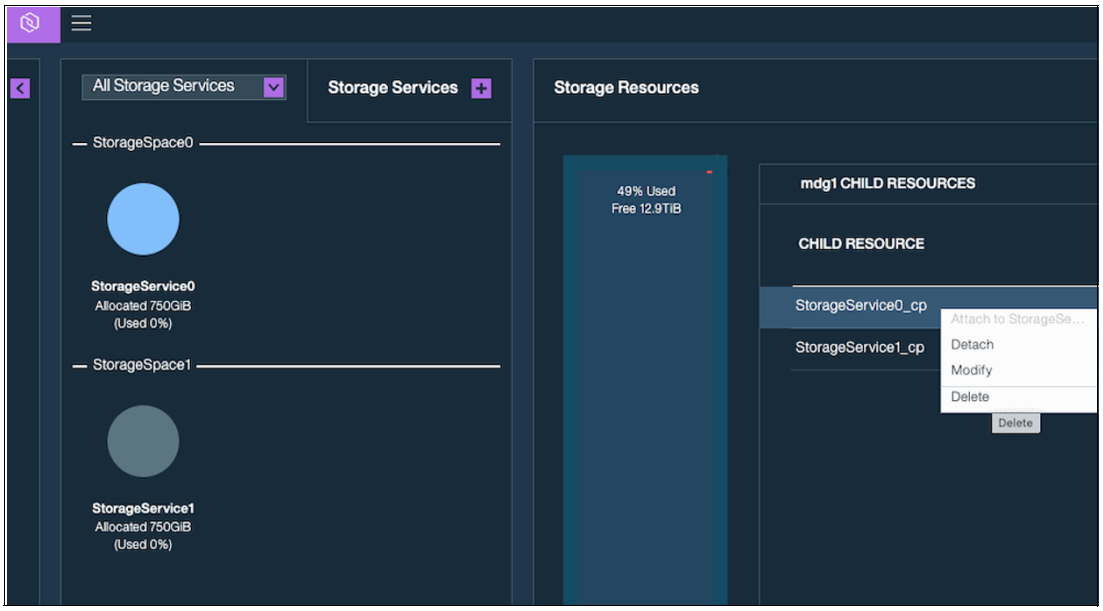


Figure 6-36 Deleting the child pools that were allocated to any vVol Storage Spaces

Alternatively, log on to the storage system CLI with a user account that has the VASA Provider role and run the following command to identify any existing vVol child pools that are used by IBM Spectrum Connect:

```
lsmdiskgrp -filtervalue owner_type=vvol_child_pool
```

Note the `mdiskgrp` name or ID, and then verify that the vVol pool is no longer required and that the name and ID are correct because there is no way to recover the pool after it is deleted. Once you are sure, run the following command to remove the child pool:

```
rmmdiskgrp <name or id>
```

Warning: Removing the pool might fail if any vVols are in the pool. You might need to manually remove any vVols in the pool before removing the pool itself. To identify any vVols that are in the pool to be deleted, run the following command:

```
lsvdisk -filtervalue mdisk_grp_name=<child pool name>
```

For each vVol, identify the VDisk ID or name and run the following command to delete the vVol.

Warning: Verify that the vVol is no longer required and that the name and ID are correct because there is no way to recover the data after the volume is deleted.

```
rmvdisk -force <vVol name or id>
```

After any lingering vVols are deleted, retry the pool removal command until all IBM Spectrum Connect vVol pools are removed.

6.5.2 Removing the user account that is used by IBM Spectrum Connect

> **Warning:** If other integration interfaces are configured, for example, vCenter or vRealize Orchestration, do not remove the user account because its removal will cause future integration commands to fail.

Identify the user account that is used by IBM Spectrum Connect by either reviewing the Storage System Credentials window in the IBM Spectrum Connect GUI or by using the CLI.

Using the GUI, select **Menu** → **Storage credentials**, as shown in Figure 6-37.

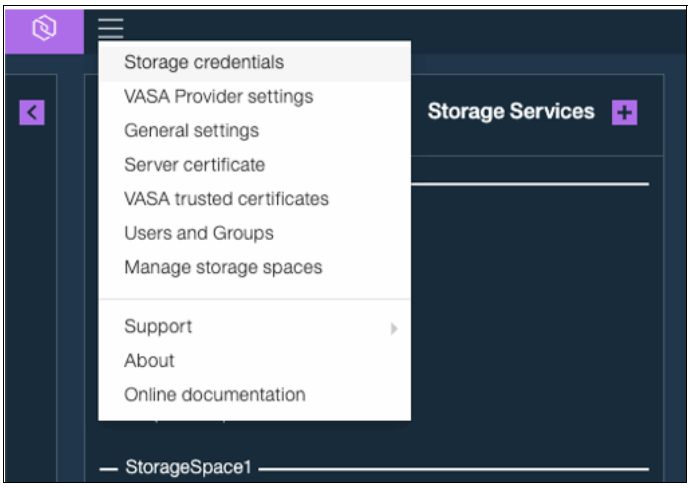


Figure 6-37 Storage credentials

You see the user account that is used by IBM Spectrum Connect, as shown in Figure 6-38.

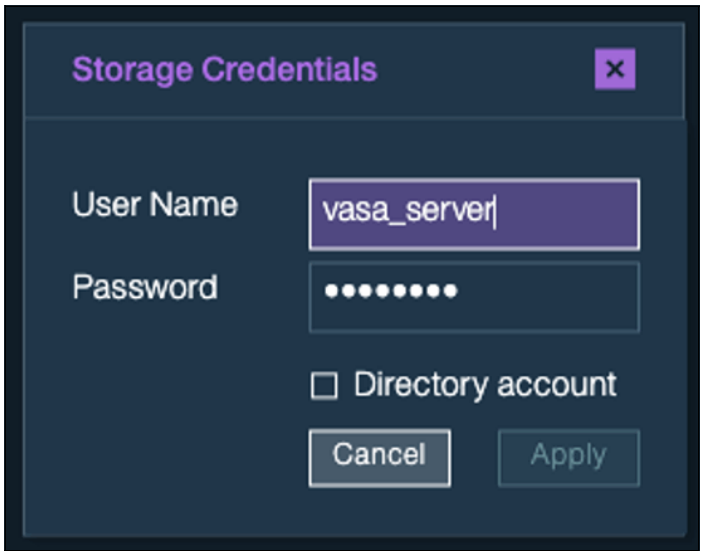


Figure 6-38 User account that is used by IBM Spectrum Connect

You also can use the command in Example 6-4 on page 161 on the storage system CLI.

Example 6-4 Identifying the user account that is used by IBM Spectrum Connect by using the command-line interface

```
IBM_2145:vvolsftw-sv1:superuser>lsuser
id name        password ssh_key remote usergrp_id usergrp_name  owner_id owner_name locked password_change_required
0  superuser   yes      yes     no     0          SecurityAdmin                    no     no
1  vasa_server yes      no      no     6          VASAUsers                        no     no
```

Remove the user account that is used by IBM Spectrum Connect, and then run the following command:

```
rmuser <user_id or name>
```

To identify the User Group that is associated with the VASA Provider role, run the command in Example 6-5 on the storage system CLI.

Example 6-5 The lsusergrp command

```
IBM_2145:vvolsftw-sv1:superuser>lsusergrp
id name            role            remote multi_factor password_sshkey_required gui_disabled cli_disabled rest_disabled
0  SecurityAdmin   SecurityAdmin   no     no           no                       no           no           no
1  Administrator   Administrator   no     no           no                       no           no           no
2  CopyOperator    CopyOperator    no     no           no                       no           no           no
3  Service         Service         no     no           no                       no           no           no
4  Monitor         Monitor         no     no           no                       no           no           no
5  RestrictedAdmin RestrictedAdmin no     no           no                       no           no           no
6  VASAUsers       VasaProvider    no     no           no                       no           no           no
```

Assuming no other user accounts are in the user group, remove the VASA Provider user group by running the following command:

```
rmusergrp <usergrp_id or name>
```

To identify the location of the metadata VDisk, run the command in Example 6-6 on the storage system CLI.

Example 6-6 The lsmetadatavdisk command

```
IBM_2145:vvolsftw-sv1:superuser>lsmetadatavdisk
vdisk_id 13
vdisk_name vdisk0
status online
```

Remove the metadata VDisk by running the following command on the storage system CLI.

Warning: The metadata VDisk contains all metadata that is associated with the vVol environment. This operation cannot be undone.

```
rmmetadatavdisk
```

6.5.3 Migrating virtual machines to the vVol data store

Now that the IBM Spectrum Connect vVol configuration is removed, complete the steps in 6.3, “Enabling vVols by using Embedded VASA Provider” on page 144 to enable and configure vVol functions through the Embedded VASA Provider.

After the new vVol data store is online, do the migration by completing the following steps:

1. Identify the VMs that will be migrated, select them, right-click them, and click **Migrate**, as shown in Figure 6-39.

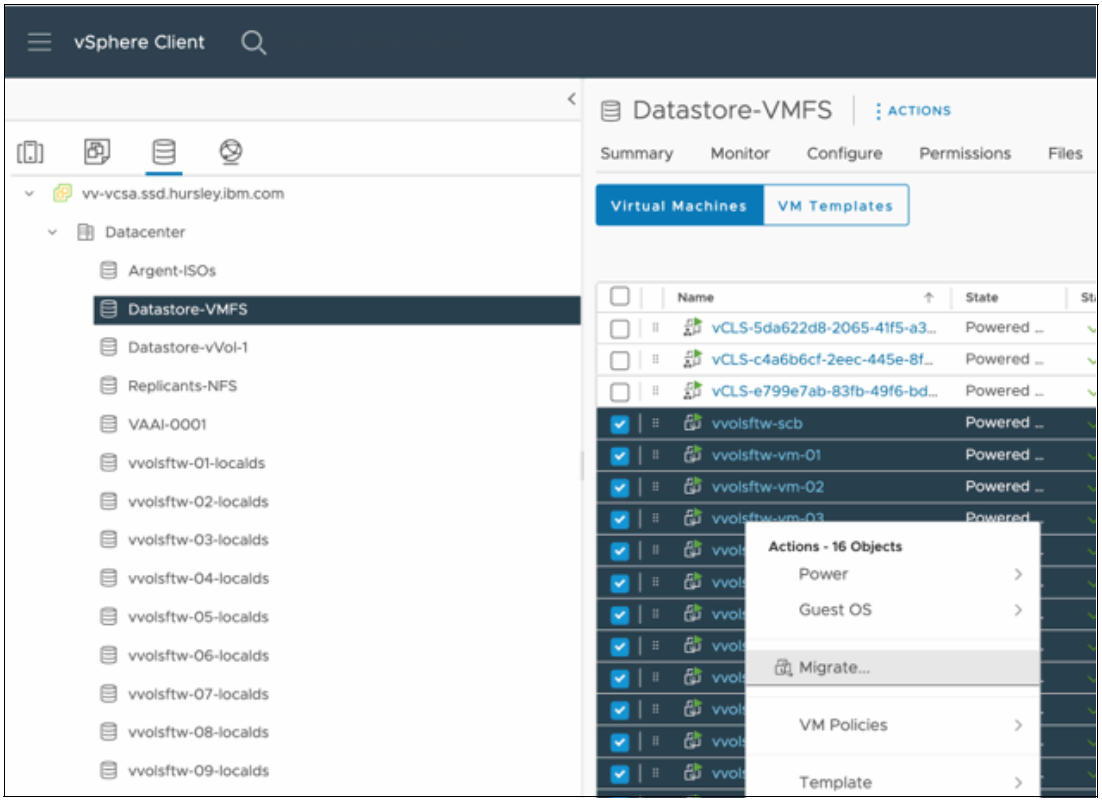


Figure 6-39 Selecting Migrate

2. In the Select Storage window, identify the newly created vVol data store and click **NEXT**, as shown in Figure 6-40 on page 163.

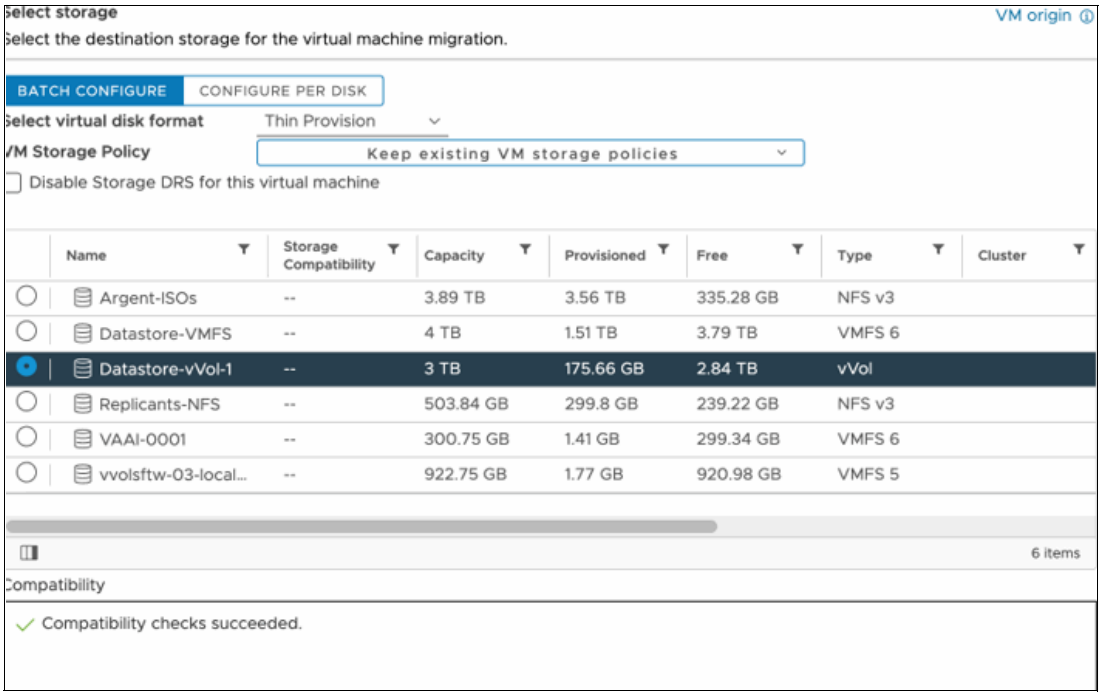


Figure 6-40 Identifying the newly created vVol data store

3. Complete the Storage vMotion workflow and review the tasks to ensure that the VMs successfully migrated (Figure 6-41).

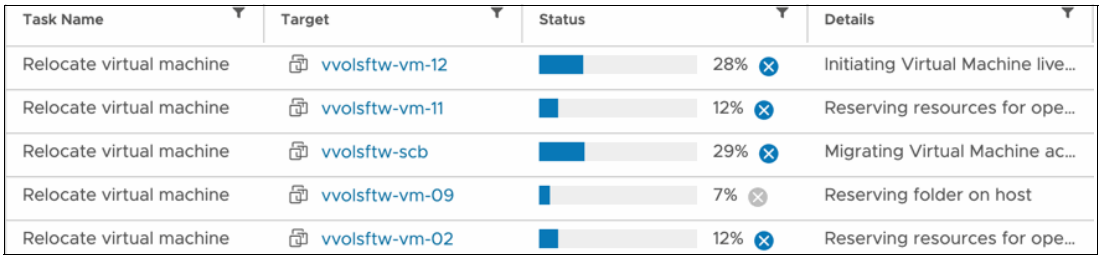


Figure 6-41 Reviewing the tasks to ensure that the VMs successfully migrated

During the migration operation, vVols automatically are created on the storage system within the child pool that was configured as a vVol storage container.

4. To review the vVol objects within IBM Spectrum Virtualize, select **Pools** → **Volumes by Pool** within the GUI, as shown in Figure 6-42.

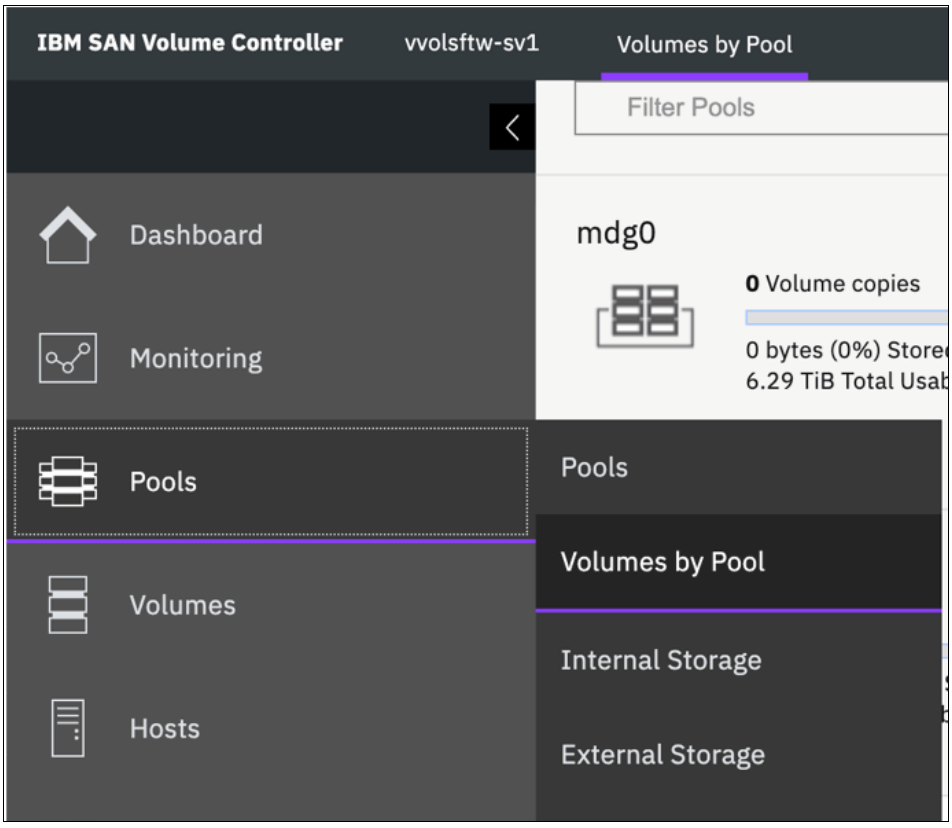


Figure 6-42 Volumes by Pool view

5. Select the vVol-enabled child pool in the list by identifying the vVol logo underneath the pool information on the left side. When you select the vVol pool, the individual vVol objects appear, as shown in Figure 6-43.

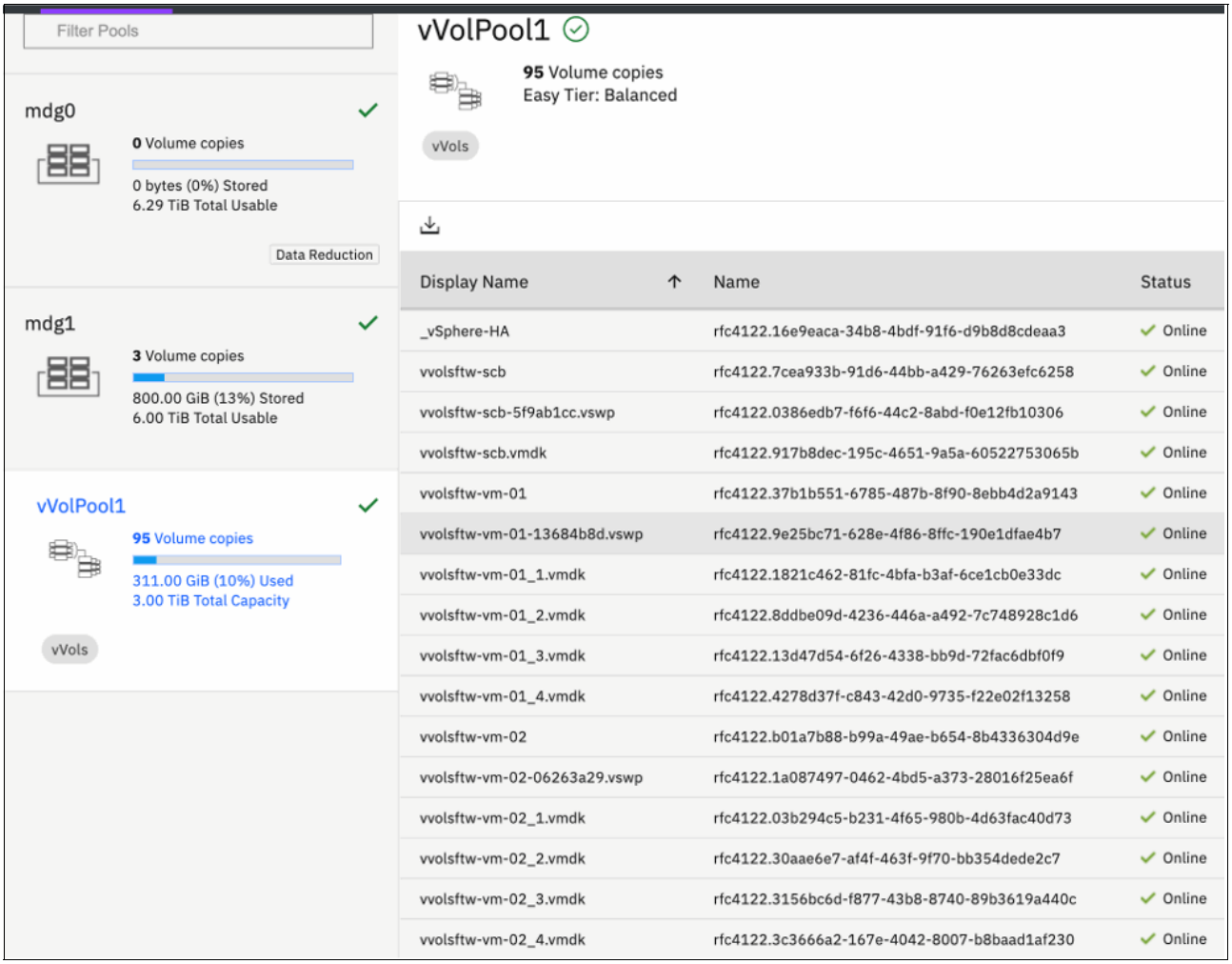


Figure 6-43 Selecting the vVol-enabled child pool

6. The Display Name column was added to provide some more information about the vVol that can help to identify the specific VM to which it belongs (Figure 6-44).

Note: By default, the Name column is not displayed in the GUI table view, but it is an optional field that can be added by right-clicking the column headers and selecting the **Name** checkbox.

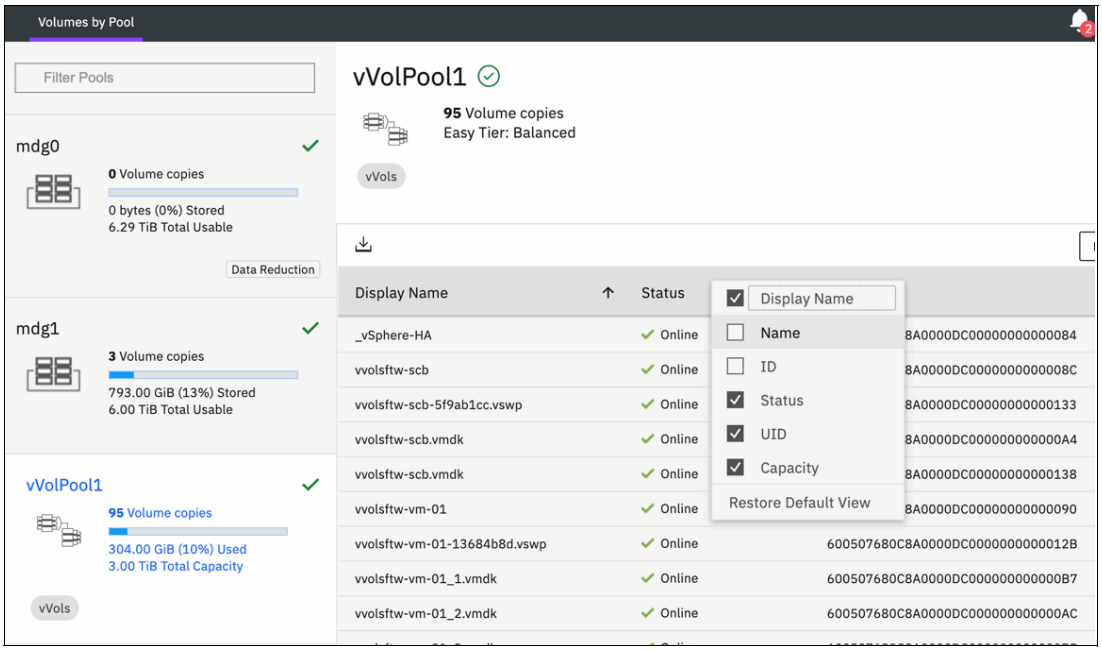


Figure 6-44 Displaying the Name column

7

IBM Storage Insights

This chapter describes IBM Storage Insights integration with VMware.

IBM Storage Insights is an IBM Cloud® software as a service (SaaS) offering that can help you monitor and optimize the storage resources in the system and across your data center.

IBM Storage Insights provides cognitive support capabilities, monitoring, and reporting for storage systems, switches and fabrics, and VMware Elastic Sky X integrated (ESXi) hosts in a single dashboard.

IBM Storage Insights provides the following features:

- Enterprise monitoring dashboard
- Device event alerting
- Performance checking
- Capacity checking
- Service ticket processing
- Streamlined uploading of diagnostic data

This chapter includes the following sections:

7.1 IBM Storage Insights editions

Two versions of IBM Storage Insights are available: IBM Storage Insights and IBM Storage Insights Pro (Table 7-1).

- IBM Storage Insights is available at no additional charge to owners of IBM block storage systems who sign up. IBM Storage Insights provides an environment overview, integration into support processes, and shows you IBM analysis results.
- The IBM Storage Insights Pro capacity-based subscription version includes all IBM Storage Insights no-charge functions, plus more information, a longer history, and more capabilities through a monthly subscription.

Table 7-1 Different features of both versions

Resource management	Functions	IBM Storage Insights	IBM Storage Insights Pro (subscription)
Monitoring	Inventory management	IBM block storage, switches, fabrics, and VMware ESXi hosts	IBM and non IBM block storage, file storage, object storage, switches, fabrics, and VMware ESXi hosts
	Logical configuration	Basic	Advanced
	Health	Call Home events	Call Home events
	Performance	Basic: ► 3 storage system metrics: I/O rate, data rate, and response times aggregated for storage systems ► 4 switch metrics: port saturation, port congestion, port hardware errors, and port logical errors ► 3 host metrics: host I/O rate, host data rate, and host response time ► 3 virtual machine (VM) metrics: VM I/O rate, VM data rate, and VM response time	Advanced: ► 100+ metrics for storage systems and their components ► 40+ metrics for switches and related components ► 10+ metrics for hosts and related components ► 10+ metrics for VMs and related components

Resource management	Functions	IBM Storage Insights	IBM Storage Insights Pro (subscription)
Monitoring (cont.)	Capacity	Basic: ► 4 metrics: used capacity, available capacity, total capacity, and compression savings aggregated for storage systems ► 2 host metrics: storage area network (SAN) capacity and used SAN capacity ► 2 VM metrics: SAN capacity and used SAN capacity	Advanced: ► 25+ metrics for storage systems and their components ► 10+ metrics for hosts and related components ► 10+ metrics for VMs and related components
	Drill-down performance workflows to enable deep troubleshooting		✓
	Explore virtualization relationships.		✓
	Explore replication relationships.		✓
	Retention of configuration and capacity data	Only the last 24 hours are shown.	2 years
	Retention of performance data	Only the last 24 hours are shown.	1 year
	Exporting performance data to a file		✓
Service	Filter events to quickly isolate trouble spots.	✓	✓
	Hassle-free log collection	✓	✓
	Simplified ticketing	✓	✓
	Show active Problem Management Reports (PMRs) and ticket history.	✓	✓

Resource management	Functions	IBM Storage Insights	IBM Storage Insights Pro (subscription)
Reporting	Inventory, capacity, performance, and storage consumption reports	► Capacity reports for block storage systems and pools ► Inventory reports for block storage systems, switches, chassis, and switch ports	All reports
Alerting and Analytics	Predictive alerts	✓	✓
	Customizable, multi-conditional alerting, including alert policies		✓
	Performance planning		✓
	Capacity planning		✓
	Business impact analysis (applications, departments, and groups)		✓
	Optimize data placement with tiering.		✓
	Optimize capacity with reclamation.		✓
Security	ISO/IEC 27001 Information Security Management standards certified	✓	✓
Entitlements		No additional charge	Capacity-based subscription

Restriction: You must have a current warranty or maintenance agreement for an IBM block storage system to open tickets and send log packages.

IBM Storage Insights for IBM Spectrum Control

IBM Storage Insights for IBM Spectrum Control is an IBM Cloud service that can help you predict and prevent storage problems before they impact your business. It is complementary to IBM Spectrum Control.

As an on-premises application, IBM Spectrum Control does not send the metadata about monitored devices off-site, which is ideal for dark shops and sites that do not want to open ports to the cloud. However, if your organization allows for communication between its network and the cloud, you can use IBM Storage Insights for IBM Spectrum Control to transform your support experience for IBM block storage.

IBM Storage Insights for IBM Spectrum Control is like IBM Storage Insights Pro in capability, and it is available for no additional cost if you have an active license with a current subscription and support agreement for IBM Virtual Storage Center, IBM Spectrum Storage Suite, or any edition of IBM Spectrum Control.

7.2 IBM Storage Insights architecture

IBM Storage Insights provides a lightweight data collector that is deployed on a Linux, Windows, or IBM AIX server, or a guest in a VM (for example, a VMware guest).

The data collector streams performance, capacity, asset, and configuration metadata to your IBM Cloud instance.

The metadata flows in one direction, that is, from your data center to IBM Cloud over HTTPS. In the IBM Cloud, your metadata is protected by physical, organizational, access, and security controls. IBM Storage Insights is ISO/IEC 27001 Information Security Management certified.

Figure 7-1 shows the architecture of the IBM Storage Insights application, the supported products, and the three main teams who can benefit from using the tool.

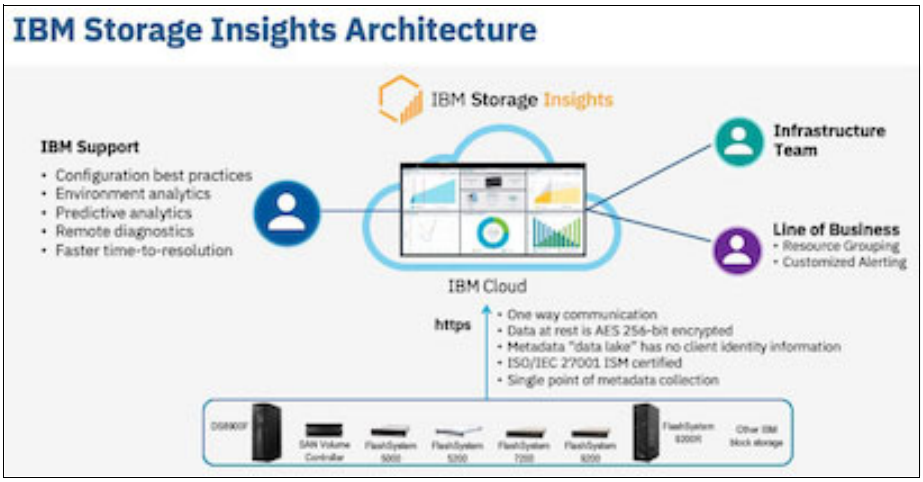


Figure 7-1 IBM Storage Insights architecture

For more information about IBM Storage Insights and to sign up and register for the no-charge service, see the following resources:

- Fact sheet
- Demonstration
- Security guide

7.3 IBM Storage Insights Monitoring

With IBM Storage Insights, you get the information that you need to monitor the health of your block storage environment and fabrics on the Operations dashboard, as shown in Figure 7-2.

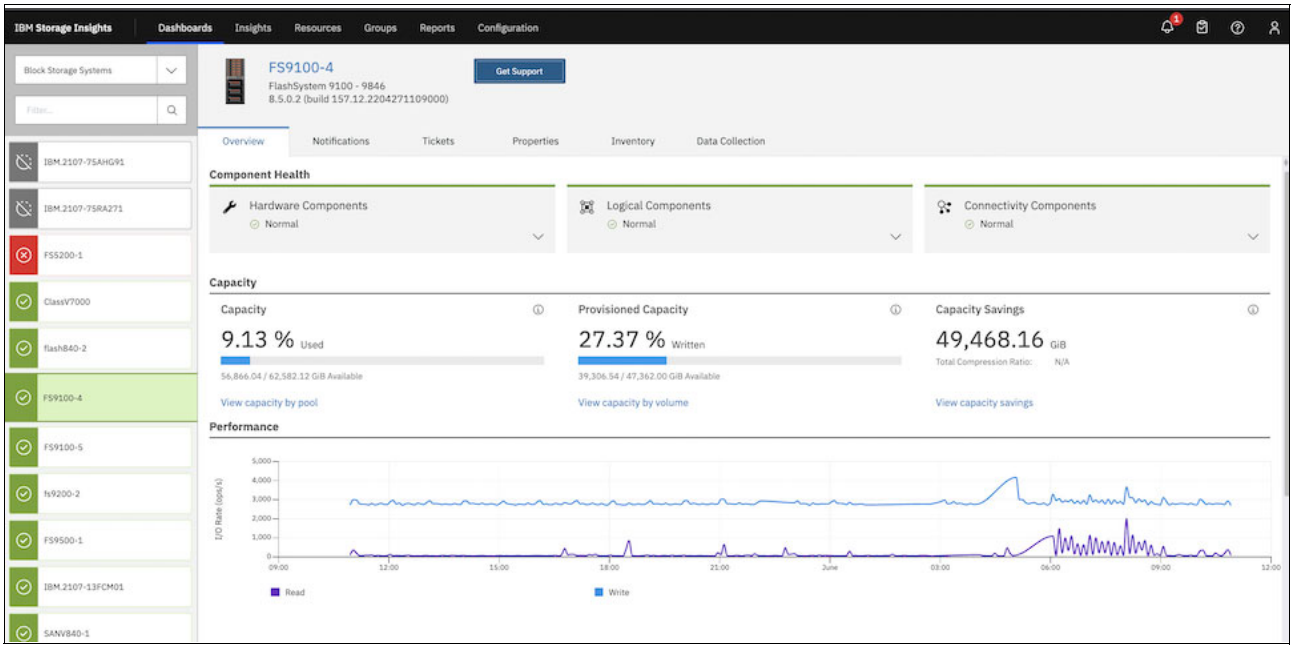


Figure 7-2 IBM Storage Insights System overview for block storage

The block storage dashboard is a default one that is shown when you go to the Operations dashboard. To take a close look at the storage systems that are being monitored, select **Dashboards** → **Operations**. Then, click the storage system in which you are interested in the list at the left of the dashboard. For example, you can see the health, capacity, and performance information for a block storage system in the Operations dashboard. The storage system is colored red because there are problems with nodes and logical components.

IBM Storage Insights supports Brocade and Cisco switches and fabrics so that you can detect and investigate performance issues throughout your storage environment. You can follow the trail of storage requests through the components in the SAN fabric to the target storage systems.

For more information about IBM Storage Insights, see Key features on IBM Storage Insights Documentation.

7.4 IBM Storage Insights VMware integration

You can add VMware ESXi hosts and their VMs for monitoring with IBM Storage Insights by specifying connection information for a vCenter Server. When you add VMware ESXi hosts, then you can collect data, generate reports, and manage storage that is related to hosts and VMs.

Identify the IP addresses and user credentials of the vCenter Servers that manage the VMware ESXi hosts and VMs that you want to monitor by completing the following steps. IBM Storage Insights uses this information to connect to the vCenter Servers and discover their managed VMware ESXi hosts and VMs. You can add multiple vCenter Servers concurrently if they share the same user credentials.

1. In the menu bar, select **Resources** → **Hosts**.
2. Click **Add vCenter Server**.
3. Enter the IP addresses or hostnames that you use to connect to the vCenter Server, and then enter the authentication credentials such as username and password that are shared by all the vCenter Servers that you are adding, as shown in Figure 7-3.

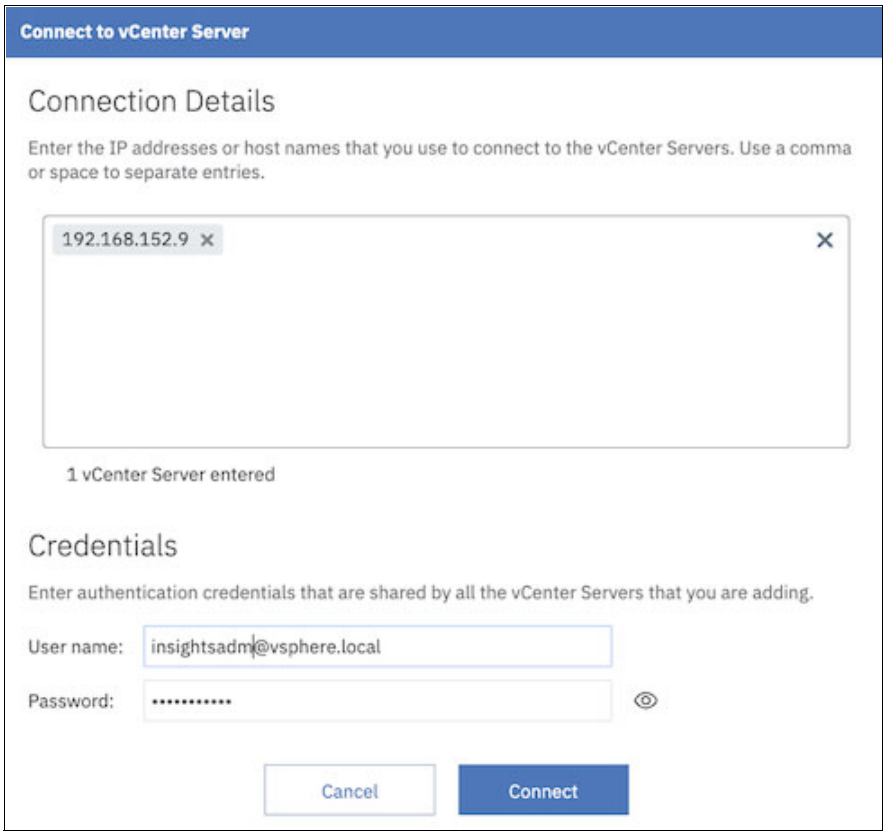


Figure 7-3 Adding vCenter Server

Note: When you add vCenter Servers for monitoring, you must specify the connection credentials of users that are used to collect metadata. The users must meet the following requirements:

- Role: Read Only (minimum). For example, the Administrator role or Virtual Machine Power User role.
- Privilege: Browse data store.

After the VCenter server initial discovery starts, all ESXi hosts and VMs that are managed by VCenter are discovered by IBM Storage Insights. You can see details of each ESXi host on the IBM Storage Insights Pro edition, as shown in Figure 7-4.

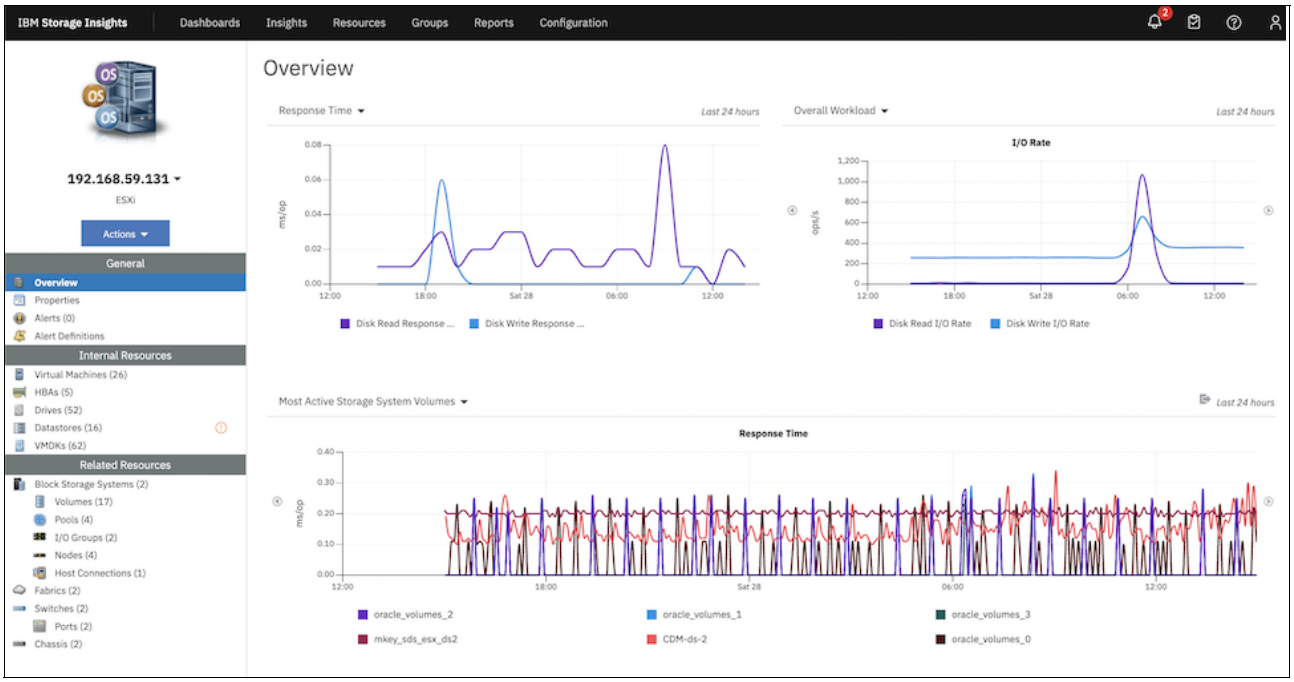


Figure 7-4 ESXi host details with daily response time, IO rate, and most active volumes on IBM Storage Insights

End-to-end SAN connectivity is visible when IBM Storage systems, SAN switches, and VMware ESXi servers are added to IBM Storage Insights.

Most of the time, it is a complex process to find which VM creates heavy I/O on a data store volume. IBM Storage Insights Pro can monitor virtual machine disk (VMDK) level I/O performance, and you can find the heavy loaded VM, as shown in Figure 7-5.

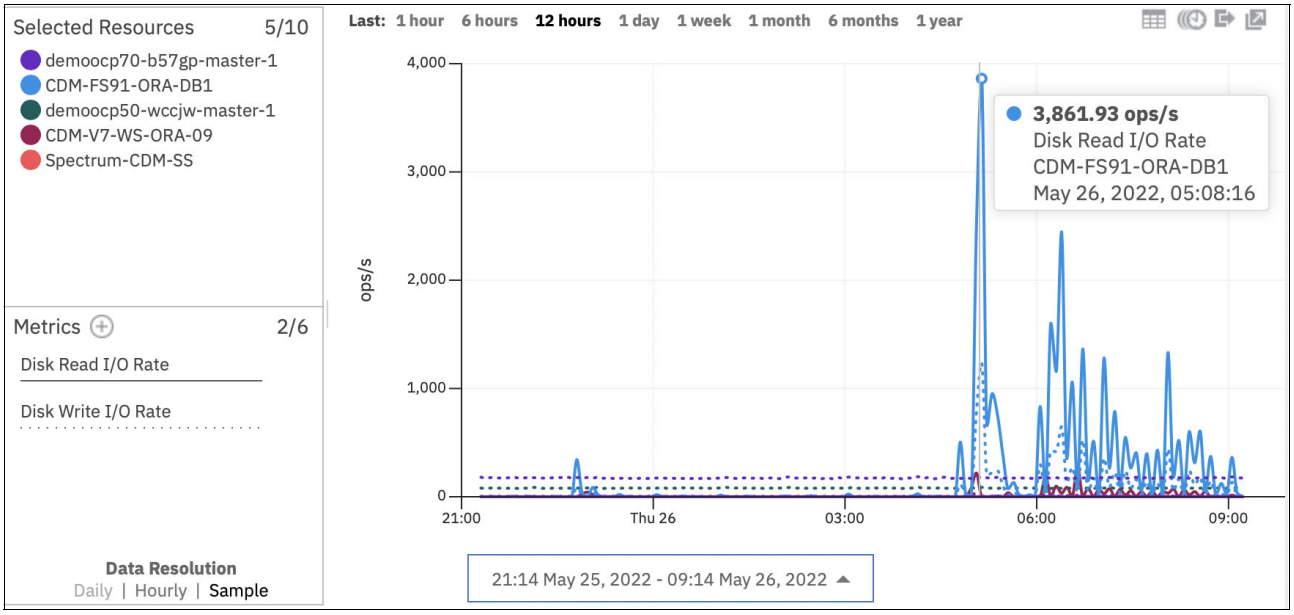


Figure 7-5 VMDK level performance monitoring on IBM Storage Insights Pro

8

Troubleshooting

This chapter provides information to troubleshoot common problems that can occur in an IBM FlashSystem and VMware Elastic Sky X integrated (ESXi) environment. It also explains how to collect the necessary problem determination data.

This chapter includes the following sections:

8.1 Collecting data for support

This section discusses the data that needs to be collected before contacting support for assistance. When interacting with support, it is important to provide a clear problem description that empowers the support engineers to help resolve the issue. A good problem description includes:

- What was expected?
- What was not expected?
- What are the resources that are involved (volumes, hosts, and so forth)?
- When did the problem take place?

8.1.1 Data collection guidelines for SAN Volume Controller and IBM FlashSystem

On SAN Volume Controller (SVC) and IBM FlashSystem, system logs can be collected in the product GUI by selecting **Settings** → **Support Package** (Figure 8-1).

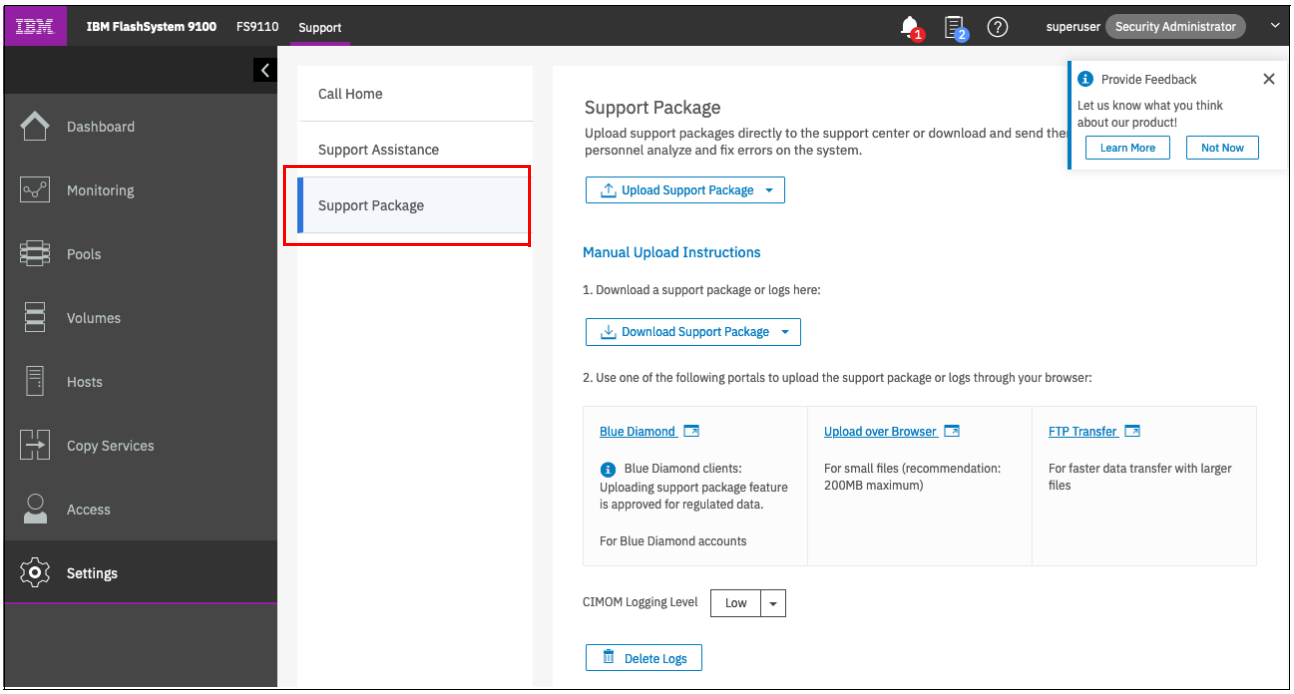


Figure 8-1 Collecting a support package in the GUI

For more information about the level of logs to collect for various issues, see What Data Should You Collect for a Problem on IBM Spectrum Virtualize systems?

For the topics covered in the scope of this document, you typically need to gather a snap (option) 4, which contains standard logs plus new statesaves. Because this data often takes a long time to collect, it might be advantageous to manually create the statesaves, and then collect the standard logs afterward. This task can be done by using the `svc_livedump` command-line interface (CLI) utility, which is available in the product command-line interface (Example 8-1 on page 179).

Example 8-1 Using svc_livedump to manually generate statesaves

```
IBM_FlashSystem:Cluster_9.42.162.160:superuser>svc_livedump -nodes all -y
Livedump - Fetching Node Configuration
Livedump - Checking for dependent VDisks
Livedump - Check Node status
Livedump - Preparing specified nodes  - this may take some time...
Livedump - Prepare node 1
Livedump - Prepare node 2
Livedump - Trigger specified nodes
Livedump - Triggering livedump on node 1
Livedump - Triggering livedump on node 2
Livedump - Waiting for livedumps to complete dumping on nodes 1,2
Livedump - Waiting for livedumps to complete dumping on nodes 2
Livedump - Successfully captured livedumps on nodes 1,2
```

After you generate the necessary statesaves, collect standard logs and the latest statesaves (option 3), and use the GUI to create a support package including the manually generated livedumps. Alternatively, you can create the support package by using the CLI (Example 8-2).

Example 8-2 Using svc_snap to generate a support package in the CLI

```
IBM_FlashSystem:Cluster_9.42.162.160:superuser>svc_snap -gui3
Collecting data
Packaging files
Snap data collected in /dumps/snap.78E35HW-2.210329.170759.tgz
```

When the support package is generated by using the command line, you can download it by using the GUI or using a Secure Copy Protocol (SCP) client.

8.1.2 Data collection guidelines for VMware ESXi

For issues involving VMware ESXi hypervisor (including storage access errors), it is vital to ensure that the logs from the host-side of the connection are collected in addition to the storage subsystem. For the VMware instructions about collecting ESXi log packages, see Collecting diagnostic information for VMware ESXi (653).

When downloading a package for an ESXi host, the default settings provide the information that is needed to analyze most problems.

8.1.3 Data collection guidelines for VMware Site Recovery Manager

Troubleshooting problems that involve VMware Site Recovery Manager usually require the analyzing of data from the following sources:

1. The storage systems associated in all related sites as shown in 8.1.1, “Data collection guidelines for SAN Volume Controller and IBM FlashSystem” on page 178.
2. The IBM Storage Replication Adapter (SRA) appliance logs in all related sites.
3. The VMware Site Recovery Manager logs.

IBM SRA log collection

Current versions of the IBM SRA are deployed inside of the VMware Site Recovery Manager (SRM) server. By default, the SRA application logs all data in `/var/log/vmware/srm` on the SRM server where SRA is deployed.

VMware SRM log collection

For the VMware instructions for creating and downloading SRM logs, see Collecting diagnostic information for VMware Site Recovery Manager (1009253).

Important: When collecting data for problems that are related to SRM, make sure to collect data from all sites associated with the problem.

8.1.4 Data collection guidelines for IBM Spectrum Connect (VASA or vVols)

For troubleshooting issues associated with VASA or VMware vSphere Virtual Volume (vVol), the following sets of data are required for troubleshooting:

1. A support package from the storage system as shown in 8.1.1, “Data collection guidelines for SAN Volume Controller and IBM FlashSystem” on page 178.
2. A support package from IBM Spectrum Connect.
3. A support package from the management application interfacing with IBM Spectrum Connect.
4. If the problem includes access to the data, ESXi logs as shown in 8.1.2, “Data collection guidelines for VMware ESXi” on page 179.

Collecting Data for IBM Spectrum Connect

IBM Spectrum Connect logs can be collected in the following two ways:

1. Using the Operating System Shell

   IBM Spectrum Connect stores data in `/var/log/sc` by default. Copy the contents of this directory off the system for use.

2. Using the IBM Spectrum Connect User Interface

   In the IBM Spectrum Connect User Interface, select **Settings** → **Support** → **Collect Log** to gather and download the log files (Figure 8-2).

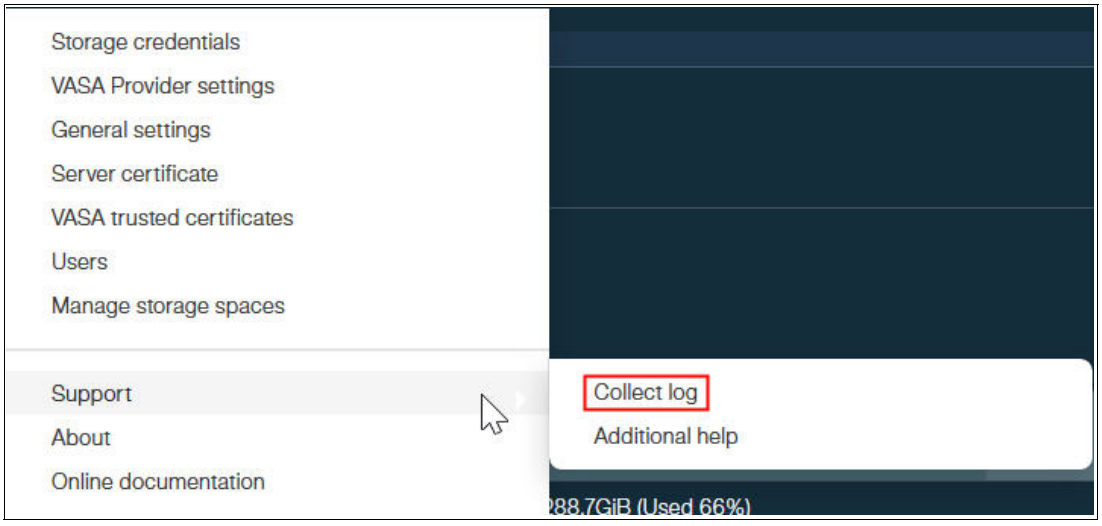


Figure 8-2 Collecting IBM Spectrum Connect logs

Collecting data for VMware vCenter

vCenter logs can be collected by using the same process as ESXi hosts, as described in 8.1.2, “Data collection guidelines for VMware ESXi” on page 179. The difference is when selecting resources for which to collect logs, select the vCenter server instead of (or in addition to) an ESXi host.

Collecting data for VMware vRealize Orchestrator

For the VMware data collection instructions for the VMware vRealize Orchestrator (vRO), see Generating a log bundle from command line for a vRealize Orchestrator 7.x appliance (2150664).

Collecting data for VMware vRealize Operations Manager

For the VMware data collection instructions for the VMware vRealize Operations (vROps) Manager, see Collecting diagnostic information from vRealize Operations (2074601).

8.2 Common support cases

This section describes topics that are commonly raised to the support center. This section is not meant to be a comprehensive guide on debugging interoperability issues between SVC, IBM FlashSystem, and VMware products.

8.2.1 Storage loss of access

When troubleshooting the loss of access to storage, it is important to properly classify how access was loss and what resources are involved.

The three general categories of storage loss of access events in VMware products are:

- All paths down (APD)
- Permanent device loss (PDL)
- Virtual machine (VM) crash

All Paths Down

An APD event takes place when all the paths to a data store are marked offline. Example 8-3 shows the `vmkernel` log signature for an APD event.

Example 8-3 ESXi All Paths Down log signature

```
cpu1:2049)WARNING: NMP: nmp_IssueCommandToDevice:2954:I/O could not be issued to
device "naa.600507681081025a1000000000000003" due to Not found
cpu1:2049)WARNING: NMP: nmp_DeviceRetryCommand:133:Device
"naa.600507681081025a1000000000000003": awaiting fast path state update for
failover with I/O blocked. No prior reservation exists on the device.
cpu1:2049)WARNING: NMP: nmp_DeviceStartLoop:721:NMP Device
"naa.600507681081025a1000000000000003" is blocked. Not starting I/O from device.
cpu1:2642)WARNING: NMP: nmpDeviceAttemptFailover:599:Retry world failover device
"naa.600507681081025a1000000000000003" - issuing command 0x4124007ba7c0
cpu1:2642)WARNING: NMP: nmpDeviceAttemptFailover:658:Retry world failover device
"naa.600507681081025a1000000000000003" - failed to issue command due to Not found
(APD), try again...
cpu1:2642)WARNING: NMP: nmpDeviceAttemptFailover:708:Logical device
"naa.600507681081025a1000000000000003": awaiting fast path state update...
```

When all paths are lost, if there is no path update lasting through the `Misc.APDTimeout` value (default of 140 seconds), then the APD condition is latched. Example 8-4 shows the log signature in the `vobd.log` file for a latched APD state.

Example 8-4 ESXi All Paths Down timeout

```
[APDCorrelator] 2682686563317us: [esx.problem.storage.apd.timeout] Device or
filesystem with identifier [11ace9d3-7bebe4e8] has entered the All Paths Down
Timeout state after being in the All Paths Down state for 140 seconds. I/Os will
now be fast failed.
```

These issues are typically the result of errors in path recovery. Corrective actions include:

- Validating the best practice multipathing configuration is in use as shown in 2.3, “Multi-path considerations” on page 18.
- Validate all server driver and firmware levels are at the latest supported level.
- Validate the network infrastructure connecting the host and storage is operating correctly.

Permanent device loss

A PDL event is the response to a unrecoverable I/O error that is returned by a storage controller. Example 8-5 shows the `vmkernel` log signature for a PDL event.

Example 8-5 ESXi permanent device loss log signature

```
cpu17:10107)WARNING: Vol3: 1717: Failed to refresh FS
4beb089b-68037158-2ecc-00215eda1af6 descriptor: Device is permanently unavailable
cpu17:10107)ScsiDeviceIO: 2316: Cmd(0x412442939bc0) 0x28, CmdSN 0x367bb6 from
world 10107 to dev "naa.600507681081025a1000000000000003" failed H:0x0 D:0x2 P:0x0
Valid sense data: 0x2/0x3e/0x1
cpu17:10107)Vol3: 1767: Error refreshing PB resMeta: Device is permanently
unavailable
```

For a list of I/O errors that trigger PDL, see Permanent Device Loss (PDL) and All-Paths-Down (APD) in vSphere 6.x and 7.x (2004684).

These types of events are often the result of a hardware failure or low-level protocol error in the server host bus adapter (HBA), storage area network (SAN), or the storage array. If hardware errors are found that match the time in which the PDL happens, PDF is likely the cause.

Virtual machine crash

If a VM fails in absence of an APD or PDL event, then this scenario should be treated as an operating system or application failure inside the VM. If the analysis of the guest VM points to a storage I/O timeout, then this analysis might point to latency in processing VM I/O requests. In such situations, it is important to review the following sets of data:

- The `vmkernel` log of the ESXi host that houses the VM that failed. Specifically, look for events that are related to the physical device backing the data store that houses the VM.
- The storage array’s performance data. Specifically, check for peak read-and-write latency during the time when the VM failed.
- Operating System and application logs for the VM that failed. Specifically, identify key timeout values and the time of the crash.

8.2.2 VMware migration task failures

Two types of migrate tasks are as follows:

- *vMotion* is a migrate task that is used to move the running state of the VM (for example, memory and compute resource) between ESXi hosts.
- *Storage vMotion* is a migrate task that is used to move the VM storage resources between data stores, for example VMDK files data stores.

vMotion tasks

vMotion tasks are largely dependent on the Ethernet infrastructure between the ESXi hosts. The only real storage interaction is at the end when file locks must move from one host to another. In this phase, it is possible for Small Computer System Interface (SCSI) Reservation Conflicts or file lock contention to result in the failing of the migration task. The following articles describe the most frequent issues:

- Investigating virtual machine file locks on ESXi hosts (10051)
- Resolving SCSI reservation conflicts (1002293)
- IBM Spectrum Virtualize APAR HU01894

Storage vMotion tasks

Storage vMotion tasks are primarily dependent on storage throughput. When moving between data stores in the same storage controller, the task is typically offloaded to the storage array by using extended copy (XCOPY) (VAAI Hardware Accelerated Move). If the migration task is between storage systems, the copy is performed by using standard read and write commands.

The default timeout for the task to complete is 100 seconds. If the migration takes longer than 100 seconds to complete, then the task fails with a timeout, as shown in Example 8-6.

Example 8-6 VMware Log Storage vMotion timeout

```
vmkernel: 114:03:25:51.489 cpu0:4100)WARNING: FSR: 690: 1313159068180024 S:
Maximum switchover time (100 seconds) reached. Failing migration; VM should resume
on source.
vmkernel: 114:03:25:51.489 cpu2:10561)WARNING: FSR: 3281: 1313159068180024 D: The
migration exceeded the maximum switchover time of 100 seconds. ESX has
preemptively failed the migration to allow the VM to continue running on the
source host.
vmkernel: 114:03:25:51.489 cpu2:10561)WARNING: Migrate: 296: 1313159068180024 D:
Failed: Maximum switchover time for migration exceeded(0xbad0109) @0x41800f61cee2
```

The task is generic by nature and the root cause behind the timeout typically requires performance analysis of the storage arrays that are involved and a detailed review of the ESXi logs for the host performing the task. In some circumstances, it might be appropriate to increase the default timeout, as described at Using Storage Motion to migrate a virtual machine with many disks fails without timeout (1010045).

Abbreviations and acronyms

ABR	array-based replication
ACS	Advanced Copy Services
ALUA	Asymmetric Logical Unit Access
APD	all paths down
API	application programming interface
ATS	Atomic Test and Set
BAU	business as usual
BE	Back-End
CAW	Compare and Write
CIM	Common Information Model
CIMOM	Common Information Model Object Manager
CLI	command-line interface
CPU	central processing unit
CSP	cloud service provider
CV	Change Volume
DR	disaster recovery
DRP	data reduction pool
DRS	Distributed Resource Scheduler
ESX	Elastic Sky X
ESXi	Elastic Sky X integrated
EZT	Eager Zero Thick
FC	Fibre Channel
FC-NVMe	Non-Volatile Memory Express over Fibre Channel
FCM	FlashCore Module
FCP	Fibre Channel Protocol
FE	Front-End
FQDN	fully qualified domain name
FT	Fault Tolerance
GMCV	Global Mirror with Change Volumes
GUID	globally unique identifier
HA	high availability
HBA	host bus adapter
HPP	high-performance plug-in
IBM	International Business Machines Corporation
IOPS	input/output operations per second
IQN	iSCSI Qualified Name
iSCSI	Internet Small Computer System Interface
iSER	iSCSI Extensions for RDMA
ITIL	IT Infrastructure Library
iWARP	Internet Wide-Area RDMA Protocol
LB-BYTES	Load Balance - Bytes
LB-IOPS	Load Balance - IOPS
LB-Latency	Load Balance - Latency
LB-RR	Load Balance - Round-Robin
LDAP	Lightweight Directory Access Protocol
LSA	Log Structured Array
LU	logical unit
LUN	logical unit number
MBR	master boot record
MDisk	managed disk
MRU	Most Recently Used
MSCS	Microsoft Cluster Server
NIC	network interface card
NL-SAS	near-line SAS
NMP	native multipathing
NPIV	N_Port ID virtualization
NTP	Network Time Protocol
NVMe-oF	NVMe over Fabrics
NVMe	Non-Volatile Memory Express
OS	operating system
PDL	permanent device loss
PE	Protocol Endpoint
PFE	product field engineer
PMP	Project Management Professional
PMR	Problem Management Report
PSA	Pluggable Storage Architecture
PSP	path selection policy or path-selection plug-in
PSS	Path Selection Scheme
PVSCSI	Paravirtual SCSI
RDM	Raw Device Mapping
RDMA	Remote Direct Memory Access
RFC	Request For Comment
RHEL	Red Hat Enterprise Linux
RNIC	RDMA-capable Ethernet NIC
RoCE	RDMA over Converged Ethernet
RR	Round-Robin

RtC	Real-time Compression
RTO	recovery time objective
RTT	round-trip time
SaaS	software as a service
SAN	storage area network
SAS	serial-attached SCSI
SCM	Storage Class Memory
SCP	Secure Copy Protocol
SCSI	Small Computer System Interface
SDRS	Storage Distributed Resource Scheduler
SDS	software-defined storage
SIOC	Storage I/O Control
SME	subject matter expert
SRA	Storage Replication Adapter
SRM	VMware Site Recovery Manager
SSD	solid-state drive
SSH	Secure Shell
SSIC	System Storage Interoperation Center
SVC	SAN Volume Controller
UI	user interface
UID	unique identifier
VAAI	vSphere Storage APIs – Array Integration
VASA	vSphere Storage APIs - Storage Awareness
vCSA	vCenter Service Appliance
VLAN	virtual local area networks
VM	virtual machine
VMDK	virtual machine disk
VMFS	Virtual Machine File System
vMSC	VMware vSphere Metro Storage Cluster
VR	vSphere Replication
vRO	VMware vRealize Orchestrator
vROps	VMware vRealize Operations
VSAN	virtual storage area network
vVol	VMware vSphere Virtual Volume
vWC	vSphere Web Client
WWNN	worldwide node name
WWPN	worldwide port name
XCOPY	extended copy

Related publications

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

IBM Redbooks

The following IBM Redbooks publications provide more information about the topics in this document. Some publications that are referenced in this list might be available in softcopy only.

- *IBM FlashSystem Best Practices and Performance Guidelines*, SG24-8503
- *IBM Spectrum Virtualize HyperSwap SAN Implementation and Design Best Practices*, REDP-5597
- *IBM Storage and the NVM Express Revolution*, REDP-5437
- *Implementing IBM FlashSystem with IBM Spectrum Virtualize V8.4*, SG24-8492
- *Introduction and Implementation of Data Reduction Pools and Deduplication*, SG24-8430

You can search for, view, download, or order these documents and other Redbooks, Redpapers, web docs, drafts, and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- Brocade Peer Zoning:

  https://docs.broadcom.com/doc/FOS-90x-Admin-AG
- *Cisco MDS 9000 NX-OS Configuration Guide*:

  https://www.cisco.com/c/en/us/td/docs/switches/datacenter/mds9000/sw/5_2/configuration/guides/fabric/nx-os/nx_os_fabric/zone.html
- Data collection instructions for the VMware vRealize Operations Manager:

  https://kb.vmware.com/s/article/2074601
- Data collection instructions for the VMware vRealize Orchestrator:

  https://kb.vmware.com/s/article/2150664
- Frequently asked questions about Storage Distributed Resource Scheduler (DRS):

  https://kb.vmware.com/s/article/2149938
- IBM Documentation for IBM FlashSystem 9100 and 9200:

  https://core.vmware.com/reference-architectures
- IBM FlashSystem v8.4 Support Matrix:

  https://www.ibm.com/support/pages/node/3543675

- IBM guidance on what level of logs to collect for various issues:

  https://www.ibm.com/support/pages/node/690179
- IBM Spectrum Virtualize Family Storage Replication Adapter Documentation:

  https://www.ibm.com/support/knowledgecenter/SSEQ4E/landing/SVF_SRA_welcome_page.html
- IBM Support article on host disconnects by using VMware vSphere 5.5.0 Update 2 and vSphere 6.0:

  http://www.ibm.com/support/docview.wss?uid=ssg1S1005201
- IBM System Storage Interoperation Center (SSIC):

  https://www.ibm.com/systems/support/storage/ssic/interoperability.wss
- Information for storage I/O control on data stores that are used by production databases:

  https://blogs.vmware.com/apps/2015/12/sioc-for-bca-good-idea.html
- Internet Small Computer System Interface (iSCSI) standard definition:

  https://tools.ietf.org/html/rfc3720
  https://www.ibm.com/support/knowledgecenter/STSLR9_8.4.0/com.ibm.fs9200_8401.doc/fs9200_ichome.html
- VMware Knowledge Base article on input/output operations per second (IOPS) limits:

  https://kb.vmware.com/s/article/2069356
- VMware Reference Architectures website:

  https://core.vmware.com/reference-architectures
- VMware Site Recovery Manager documentation:

  https://www.vmware.com/support/pubs/srm_pubs.html
- VMware guidance on collecting ESXi log packages:

  https://kb.vmware.com/s/article/653

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Redbooks

IBM FlashSystem and VMware Implementation and Best Practices Guide

(0.2”spine)
0.17”<->0.473”
90<->249 pages

SG24-8505-01

ISBN 0738460869

Printed in U.S.A.