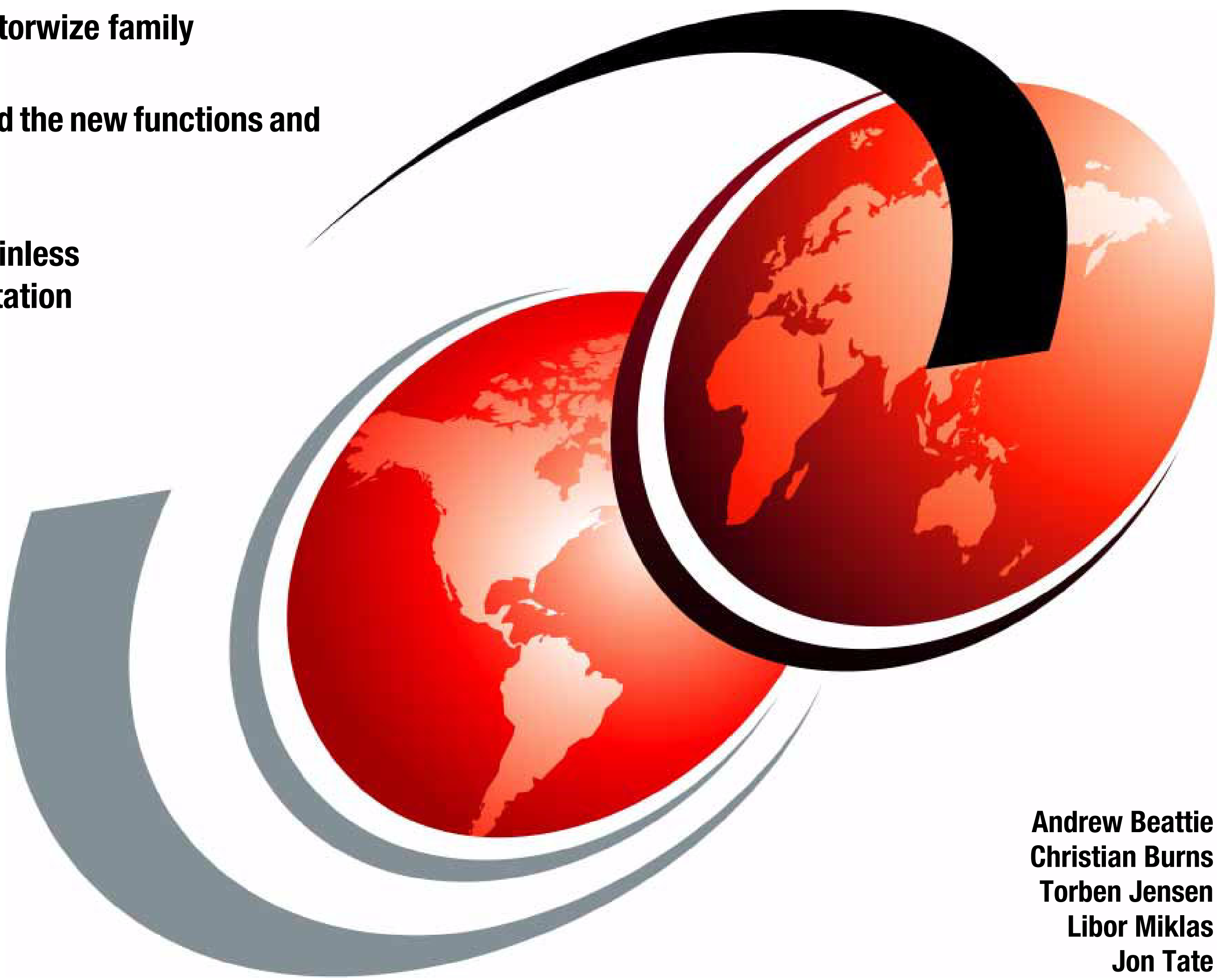
IBM SAN Volume Controller 2145-DH8 Introduction and Implementation

Learn about the latest addition to the SVC and Storwize family

Understand the new functions and features

Enjoy a painless implementation



Andrew Beattie
Christian Burns
Torben Jensen
Libor Miklas
Jon Tate

Redbooks

ibm.com/redbooks

International Technical Support Organization

IBM SAN Volume Controller 2145-DH8 Introduction and Implementation

October 2014

SG24-8229-00

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

First Edition (October 2014)

This edition applies to software version 7.3 of the IBM System Storage SAN Volume Controller when used on the SAN Volume Controller 2145-DH8.

Contents

Notices

This information was developed for products and services offered in the U.S.A.

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, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: 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 states 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 supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

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.

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 the names and addresses used by an actual business enterprise 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.

Trademarks

Preface

Data is the new currency of business, the most critical asset of the modern organization. In fact, enterprises that can gain business insights from their data are twice as likely to outperform their competitors; yet, 72 percent of them have not started or are only planning big data activities. In addition, organizations often spend too much money and time managing where their data is stored. The average firm purchases 24% more storage every year, but uses less than half of the capacity it already has.

A member of the IBM® Storwize® family, IBM SAN Volume Controller (SVC) Data Platform is a storage virtualization system that enables a single point of control for storage resources to help support improved business application availability and greater resource utilization. The objective is to manage storage resources in your IT infrastructure and to make sure they are used to the advantage of your business, and do it quickly, efficiently, and in real time, while avoiding increases in administrative costs.

Virtualizing storage with SVC Data Platform helps make new and existing storage more effective. SVC Data Platform includes many functions traditionally deployed separately in disk systems. By including these in a virtualization system, SVC Data Platform standardizes functions across virtualized storage for greater flexibility and potentially lower costs.

SVC Data Platform functions benefit all virtualized storage. For example, IBM Easy Tier® optimizes use of flash storage. And IBM Real-time Compression™ enhances efficiency even further by enabling the storage of up to five times as much active primary data in the same physical disk space. Finally, high-performance thin provisioning helps automate provisioning. These benefits can help extend the useful life of existing storage assets, reducing costs.

Integrating these functions into SVC Data Platform also means that they are designed to operate smoothly together, reducing management effort.

In this IBM Redbooks® publication, we discuss the latest features and functions of the SVC 2145-DH8 and software version 7.3, implementation, architectural improvements, and Easy Tier.

Authors

This book was produced by a team of specialists from around the world working at the International Technical Support Organization, San Jose Center.

Andrew Beattie is a Storage Technical Sales Specialist for IBM Systems & Technology Group, in Brisbane, Australia. Before joining IBM in 2010, he worked with a number of IBM Business Partners, in both sales and technical roles with a focus on storage and virtual infrastructure solutions. Andrew has 15 years experience in infrastructure solution design and implementation.

Christian Burns is an IBM Storage Solution Architect based in New Jersey. As a member of the Storage Solutions Engineering team in Littleton, MA, he works with clients, IBM Business Partners, and IBMers worldwide, designing and implementing storage solutions that include various IBM products and technologies. Christian's areas of expertise include IBM Real-time Compression, SVC, IBM XIV®, and IBM FlashSystem™. Before joining IBM, Christian was the Director of Sales Engineering at IBM Storwize, before its becoming IBM Storwize. He brings over a decade of industry experience in the areas of sales engineering, solution design, and software development. Christian holds a BA degree in Physics and Computer Science from Rutgers College.

Torben Jensen is an IT Specialist at IBM Global Technology Services®, Copenhagen, Denmark. He joined IBM in 1999 for an apprenticeship as an IT-System Supporter. From 2001 to 2005, he was the client representative for IBM Internal Client platforms in Denmark. Torben started work with the SAN/DISK for open systems department in March 2005. He provides daily and ongoing support to customers as well as colleagues, and works with SAN designs and solutions for customers. Torben has written several Redbooks within the Storage virtualization area in the past.

Libor Miklas is an IT Architect working at the IBM Integrated Delivery Center in the Czech Republic. He demonstrates 10 years of extensive experience within the IT industry. During the last eight years, his main focus has been on the data protection solutions and on storage management and provisioning. Libor and his team design, implement, and support midrange and enterprise storage environments for various global and local clients, worldwide. He is an IBM Certified Deployment Professional for the IBM Tivoli® Storage Manager family of products and holds a Master's degree in Electrical Engineering and Telecommunications.

Jon Tate is a Project Manager for IBM System Storage® SAN Solutions at the International Technical Support Organization (ITSO), San Jose Center. Before joining the ITSO in 1999, he worked in the IBM Technical Support Center, providing Level 2 support for IBM storage products. Jon has 28 years of experience in storage software and management, services, and support, and is both an IBM Certified IT Specialist and an IBM SAN Certified Specialist. He is also the UK Chairman of the Storage Networking Industry Association.

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

Ian Boden
John Fairhurst
Katja Gebuhr
Paul Merrison
Gareth Nicholls
Evelyn Perez
Lee Sanders
Barry Whyte
IBM Hursley, UK

Nick Clayton
IBM Systems & Technology Group, UK

Bill Turnau
IBM Global Technology Services, North Carolina, US

Special thanks to the Brocade Communications Systems staff in San Jose, California for their unparalleled support of this residency in terms of equipment and support in many areas:

Jim Baldyga
Silviano Gaona
Brian Steffler
Marcus Thordal
Brocade Communications Systems

Now you can become a published author, too!

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

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

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

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

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  https://twitter.com/ibmredbooks

- Look for us on LinkedIn:

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

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

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

- Stay current on recent Redbooks publications with RSS Feeds:

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

1

Introduction to IBM storage virtualization

In this chapter, we introduce the concept of storage virtualization. Then, we present an overview explaining how you can apply virtualization to help address challenging storage requirements. Finally, we describe in detail all changes and enhancements of the IBM SAN Volume Controller version 7.3 and its associated hardware as the key milestone in the product development roadmap.

For a comparison, the brief overview and list of changes in the whole product path are outlined in Chapter 2, “Change history” on page 23.

1.1 Storage virtualization terminology

Although *storage virtualization* is a term that is used extensively throughout the storage industry, it can be applied to a wide range of technologies and underlying capabilities. In reality, most storage devices can technically claim to be virtualized in one form or another. Therefore, we must start by defining the concept of storage virtualization as used in this publication.

IBM interprets storage virtualization in the following manner:

- Storage virtualization is a technology that makes one set of resources look and feel like another set of resources, preferably with more desirable characteristics.
- It is a logical representation of resources that are not constrained by physical limitations:
  - It hides part of the complexity.
  - It adds or integrates new function with existing services.
  - It can be nested or applied to multiple layers of a system.

When discussing storage virtualization, it is important to understand that virtualization can be implemented at various layers within the I/O stack. We have to clearly distinguish between virtualization at the disk layer and virtualization at the file system layer.

The focus of this book is virtualization at the disk layer, which is referred to as *block-level virtualization*, or the *block aggregation layer*. A discussion of file system virtualization is beyond the scope of this book.

However, if you are interested in file system virtualization, see IBM General Parallel File System (GPFS™) or IBM Scale Out Network Attached Storage, which is based on GPFS.

To obtain more information and an overview of GPFS, visit the following website:

http://www.ibm.com/systems/technicalcomputing/platformcomputing/products/gpfs

More information about Scale Out Network Attached Storage is available at the following link:

http://www.ibm.com/systems/storage/network/sonas

The Storage Networking Industry Association’s (SNIA) block aggregation model (Figure 1-1 on page 3) provides a useful overview of the storage domain and its layers. The figure shows the three layers of a storage domain: the file, the block aggregation, and the block subsystem layers.

The model splits the block aggregation layer into three sublayers. Block aggregation can be realized within hosts (servers), in the storage network (storage routers and storage controllers), or in storage devices (intelligent disk arrays).

The IBM implementation of a block aggregation solution is the IBM System Storage SAN Volume Controller (SVC). The SVC is implemented as a clustered appliance in the storage network layer.

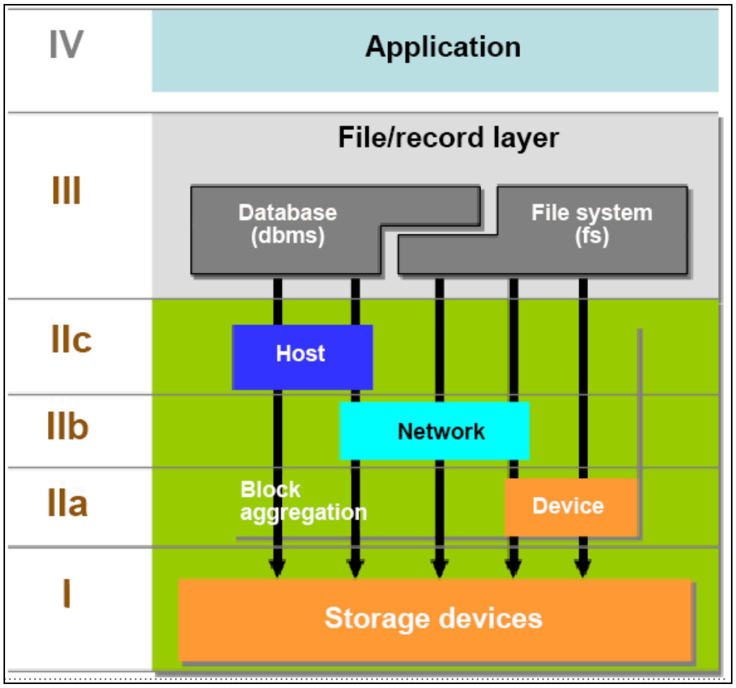


Figure 1-1 SNIA block aggregation model[1]

The key concept of virtualization is to decouple the storage from the storage functions that are required in the storage area network (SAN) environment.

Decoupling means abstracting the physical location of data from the logical representation of the data. The virtualization engine presents logical entities to the user and internally manages the process of mapping these entities to the actual location of the physical storage.

The actual mapping that is performed depends on the specific implementation, such as the granularity of the mapping, which can range from a small fraction of a physical disk, up to the full capacity of a physical disk. A single block of information in this environment is identified by its *logical unit number* (LUN), which is the physical disk, and an offset within that LUN, which is known as a *logical block address* (LBA).

The term *physical disk* is used in this context to describe a piece of storage that might be carved out of a Redundant Array of Independent Disks (RAID) array in the underlying disk subsystem.

Specific to the SVC implementation, the address space that is mapped between the logical entity is referred to as *volume*, and the physical disk is referred to as *managed disks* (MDisks).

Figure 1-2 on page 4 shows an overview of block-level virtualization.

[1] This figure is produced by the Storage Networking Industry Association.

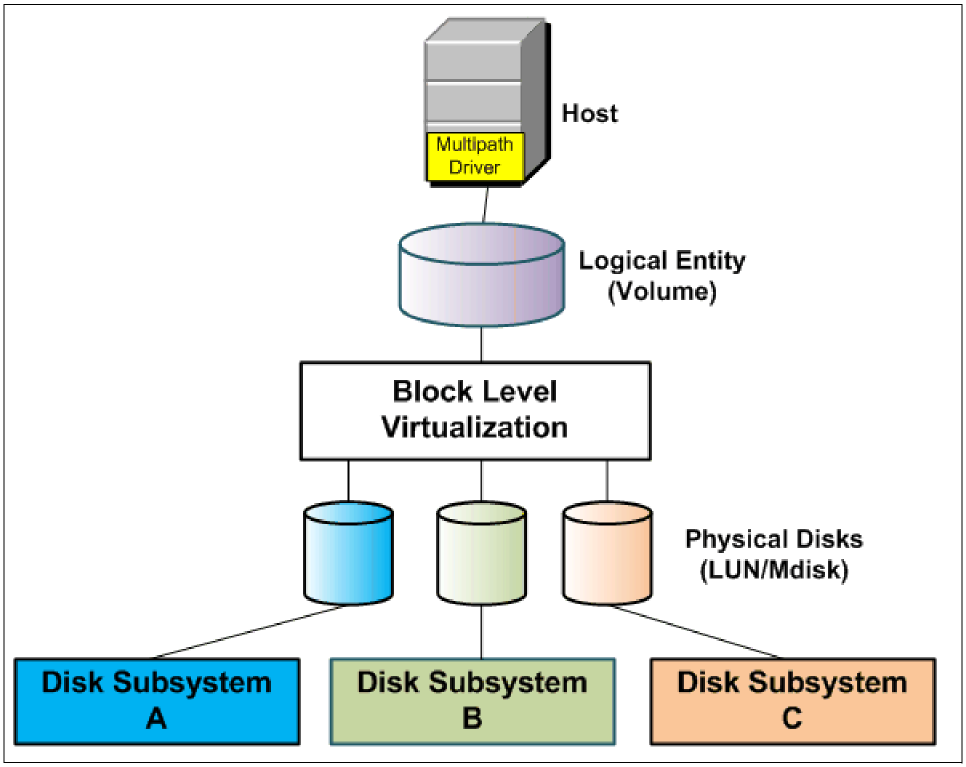


Figure 1-2 Block-level virtualization overview

The server and application are only aware of the logical entities, and they access these entities using a consistent interface that is provided by the virtualization layer.

The functionality of a volume that is presented to a server, such as expanding or reducing the size of a volume, mirroring a volume, creating an IBM FlashCopy®, and thin provisioning, is implemented in the virtualization layer. It does not rely in any way on the functionality that is provided by the underlying disk subsystem. Data that is stored in a virtualized environment is stored in a location-independent way, which allows a user to move or migrate data between physical locations, which are referred to as *storage pools*.

We refer to block-level storage virtualization as the *cornerstones of virtualization*. These cornerstones of virtualization are the core benefits that a product, such as the SAN Volume Controller, can provide over the traditional directly attached or SAN-attached storage.

The IBM SAN Volume Controller provides the following benefits:

- The SVC provides online volume migration while applications are running, which is possibly the greatest single benefit for storage virtualization. This capability allows data to be migrated on and between the underlying storage subsystems without any impact to the servers and applications. In fact, this migration is performed without the knowledge of the servers and applications that it even occurred.
- It simplifies storage management by providing a single image for multiple controllers and a consistent user interface for provisioning heterogeneous storage.
- The SVC provides enterprise-level Copy Services functions. Performing the Copy Services functions within the SVC removes dependencies on the storage subsystems, therefore enabling the source and target copies to be on other storage subsystem types.
- Storage utilization can be increased by pooling storage across the SAN and by enabling space reduction techniques such an IBM Real-time Compression or Thin Provisioning.

- System performance is often improved with the SVC as a result of volume striping across multiple arrays or controllers and the additional cache that it provides.

The SVC delivers these functions in a homogeneous way on a scalable and highly available platform, over any attached storage, and to any attached server.

1.2 Requirements driving storage virtualization

Today, an emphasis exists on a dynamic infrastructure being able to adapt the company needs for diversed data operations. Thus, there is a need for a storage environment that is as flexible as the application and server mobility. Business demands change quickly.

These key client concerns drive storage virtualization:

- Growth in data center costs
- Inability of IT organizations to respond quickly to business demands
- Poor asset utilization
- Poor availability or service levels
- Lack of skilled staff for storage administration

You can see the importance of addressing the complexity of managing storage networks by applying the total cost of ownership (TCO) metric to storage networks. Industry analyses show that storage acquisition costs are only about 20% of the TCO. Most of the remaining costs relate to managing the storage system.

But how much of the management of multiple systems, with separate interfaces, can be handled as a single entity? In a non-virtualized storage environment, every system is an “island” that needs to be managed separately.

1.2.1 Benefits of using IBM SAN Volume Controller

A storage virtualization is no longer merely a concept or an unproven technology. All major storage vendors offer storage virtualization products. Using storage virtualization as the foundation for a flexible and reliable storage solution helps enterprises to better align business and IT by optimizing the storage infrastructure and storage management to meet business demands.

The IBM System Storage SAN Volume Controller is a mature, eighth-generation virtualization solution that uses open standards and is consistent with the Storage Networking Industry Association (SNIA) storage model. The SVC is an appliance-based in-band block virtualization process, in which intelligence, including advanced storage functions, is migrated from individual storage devices to the storage network.

The IBM System Storage SAN Volume Controller can improve the utilization of your storage resources, simplify your storage management, and improve the availability of your applications.

SVC can reduce the number of separate environments that need to be managed down to a single environment. It provides a single interface for storage management. After the initial configuration of the storage subsystems, all of the day-to-day storage management operations are performed from the SVC.

Because the SVC provides advanced functions, such as mirroring and FlashCopy, there is no need to purchase them again for each new disk subsystem.

Today, it is typical that open systems run at less than 50% of the usable capacity that is provided by the RAID disk subsystems. Using the installed raw capacity in the disk subsystems will, depending on the RAID level that is used, show utilization numbers of less than 35%. A block-level virtualization solution, such as the SAN Volume Controller, can allow capacity utilization to increase to approximately 75 - 80%. With SVC, free space does not need to be maintained and managed within each storage subsystem, which further increases capacity utilization.

1.3 SAN Volume Controller architecture

The IBM System Storage SAN Volume Controller is a SAN block aggregation virtualization appliance that is designed for attachment to various host computer systems.

There are two major approaches in use today to consider for the implementation of block-level aggregation and virtualization:

- Symmetric: In-band appliance

  The device is a SAN appliance that sits in the data path, and all I/O flows through the device. This implementation is also referred to as *symmetric virtualization* or *in-band*.

  The device is both target and initiator. It is the target of I/O requests from the host perspective, and the initiator of I/O requests from the storage perspective. The redirection is performed by issuing new I/O requests to the storage. The SVC uses symmetric virtualization.

- Asymmetric: Out-of-band or controller-based

  The device is usually a storage controller that provides an internal switch for external storage attachment. In this approach, the storage controller intercepts and redirects I/O requests to the external storage as it does for internal storage. The actual I/O requests are themselves redirected. This implementation is also referred to as *asymmetric virtualization* or *out-of-band*.

Figure 1-3 on page 7 shows variations of the two virtualization approaches.

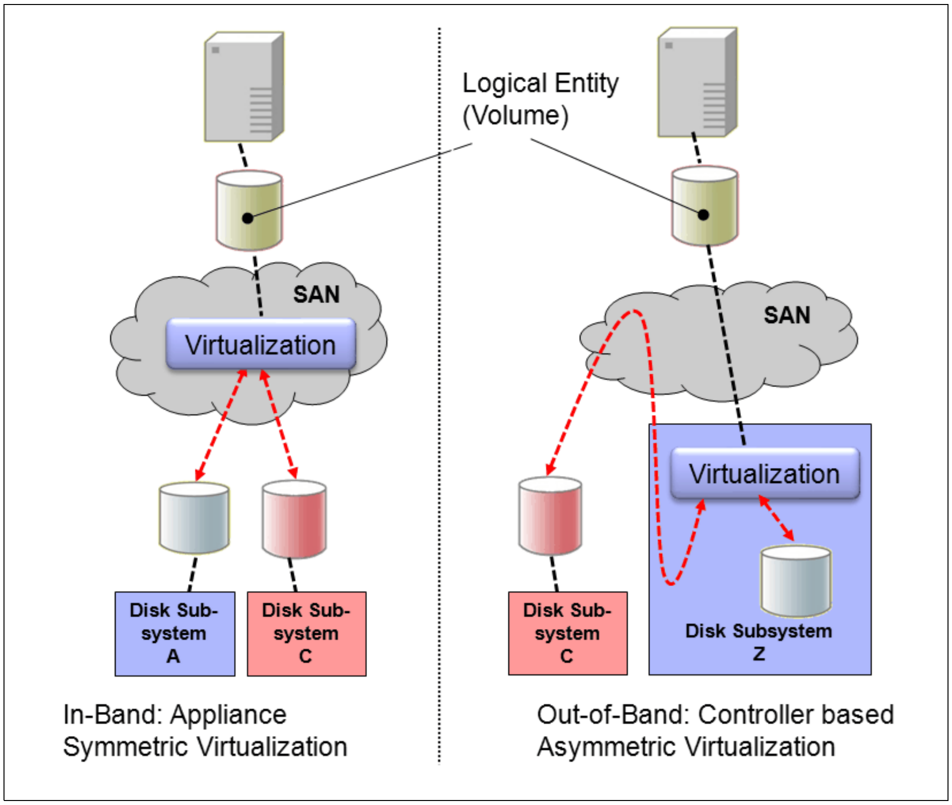


Figure 1-3 Overview of block-level virtualization architectures

Although these approaches provide essentially the same cornerstones of virtualization, there can be interesting side effects, as discussed here.

The controller-based approach has high functionality, but it fails in terms of scalability or upgradeability. Because of the nature of its design, there is no true decoupling with this approach, which becomes an issue for the lifecycle of this solution, such as with a controller. Data migration issues and questions are challenging, such as how to reconnect the servers to the new controller, and how to reconnect them online without any effect on your applications.

Be aware that with this approach, you not only replace a controller but also implicitly replace your entire virtualization solution. In addition to replacing the hardware, it might be necessary to update or repurchase the licenses for the virtualization feature, advanced copy functions, and so on.

With a SAN or fabric-based appliance solution that is based on a scale-out cluster architecture, lifecycle management tasks, such as adding or replacing new disk subsystems or migrating data between them, are extremely simple. Servers and applications remain online, data migration takes place transparently on the virtualization platform, and licenses for virtualization and copy services require no update; that is, they require no additional costs when disk subsystems are replaced.

Only the fabric-based appliance solution provides an independent and scalable virtualization platform that can provide enterprise-class copy services, is open for future interfaces and protocols, allows you to choose the disk subsystems that best fit your requirements, and does not lock you into specific SAN hardware.

For these reasons, IBM has chosen the SAN or fabric-based appliance approach for the implementation of the IBM System Storage SAN Volume Controller.

The SAN Volume Controller possesses the following key characteristics:

- It is highly scalable, providing an easy growth path to two-*n* nodes (pair of nodes).
- It is SAN interface-independent. It supports FC and FCoE and iSCSI, but it is also open for future enhancements.
- It is host-independent, for fixed block-based Open Systems environments.
- It is external storage RAID controller-independent, providing a continuous and ongoing process to qualify additional types of controllers.
- It can use disks that are internally located within the nodes or in specific expansion units attached to nodes (flash drives).
- It can use disks that are locally attached to the nodes (SAS and flash drives).

On the SAN storage that is provided by the disk subsystems, the SVC can offer the following services:

- It can create and manage a single pool of storage that is attached to the SAN.
- It can manage multiple tiers of storage.
- It provides block-level virtualization (logical-unit virtualization).
- It provides automatic block-level or sub-LUN-level data migration between storage tiers.
- It provides advanced functions to the entire SAN:
  - Large scalable cache
  - Advanced Copy Services:
    - FlashCopy (point-in-time copy)
    - Metro Mirror and Global Mirror (remote copy, synchronous, and asynchronous) These mirror functions can be either FC or IP based.
- It provides nondisruptive and concurrent data migration.

This list of features grows with each release because the layered architecture of the SVC can easily implement new storage features.

1.3.1 SAN Volume Controller conceptual model

A SAN-based storage is managed by the IBM SAN Volume Controller (SVC) in one or more "pairs" of SVC hardware nodes, referred to as a *clustered system or system*. These nodes are attached to the SAN fabric, along with RAID controllers and host systems. The SAN fabric is zoned to allow the SVC to "see" the RAID controllers, and for the hosts to see the SVC. The hosts are not allowed to see or operate on the same physical storage (LUN) from the RAID controller that has been assigned to the SVC. Storage controllers can be shared between the SVC and direct host access if the same LUNs are not shared. The zoning capabilities of the SAN switch must be used to create distinct zones to ensure that this rule is enforced. SAN fabrics can include standard FC, FC over Ethernet, iSCSI over Ethernet, or possible future types.

Figure 1-4 on page 9 shows a conceptual diagram of a storage system using the SAN Volume Controller. It shows a number of hosts that are connected to a SAN fabric or LAN. In practical implementations that have high-availability requirements (the majority of the target clients for SVC), the SAN fabric "cloud" represents a redundant SAN. A *redundant SAN* consists of a fault-tolerant arrangement of two or more counterpart SANs, therefore providing alternate paths for each SAN-attached device.

Both scenarios (using a single network and using two physically separate networks) are supported for iSCSI-based and LAN-based access networks to the SVC. Redundant paths to volumes can be provided in both scenarios.

For simplicity, Figure 1-4 shows only one SAN fabric and two zones, namely host and storage. In a real environment, it is a preferred practice to use two redundant SAN fabrics. The SVC can be connected to up to four fabrics.

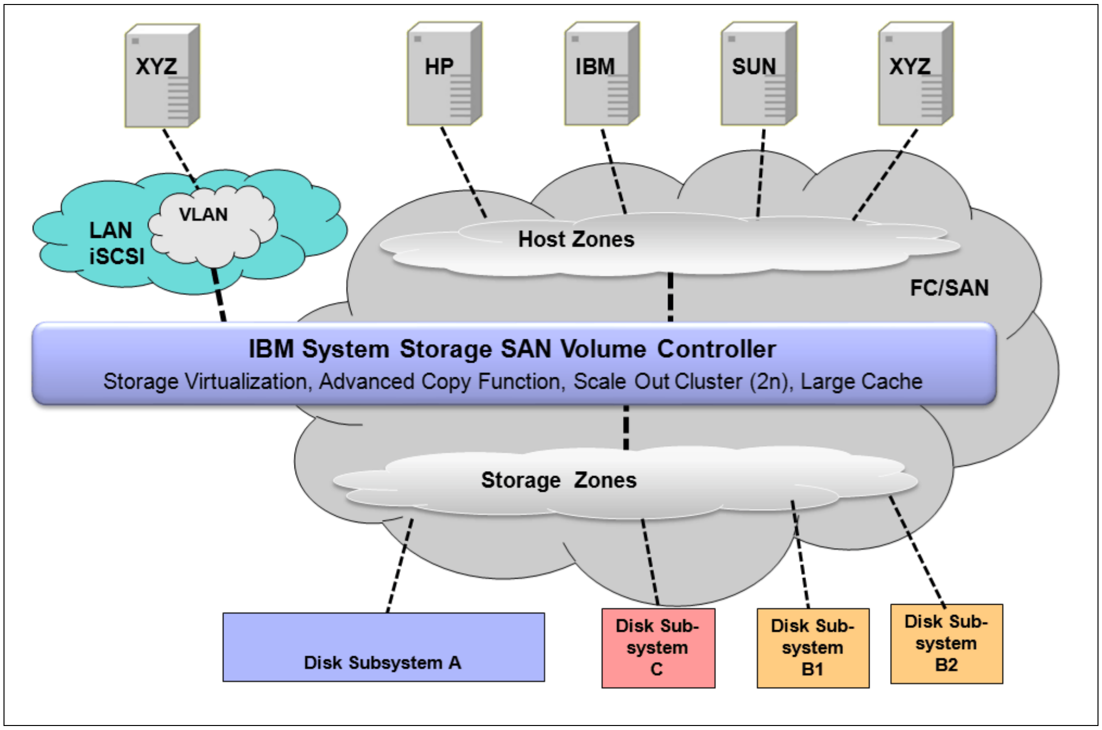


Figure 1-4 SVC conceptual and topology overview

A clustered system of SAN Volume Controller nodes that are connected to the same fabric presents *logical disks* or volumes to the hosts. These volumes are created from managed LUNs or managed disks (MDisks) that are presented by the RAID disk subsystems. There are two distinct zones shown in the fabric:

- A host zone, in which the hosts can see and address the SVC nodes
- A storage zone, in which the SVC nodes can see and address the MDisks/logical unit numbers (LUNs) that are presented by the RAID subsystems

Hosts are not permitted to operate on the RAID LUNs directly, and all data transfer happens through the SVC nodes. This design is commonly described as *symmetric virtualization*. LUNs that are not processed by the SVC can still be provided to the hosts.

For iSCSI-based access, using two networks and separating iSCSI traffic within the networks by using a dedicated virtual local area network (VLAN) path for storage traffic prevents any IP interface, switch, or target port failure from compromising the host servers' access to the volumes LUNs.

1.3.2 SAN Volume Controller components

The SAN Volume Controller (SVC) product provides block-level aggregation and volume management for attached disk storage. In simpler terms, the SVC manages a number of back-end storage controllers or locally attached disks and maps the physical storage within those controllers or disk arrays into logical disk images, or volumes, that can be seen by application servers and workstations in the SAN.

The SAN is zoned so that the application servers cannot see the back-end physical storage, which prevents any possible conflict between the SVC and the application servers both trying to manage the back-end storage. In this section, we briefly explain basic architecture components of SVC.

Nodes

Each SVC hardware unit is called a *node*. The node provides the virtualization for a set of volumes, cache, and copy services functions. SVC nodes are deployed in pairs and multiple pairs make up a *clustered system or system*. A system can consist of between one and four SVC node pairs.

One of the nodes within the system is known as the *configuration node*. The configuration node manages the configuration activity for the system. If this node fails, the system chooses a new node to become the configuration node.

Because the nodes are installed in pairs, each node provides a failover function to its partner node in the event of a node failure.

I/O Groups

Each pair of SVC nodes is also referred to as an *I/O Group*. An SVC clustered system can have from one to four I/O Groups. A specific *volume* is always presented to a host server by a single I/O Group of the system.

When a host server performs I/O to one of its volumes, all the I/Os for a specific volume are directed to one specific I/O Group in the system. Also, under normal conditions, the I/Os for that specific volume are always processed by the same node within the I/O Group. This node is referred to as the *preferred node* for this specific volume.

Both nodes of an I/O Group act as the preferred node for their own specific subset of the total number of volumes that the I/O Group presents to the host servers. A maximum of 2048 volumes per I/O Group is allowed. However, both nodes also act as failover nodes for their respective partner node within the I/O Group. Therefore, a node takes over the I/O workload from its partner node, if required.

Thus, in an SVC-based environment, the I/O handling for a volume can switch between the two nodes of the I/O Group. For this reason, it is mandatory for servers that are connected through FC to use multipath drivers to be able to handle these failover situations.

The SVC I/O Groups are connected to the SAN so that all application servers accessing volumes from this I/O Group have access to this group. Up to 512 hosts can be defined per I/O Group (models DH8, CG8, and CF8). The host server objects can access volumes that are provided by this specific I/O Group.

If required, host servers can be mapped to more than one I/O Group within the SVC system; therefore, they can access volumes from separate I/O Groups. You can move volumes between I/O Groups to redistribute the load between the I/O Groups; however, moving volumes between I/O Groups cannot be done concurrently with host I/O and requires a brief interruption to remap the host.

System

The system or clustered system consists of between one and four I/O Groups. Certain configuration limitations are then set for the individual system. For example, the maximum number of volumes supported per system is 8192 (having a maximum of 2048 volumes per I/O Group), or the maximum managed disk supported is 32 PB per system.

All configuration, monitoring, and service tasks are performed at the system level. Configuration settings are replicated to all nodes in the system. To facilitate these tasks, a management IP address is set for the system.

A process is provided to back up the system configuration data onto disk so that it can be restored in the event of a disaster. Note that this method does not back up application data. Only SVC system configuration information is backed up. For the purposes of remote data mirroring, two or more systems must form a *partnership* before creating relationships between mirrored volumes.

For details about the maximum configurations that are applicable to the system, I/O Group, and nodes, see the following link:

http://www.ibm.com/support/docview.wss?uid=ssg1S1004510

Stretched system

Normally, a pair of nodes from the same I/O Group is physically located within the same rack, in the same computer room. Since IBM SAN Volume Controller V5.1, to provide protection against failures that affect an entire location (for example, a power failure), you can split a single system between two physical locations, up to 10 km (6.2 miles) apart. All inter-node communication between SVC node ports in the same system must not cross inter-switch links (ISLs). Also, all inter-node communication between the SVC and back-end disk controllers must not cross ISLs.

Therefore, the FC path between sites cannot use an inter-switch ISL path. The remote node must have a direct path to the switch to which its partner and other system nodes connect. Starting with SVC 6.3, the distance limit has been extended to Metro Mirror distance (about 300 km or 186.4 miles).

In various information sources, you might see also the term Stretched cluster, Split system, or Split I/O Group. They all refer to the same concept.

MDisks

The IBM SAN Volume Controller system and its I/O Groups view the storage that is presented to the SAN by the back-end controllers as a number of disks or LUNs, known as *managed disks* or *MDisks*. Because the SVC does not attempt to provide recovery from physical disk failures within the back-end controllers, an MDisk is usually provisioned from a RAID array. The application servers, however, do not see the MDisks at all. Instead, they see a number of logical disks, known as *virtual disks* or *volumes*, which are presented by the SVC I/O Groups through the SAN (FC/FCoE) or LAN (iSCSI) to the servers.

The MDisks are placed into storage pools where they are divided into a number of extents, which can range in size 16 - 8182 MB, as defined by the SVC administrator. See the following link for an overview of the total storage capacity that is manageable per system regarding the selection of extents:

http://www.ibm.com/support/docview.wss?uid=ssg1S1004368#_Extents

A volume is host-accessible storage that has been provisioned out of one *storage pool*, or if it is a mirrored volume, out of two storage pools.

The maximum size of an MDisk is 1 PB. An SVC system supports up to 4096 MDisks (including internal RAID arrays). At any point in time, an MDisk is in one of the following three modes:

- Unmanaged MDisk

  An MDisk is reported as unmanaged when it is not a member of any storage pool. An unmanaged MDisk is not associated with any volumes and has no metadata stored on it.

  The SVC does not write to an MDisk that is in unmanaged mode, except when it attempts to change the mode of the MDisk to one of the other modes. The SVC can see the resource, but the resource is not assigned to a storage pool.

- Managed MDisk

  Managed mode MDisks are always members of a storage pool, and they contribute extents to the storage pool. Volumes (if not operated in image mode) are created from these extents. MDisks operating in managed mode might have metadata extents allocated from them and can be used as *quorum disks*. This mode is the most common and normal mode for an MDisk.

- Image mode MDisk

  Image mode provides a direct block-for-block translation from the MDisk to the volume by using virtualization. This mode is provided to satisfy three major usage scenarios:

  – Image mode allows the virtualization of MDisks already containing data that was written directly and not through an SVC; rather, it was created by a direct-connected host. This mode allows a client to insert the SVC into the data path of an existing storage volume or LUN with minimal downtime. The image mode is typically used for data migration from old storage systems to new.

  – Image mode allows a volume that is managed by the SVC to be used with the native copy services function provided by the underlying RAID controller. To avoid the loss of data integrity when the SVC is used in this way, it is important that you disable the SVC cache for the volume.

  – SVC provides the ability to migrate to image mode, which allows the SVC to export volumes and access them directly from a host without the SVC in the path.

  Each MDisk presented from an external disk controller has an online path count that is the number of nodes having access to that MDisk. The *maximum count* is the maximum number of paths detected at any point in time by the system. The *current count* is what the system sees at this point in time. A current value less than the maximum can indicate that SAN fabric paths have been lost.

  Solid-state drives (flash drives) that are located in SVC 2145-24F expansion unit (or directly in disk slots of previous 2145-CG8 nodes) are presented to the cluster as MDisks. To determine whether the selected MDisk is a flash drive, click the link on the MDisk name to display the Viewing MDisk Details panel. If the selected MDisk is a flash drive that is located in 2145-24F expansion unit or in an SVC 2145-CG8 node, the Viewing MDisk Details panel displays values for the Node ID, Node Name, and Node Location attributes. Alternatively, you can select **Work with Managed Disks → Disk Controller Systems** from the portfolio. On the Viewing Disk Controller panel, you can match the MDisk to the disk controller system that has the following values for these attributes.

  Important: With previous SVC models CG8 and older, you can only mirror flash drives between I/O group nodes. With the latest model DH8, you are able to create an array across flash drives located in both 2145-24F expansion units. Each expansion unit must be connected to both nodes.

Quorum disk

A *quorum disk* is a managed disk (MDisk) that contains a reserved area for use exclusively by the system. The system uses quorum disks to break a tie when exactly half the nodes in the system remain after a SAN failure: this situation is referred to as “split brain.” Quorum functionality is not supported on flash drives within SVC nodes. There are three candidate quorum disks. However, only one quorum disk is active at any time.

Disk tier

It is likely that the MDisks (LUNs) presented to the SVC system have various performance attributes due to the type of disk or RAID array on which they reside. The MDisks can be on 15 K disk revolutions per minute (RPMs) Fibre Channel or SAS disk, Nearline SAS, or SATA, or even flash drives.

Therefore, a storage tier attribute is assigned to each MDisk, with the default being generic_hdd. Starting with SVC V6.1, a new tier 0 (zero) level disk attribute is available for flash drives, and it is known as generic_ssd.

Storage pool

A *storage pool* is a collection of up to 128 MDisks that provides the pool of storage from which volumes are provisioned. A single system can manage up to 128 storage pools. The size of these pools can be changed (expanded or shrunk) at run time by adding or removing MDisks, without taking the storage pool or the volumes offline.

At any point in time, an MDisk can only be a member in one storage pool, except for image mode volumes.

Each MDisk in the storage pool is divided into a number of extents. The size of the extent is selected by the administrator at the creation time of the storage pool and cannot be changed later. The size of the extent ranges 16 - 8192 MB.

It is a preferred practice to use the same extent size for all storage pools in a system. This approach is a prerequisite for supporting volume migration between two storage pools. If the storage pool extent sizes are not the same, you must use volume mirroring.

Figure 1-5 on page 14 illustrates the relationships of the SAN Volume Controller entities to each other.

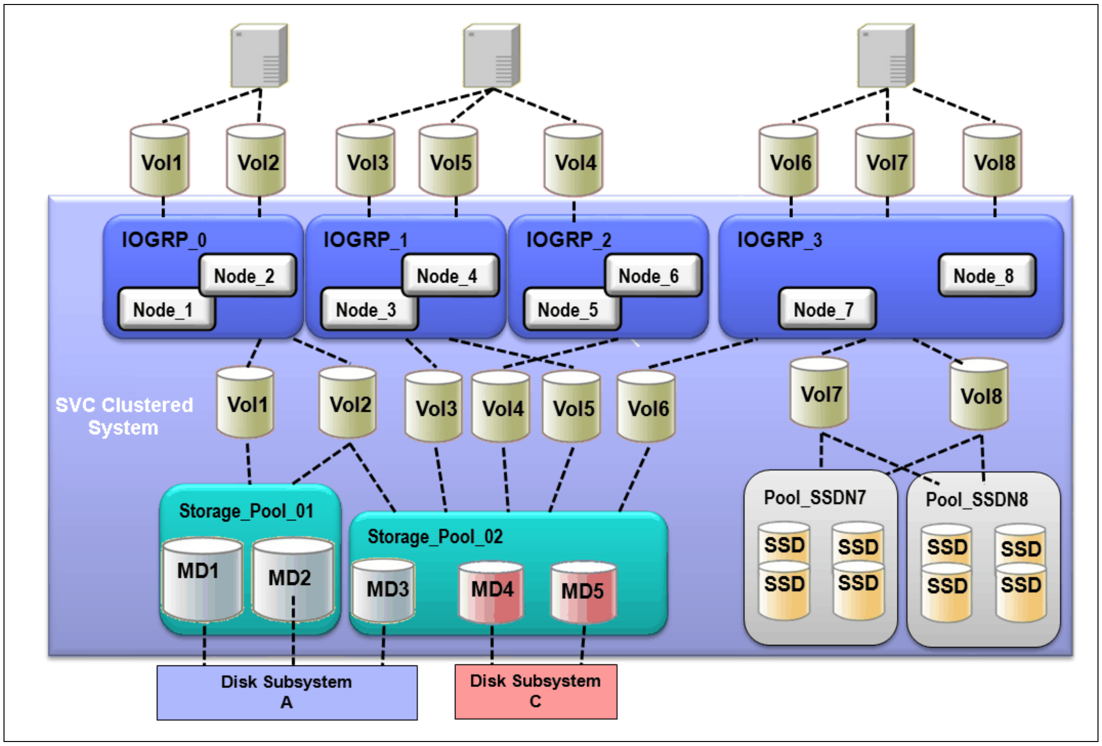


Figure 1-5 Overview of SAN Volume Controller clustered system with I/O Group

The SAN Volume Controller limits the number of extents in a system to 2^{22} = ~4 million. Because the number of addressable extents is limited, the total capacity of a SAN Volume Controller system depends on the extent size that is chosen by the SAN Volume Controller administrator. The capacity numbers that are specified in Table 1-1 for a SAN Volume Controller system assume that all defined storage pools have been created with the same extent size.

Table 1-1 Extent size-to-addressability matrix

Extent size maximum	System capacity	Extent size maximum	System capacity
16 MB	64 TB	512 MB	2 PB
32 MB	128 TB	1024 MB	4 PB
64 MB	256 TB	2048 MB	8 PB
128 MB	512 TB	4096 MB	16 PB
256 MB	1 PB	8192 MB	32 PB

For most systems, a capacity of 1 - 2 PB is sufficient. A preferred practice is to use 256 MB or for larger clustered systems 512 MB as the standard extent size.

Volumes

Volumes are logical disks that are presented to the host or application servers by the SAN Volume Controller. The hosts cannot see the MDisks; they can only see the logical volumes created from combining extents from a storage pool.

There are three types of volumes: striped, sequential, and image. These types are determined by how the extents are allocated from the storage pool, as explained here:

- A volume created in striped mode has extents allocated from each MDisk in the storage pool in a round-robin fashion.
- With a sequential mode volume, extents are allocated sequentially from an MDisk.
- Image mode is a one-to-one mapped extent mode volume.

Using striped mode is the best method to use for most cases. However, sequential extent allocation mode can slightly increase the sequential performance for certain workloads.

Figure 1-6 shows the striped volume mode and sequential volume mode, and it illustrates how the extent allocation from the storage pool differs.

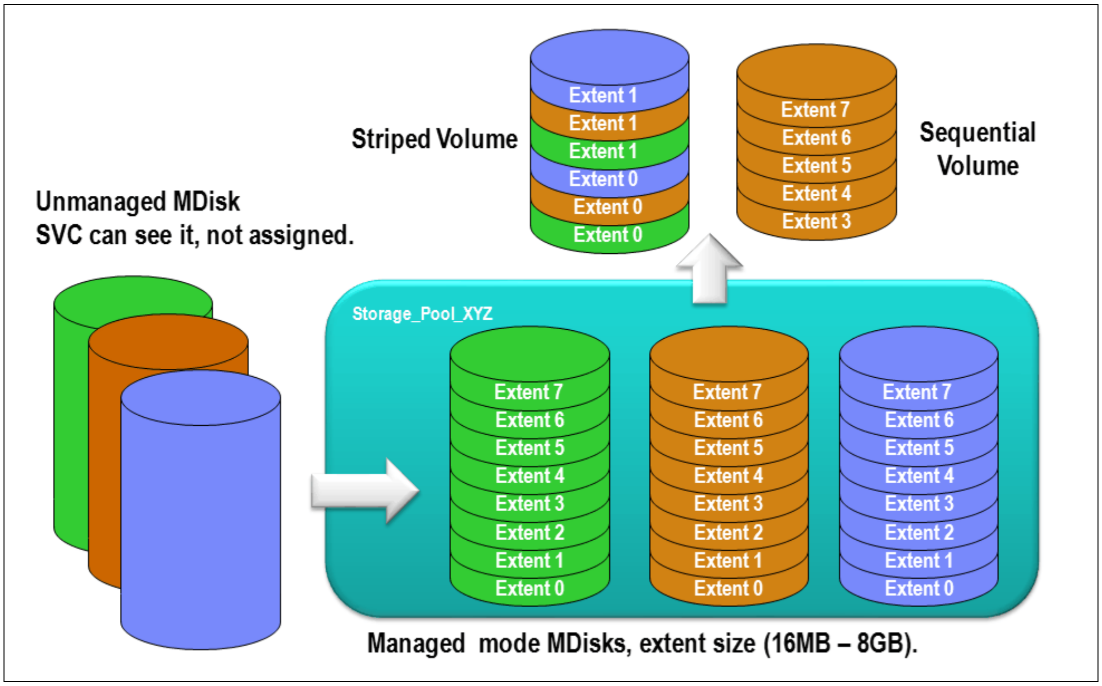


Figure 1-6 Storage pool extents overview

You can allocate the extents for a volume in many ways. The process is under full user control at volume creation time and can be changed at any time by migrating single extents of a volume to another MDisk within the storage pool.

Hosts

Volumes can be mapped to a *host* to allow access for a specific server to a set of volumes. A host within the SAN Volume Controller is a collection of host bus adapter (HBA) worldwide port names (WWPNs) or iSCSI qualified names (IQNs), defined on the specific server. Note that iSCSI names are internally identified by “fake” WWPNs, or WWPNs that are generated by the SAN Volume Controller. Volumes can be mapped to multiple hosts, for example, a volume that is accessed by multiple hosts of a server system.

iSCSI is an alternative means of attaching hosts. However, all communication with back-end storage subsystems, and with other SAN Volume Controller systems, is still through FC.

Node failover can be handled without having a multipath driver installed on the iSCSI server. An iSCSI-attached server can simply reconnect after a node failover to the original target IP address, which is now presented by the partner node. To protect the server against link failures in the network or HBA failures, using a multipath driver is mandatory.

Volumes are LUN-masked to the host's HBA WWPNs by a process called *host mapping*. Mapping a volume to the host makes it accessible to the WWPNs or iSCSI names (IQNs) that are configured on the host object.

For a SCSI over Ethernet connection, the IQN identifies the iSCSI target (destination) adapter. Host objects can have both IQNs and WWPNs.

Easy Tier

IBM Easy Tier is a performance function that automatically migrates or moves extents off a volume to, or from, one MDisk storage tier to another MDisk storage tier. Starting with the SVC V7.3, the Easy Tier automatically moves extents between highly and less utilized MDisks within the same storage tier. This function is called *Storage Pool Balancing* and it is enabled by default without any need for licensing. It cannot be disabled by user. Easy Tier monitors the host I/O activity and latency on the extents of all volumes with the Easy Tier function turned on in a multitier storage pool over a 24-hour period.

New in SVC 7.3: Easy Tier V3 integrates the automatic functionality to balance the workloads between highly and less utilized MDisks within the same tier. It is enabled by default, cannot be disabled by user, and does not need an Easy Tier license.

Next, it creates an extent migration plan based on this activity and then dynamically moves high activity or hot extents to a higher disk tier within the storage pool. It also moves extents whose activity has dropped off or cooled from the high-tier MDisks back to a lower-tiered MDisk.

Easy Tier: The Easy Tier function can be turned on or off at the storage pool level and volume level. It supports any combination of three tiers within the system. Flash drives are always marked as Tier 0. Turning off Easy Tier does not disable Storage Pool Balancing.

To experience the potential benefits of using Easy Tier in your environment before actually installing expensive flash drives, you can turn on the Easy Tier function for a *single-level* storage pool. Next, turn on the Easy Tier function for the volumes within that pool. Easy Tier then starts monitoring activity on the volume extents in the pool.

Easy Tier creates a report every 24 hours, providing information about how Easy Tier behaves if the tier were a multitiered storage pool. So, even though Easy Tier extent migration is not possible within a single-tiered pool, the Easy Tier statistical measurement function is available.

The Easy Tier function can make it more appropriate to use smaller storage pool extent sizes. The usage statistics file can be offloaded from the SAN Volume Controller nodes. Then, you can use an IBM Storage Advisor Tool to create a summary report.

Thin provisioning

Volumes can be configured to be either "*thin-provisioned*" or "*fully allocated*". A thin-provisioned volume behaves concerning application reads and writes as though they were fully allocated. When creating a thin-provisioned volume, the user specifies two capacities: the real physical capacity allocated to the volume from the storage pool, and its virtual capacity available to the host. In a fully allocated volume, these two values are the same.

Thus, the real capacity determines the quantity of MDisk extents that is initially allocated to the volume. The virtual capacity is the capacity of the volume reported to all other SAN Volume Controller components (for example, FlashCopy, Cache, and remote copy) and to the host servers.

The real capacity is used to store both the user data and the metadata for the thin-provisioned volume. The real capacity can be specified as an absolute value or a percentage of the virtual capacity.

Thin-provisioned volumes can be used as volumes assigned to the host, by FlashCopy to implement thin-provisioned FlashCopy targets, and also with the mirrored volumes feature.

When a thin-provisioned volume is initially created, a small amount of the real capacity is used for initial metadata. Write I/Os to grains of the thin volume that were not previously written to cause grains of the real capacity to be used to store metadata and the actual user data. Write I/Os to grains that were previously written to update the grain where data was previously written. The grain size is defined when the volume is created and can be 32 KB, 64 KB, 128 KB, or 256 KB. The default grain size is 256 KB, and is the strongly recommended option. If you select 32 KB for the grain size, the volume size cannot exceed 260,000 GB. The grain size cannot be changed after the thin-provisioned volume has been created. Generally, smaller grain sizes save space but require more metadata access, which can adversely impact performance. If you are not going to use the thin-provisioned volume as a FlashCopy source or target volume, use 256 KB to maximize performance. If you are going to use the thin-provisioned volume as a FlashCopy source or target volume, specify the same grain size for the volume and for the FlashCopy function.

Figure 1-7 illustrates the thin-provisioning concept.

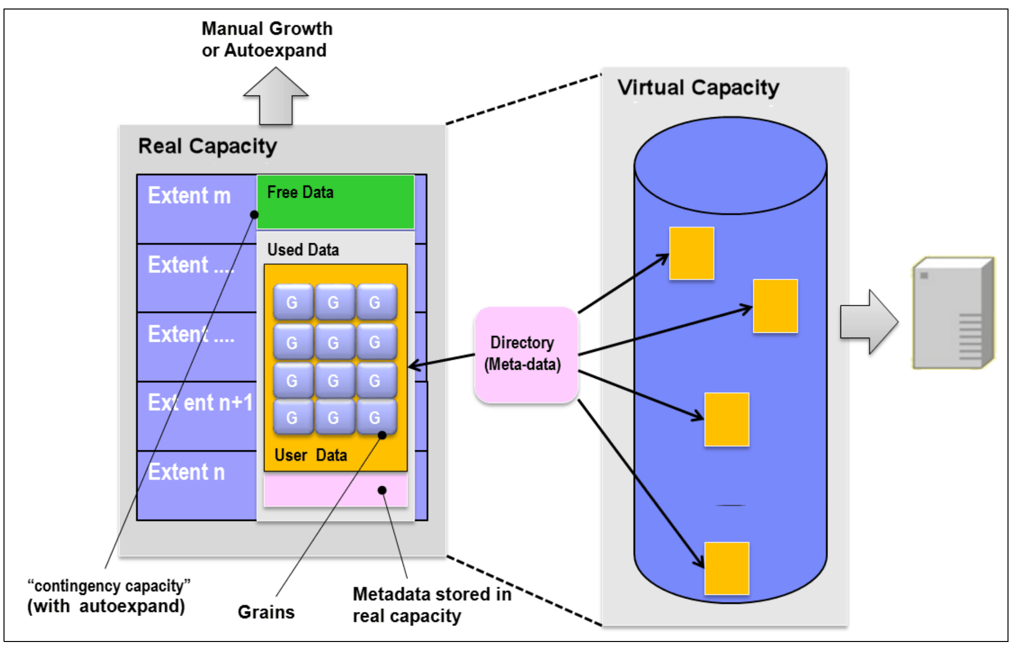


Figure 1-7 Conceptual diagram of thin-provisioned volume

Thin-provisioned volumes store both user data and metadata. Each grain of data requires metadata to be stored. Therefore, the I/O rates that are obtained from thin-provisioned volumes are less than the I/O rates that are obtained from fully allocated volumes.

The metadata storage overhead is never greater than 0.1% of the user data. The overhead is independent of the virtual capacity of the volume. If you are using thin-provisioned volumes in a FlashCopy map, for the best performance, use the same grain size as the map grain size. If you are using the thin-provisioned volume directly with a host system, use a small grain size.

Thin-provisioned volume format: Thin-provisioned volumes do not need formatting. A read I/O, which requests data from deallocated data space, returns zeros. When a write I/O causes space to be allocated, the grain is zeroed before use. However, if the node is a model CF8 or CG8, the space is not allocated for a host write that contains all zeros. The formatting flag is ignored when a thin volume is created or the real capacity is expanded; the virtualization component never formats the real capacity of a thin-provisioned volume.

The real capacity of a thin volume can be changed if the volume is not in image mode. Increasing the real capacity allows a larger amount of data and metadata to be stored on the volume. Thin-provisioned volumes use the real capacity that is provided in ascending order as new data is written to the volume. If the user initially assigns too much real capacity to the volume, the real capacity can be reduced to free storage for other uses.

A thin-provisioned volume can be configured to *autoexpand*. This feature causes the SAN Volume Controller to automatically add a fixed amount of additional real capacity to the thin volume as required. Autoexpand therefore attempts to maintain a fixed amount of unused real capacity for the volume. This amount is known as the *contingency capacity*.

The contingency capacity is initially set to the real capacity that is assigned when the volume is created. If the user modifies the real capacity, the contingency capacity is reset to be the difference between the used capacity and real capacity.

A volume that is created without the autoexpand feature, and thus has a zero contingency capacity, will go offline as soon as the real capacity is used and needs to expand.

Autoexpand will not cause the real capacity to grow much beyond the virtual capacity. The real capacity can be manually expanded to more than the maximum that is required by the current virtual capacity, and the contingency capacity will be recalculated.

To support the auto expansion of thin-provisioned volumes, the storage pools from which they are allocated have a configurable capacity warning. When the used capacity of the pool exceeds the warning capacity, a warning event is logged. For example, if a warning of 80% has been specified, the event will be logged when 20% of the free capacity remains.

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

The fully allocated to thin-provisioned migration procedure uses a zero-detection algorithm so that grains containing all zeros do not cause any real capacity to be used.

Real-time Compression

Compressed volumes are a special type of volume where data is compressed as it is written to disk, saving additional space. To use the compression function, you must obtain the IBM Real-time Compression license, and the hardware level for both nodes within the I/O group must be either IBM SAN Volume Controller 2145-DH8, CG8, or CF8. The SVC model DH8 nodes must have two processors, 64 GB memory, and at least one compression acceleration adapter installed in order to use compression. Enabling compression on SAN Volume Controller 2145-DH8 nodes does not affect non-compressed host to disk I/O performance. Like thin-provisioned volumes, compressed volumes have virtual, real, and used capacities. Use the following guidelines before working with compressed volumes:

- Real capacity is the extent space that is allocated from the storage pool. The real capacity is also set when the volume is created, and like thin-provisioned volumes, can be expanded or shrunk down to the used capacity.
- Virtual capacity is available to hosts. The virtual capacity is set when the volume is created and can be expanded or shrunk afterward.
- Used capacity is the amount of real capacity used to store client data and metadata after compression.
- Capacity before compression is the amount of client data that has been written to the volume and then compressed. The capacity before compression does not include regions where zero data is written to deallocated space.
- An I/O group can contain a maximum of 200 compressed volumes and compressed volume mirrors. By installation of the secondary compression accelerator card into each DH8 node, the I/O group supports up to 512 compressed volumes.

You can also monitor information about compression usage to determine the savings to your storage capacity when volumes are compressed. To monitor system-wide compression savings and capacity, select **Monitoring** → **System** and either select the system name or Compression View. You can compare the amount of capacity used before compression is applied to the capacity that is used for all compressed volumes. In addition, you can view the total percentage of capacity savings when compression is used on the system. In addition, you can also monitor compression savings across individual pools and volumes. For volumes, you can use these compression values to determine which volumes have achieved the highest compression savings.

More details about SVC Real-time Compression are available in Chapter 7, “Real-time Compression and the SVC 2145-DH8 Node” on page 139.

Cache

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

The new 2145-DH8 nodes combined with SAN Volume Controller V7.3 provide 32 GB (and an additional 32 GB with the second processor card, which offers more processor power and memory for the Real-time Compression feature) memory per node, or 64 GB (128 GB) per I/O Group, or 256 GB (512 GB) per SAN Volume Controller system. The SVC provides a semi-flexible cache model, and the node’s memory can be used as read or write cache, either as an I/O workload cache (28 GB without RtC or 26 GB with RtC enabled) or as compression cache (36 GB). The size of the write cache is maximally 12 GB of the node’s memory. The remaining part of the memory is split between read cache allocation and compression allocation. The cache allocation mechanism is depicted in Table 1-2.

Table 1-2 Memory allocation for cache

Cache allocation per node	Write cache	Read cache	Compression cache	Not used
1 CPU, 32 GB, RtC not available	12 GB	16 GB	N/A	0 GB
2 CPU, 64 GB, RtC disabled	12 GB	16 GB	N/A	36 GB
2 CPU, 64 GB, RtC enabled	12 GB	16 GB	36 GB	0 GB

When data is written by the host, the preferred node saves the data in its cache. Before the cache returns completion to the host, the write must be mirrored to the partner node, or copied into the cache of its partner node, for availability reasons. After having a copy of the written data, the cache returns completion to the host. A volume that has not received a write update during the last 2 minutes will automatically have all modified data destaged to disk.

Starting with SAN Volume Controller V7.3, the concept of the cache architecture has been changed. The SVC now distinguishes between Upper and Lower cache that allows system to be more scalable:

- Required for support beyond 8192 volumes
- Required for support beyond eight node clusters
- Required for 64-bit addressing beyond 28 GB
- Required for larger memory in nodes
- Required for more processor cores
- Required for improved performance and stability

The architectural overview is presented in Figure 1-8.

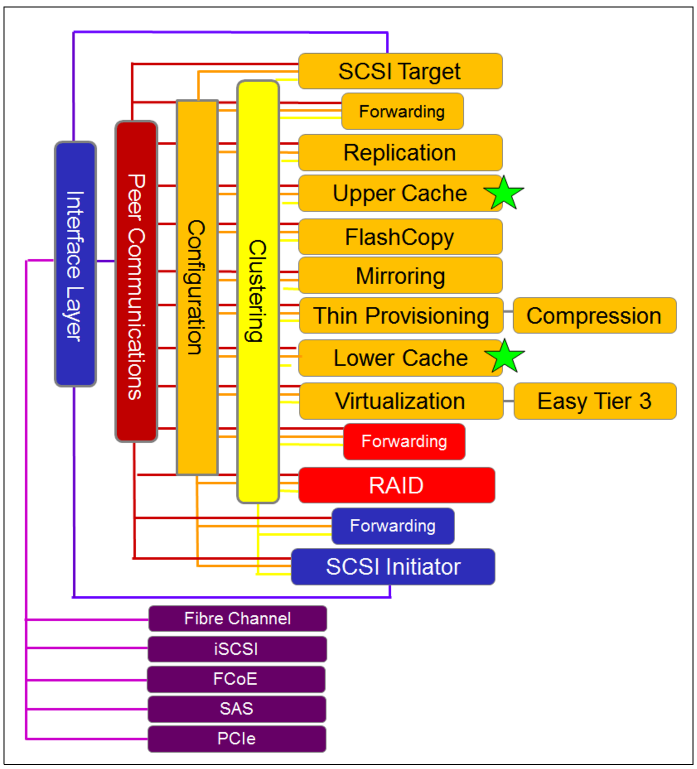


Figure 1-8 New cache architecture

If one node of an I/O Group is missing, due to a restart or a hardware failure, the remaining node empties all of its write cache and proceeds in operation mode, which is referred to as *write-through mode*. A node operating in write-through mode writes data directly to the disk subsystem before sending an `I/O complete` status message back to the host. Running in this mode can degrade the performance of the specific I/O Group.

Write cache is partitioned by storage pool. This feature restricts the maximum amount of write cache that a single storage pool can allocate in a system. Table 1-3 on page 21 shows the upper limit of write-cache data that a single storage pool in a system can occupy.

Table 1-3 Upper limit of write cache per storage pool

One storage pool	Two storage pools	Three storage pools	Four storage pools	More than four storage pools
100%	66%	40%	33%	25%

A SAN Volume Controller node will treat part of its physical memory as non-volatile. *Non-volatile* means that its contents are preserved across power losses and resets. Bitmaps for FlashCopy and Remote Mirroring relationships, the virtualization table, and the write cache are items in the non-volatile memory.

In the event of a disruption or external power loss, the physical memory is copied to a file in the file system on the node's internal disk drive, so that the contents can be recovered when external power is restored. The functionality of uninterruptible power supply units is provided by internal batteries, which are delivered with each node's hardware. They ensure that there is sufficient internal power to keep a node operational to perform this dump when the external power is removed. After dumping the content of the non-volatile part of the memory to disk, the SAN Volume Controller node shuts down.

1.4 More information

This publication covers key aspects of the implementation of the IBM SAN Volume Controller (SVC) focused on version 7.3 and its associated hardware upgrade (2145-DH8). For a more detailed description about the concept of storage virtualization and how IBM leverages SVC in data centers, study the following materials:

- *Implementing the IBM System Storage SAN Volume Controller V7.2*, SG24-7933
- *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521
- *IBM System Storage SAN Volume Controller and Storwize V7000 Replication Family Services*, SG24-7574
- *IBM SAN Volume Controller Enhanced Stretched Cluster with VMware*, SG24-8211
- *Implementing IBM Real-time Compression in SAN Volume Controller and IBM Storwize V7000*, TIPS1083
- *Implementing FlashSystem 840 with SAN Volume Controller*, TIPS1137
- *Implementing IBM Easy Tier with IBM Real-time Compression*, TIPS1072

2

Change history

This chapter summarizes the enhancements in the IBM System Storage SAN Volume Controller (SVC) since V4.3. It also explains the terminology that changed over previous releases of SAN Volume Controller.

It includes the following sections in a reverse-chronological order:

- Enhancements and changes in SAN Volume Controller V7.3
- Enhancements and changes in SAN Volume Controller V7.2
- Enhancements and changes in SAN Volume Controller V7.1
- Enhancements and changes in SAN Volume Controller V6.4
- Enhancements and changes in SAN Volume Controller V6.3
- Enhancements and changes in SAN Volume Controller V6.2
- Enhancements and changes in SAN Volume Controller V6.1
- Enhancements and changes in SAN Volume Controller V5.1

2.1 Enhancements and changes in SAN Volume Controller V7.3

The IBM SAN Volume Controller V7.3 and its related HW upgrade is another important milestone in the product line development. The product internal architecture is significantly rebuilt, enabling the system to break the previous limitations in terms of scalability and flexibility. It includes:

- Significant upgrade of the 2145-DH8 hardware

  The new 2145-DH8 introduces 2U server based on IBM x3650 M4 series and integrates the following features:

  – Minimum 8-core processors with 32 GB memory for SVC. Optional secondary processor with additional 32 GB memory when a third I/O card is needed. Secondary processor is compulsory when Real-time Compression is enabled.

  – Integrated dual-battery pack as an uninterruptible power supply in case of power outage. External UPS device is no longer needed, avoiding mis-cabling issues.

  – Dual, redundant power supplies, therefore no external power switch required.

  – Removed front panel. The majority of its actions were moved to the functionality of rear Technician Ethernet port with enabled DHCP for instant access.

  – Two boot drives with data mirrored across drives node will still boot in the event of a drive failure. Dump data is striped for performance reasons.

  – Enhanced scalability with up to three PCI Express slot capabilities. This allows users to install up to three four-port 8 Gbps FC HBA (12 ports). It supports one four-port 10GbE card (iSCSI or FCoE) and 1 dual-port 12 Gbps SAS card for flash drives expansion unit attachment (model 2145-24F).

  – Improved Real-time Compression engine (RACE) with the processing offloaded to the secondary dedicated processor and using 36 GB of dedicated memory cache. At minimum one Compression Accelerator card needs to be installed (up to 200 compressed volumes) or two Compression Accelerators allow up to 512 compressed volumes.

  – Optional 2U expansion enclosure 2145-24F with up to 24 flash drives (200, 400, or 800 GB) for tier 0 when the Easy Tier function is enabled. The RAID array can be built across both expansions attached to each node in the I/O group. The 12 Gbps SAS four-port adapter needs to be installed in each SVC node.

  – Extends the functionality of Easy Tier by Storage Pool Balancing mode within the same tier. It moves or exchanges extents between highly and low utilized MDisks within a storage pool, thus increasing the read and write performance of the volumes. This function is enabled automatically in SVC and does not need any license. It cannot be disabled by the administrator.

  – The SVC cache rearchitecture splits the original single cache into the Upper and Lower cache of different size. Upper cache uses up to 256 MB while lower cache uses up to 64 GB of installed memory allocated to both processors (if installed). 36 GB of memory is always allocated for Real-time Compression if enabled.

  – Near instant prepare for FlashCopy due to presence of lower cache. Multiple snapshots of golden image now share cache data (instead of number of N copies).

More information about new hardware and software enhancements is summarized in Chapter 3, “2145-DH8 and SAN Volume Controller software version 7.3” on page 37, and additional technical details about each of its functions are summarized in the remaining chapters of this publication.

2.2 Enhancements and changes in SAN Volume Controller V7.2

SAN Volume Controller V7.2 has the following enhancements and changes:

- Remote Mirroring over IP communication links

  The Remote Mirroring function (also referred to as *Metro/Global Mirror*) is now supported using Ethernet communication links. IBM Storwize Family Software IP replication uses innovative Bridgeworks SANSlide technology to optimize network bandwidth and utilization. This new function enables the use of lower speed and lower-cost networking infrastructure for data replication. Bridgeworks' SANSlide technology integrated into IBM Storwize Family Software uses artificial intelligence to help optimize network bandwidth utilization and adapt to changing workload and network conditions. This technology can improve remote mirroring network bandwidth utilization up to three times, which may enable clients to deploy a less costly network infrastructure or speed remote replication cycles to enhance disaster recovery effectiveness.

- Enhanced Stretched Cluster for SAN Volume Controller

  Before this release, stretched cluster configurations did not provide manual failover capability, and data being sent across a long-distance link had the potential to be sent twice. The addition of “site awareness” in Storwize Family Software V7.2 routes I/O traffic between SVC nodes and storage controllers to optimize the data flow, and it polices I/O traffic during a failure condition to allow for a manual cluster invocation to ensure consistency. The use of Stretched Cluster continues to follow all the same hardware installation guidelines as previously announced and found in the product documentation. Use of Enhanced Stretched Cluster is optional, and existing Stretched Cluster configurations will continue to be supported.

- Performance improvements for asynchronous remote mirroring

  Enhancements to the asynchronous remote mirroring function enable improved throughput of remotely replicated data.

- Improved efficiency in drive firmware update process

  In prior software releases, updating internal drive firmware was a serial process that required the user to update each drive individually. V7.2 introduces a drive firmware update command, **`svctask applydrivesoftware`**, that allows for the updating of multiple drives with a single command-line interface (CLI) command.

- vSphere API for Storage Awareness (VASA)

  Storwize Family Software V7.2 enables users to get more capability out of their VMware environments by being a provider for the vSphere API for Storage Awareness.

- Data Migration using SAS connectivity on Storwize V3500, V3700, and V5000

  Data migration support is standard on all Storwize V3500, V3700, and V5000 systems and this function can now be performed using SAS connectivity to help you easily and non-disruptively migrate data from IBM System Storage DS3200 and DS3500 systems onto Storwize V3500, V3700, and V5000.

- Enhanced monitoring capabilities with the new IBM Storage Mobile Dashboard

  With V7.2, a new mobile application was released that allows for monitoring and health check functionality for SVC and Storwize Family storage systems. The application is available for free from the Apple App Store.

- Improved performance and efficiency for Real-time Compression with the introduction of the Random Access Compression Engine (RACE) 2.2

  V7.2 introduces several improvements to the Real-time Compression functionality that allow for:

  - Up to 3x higher sequential write throughput, allowing for faster VMware vMotion operations and sequential copy operations, as well as more VMware vMotion sessions in parallel
  - 35% higher throughput (IOPS) in intensive DB OLTP workloads
  - 35% lower compression processor usage for the same workload compared to V7.1
- Interoperability

  For an updated list, see “V7.2 Supported Hardware List, Device Driver, Firmware and Recommended Software Levels for SAN Volume Controller” at:

  https://www.ibm.com/support/docview.wss?uid=ssg1S1004453

2.3 Enhancements and changes in SAN Volume Controller V7.1

SAN Volume Controller V7.1 has the following enhancements and changes:

- Increased Number of Host Objects

  SAN Volume Controller V7.1increases the number of host objects per I/O group from 256 to 512 and the per cluster limit from 1024 to 2048. The increased host objects can be used for any host type subject to limit restrictions for that host type, for example, iSCSI names/IQNs (iSCSI Qualified Names).
- Increased Number of Host worldwide port names (WWPNs)

  SAN Volume Controller V7.1 increases the officially supported number of host WWPNs per I/O group to 2048 and per cluster to 8192. This specifically benefits IBM AIX® LPM configurations and environments using NPIV to map volumes to virtual WWPNs. This increase applies to native FC and FCoE WWPNs.
- Increased Number of Volumes per Host

  SAN Volume Controller V7.1 increases volumes per host from 512 to 2048, and the increase in volumes per host is available to any host operating system, subject to that host’s OS limits. Increase in volumes per host is applicable to FC and FCoE host attachment types, and does not apply to iSCSI attached servers/hosts.
- Support for the direct attachment of AIX hosts

  SAN Volume Controller V7.1.X introduces support of direct AIX host attachment.
- Support for additional drive types

  SAN Volume Controller V7.1 introduces support for:

  - 4 TB NL_SAS 7.2 K RPM 3.5” LFF drives (supported on Storwize V7000 2076-x12 LFF enclosures only; Supported with Flex System V7000 when using external expansion enclosure model 2076-212 connected to the Flex System V7000 control enclosure; Not supported on Storwize V3700/3500 models)
  - 1.2 TB SAS 10 K RPM 2.5” SFF drive supported on Storwize V7000, Storwize V3700/3500, and Flex System V7000 SFF enclosures. For Storwize V3700/3500 LFF control or expansion enclosures this 2.5” drive is available on an LFF carrier as is the 2.5” SAS 15 K RPM 300 GB drive and the 2.5” SAS 10 K RPM 900 GB drive

- Second Fibre Channel HBA support

  SAN Volume Controller V7.1 adds support for the additional 4-port 8 Gbps Fibre Channel HBA that is available in feature code AHA7 on 2145-CG8 hardware.

- Port masking

  The addition of more Fibre Channel HBA ports that are introduced with feature code AHA7 allow clients to optimize their SVC configuration by using dedicated ports for certain system functions. However, the addition of these ports necessitates the ability to ensure traffic isolation. As such, SAN Volume Controller V7.1 introduces port masking.

  The following are two examples of traffic types that you might want to isolate using port masking:

  – Local node-to-node communication
  – Replication traffic

- Support for Easy Tier with compressed volumes

  Easy Tier is a performance optimization function that automatically migrates “hot” extents belonging to a volume to MDisks that better meet the performance requirements of that extent. The Easy Tier function can be turned on or off at the storage pool level and at the volume.

  Real-time Compression is a feature of SVC that addresses all the requirements of primary storage data reduction, including performance, using a purpose-built compression technology, allowing for data reduction of up to 80%.

  In practice, clients have found that their target workloads for these two features have a significant overlap. Before SVC Storage Software version 7.1, the use of these two features was mutually exclusive at the volume level. SAN Volume Controller V7.1 introduces support for the concurrent use of Easy Tier and Real-time Compression on the same volume.

- Enhanced flexibility in modifying Remote Copy relationships

  SAN Volume Controller V7.1introduces the ability to change between Metro Mirror and Global Mirror (with or without change volumes) without requiring a full resync of all data from the primary volume to the secondary volume.

- Storwize V3700 support for Remote Copy

  SAN Volume Controller V7.1introduces support for Remote Copy on Storwize V3700 systems, allowing for remote replication between any combination of the following systems:

  – SVC
  – Storwize V7000
  – Flex System V7000
  – Storwize V3700
  – Storwize V3500

- Interoperability

  For an updated list, see “V7.1 Supported Hardware List, Device Driver, Firmware and Recommended Software Levels for SAN Volume Controller” at the following link:

  https://www.ibm.com/support/docview.wss?uid=ssg1S1004392

2.4 Enhancements and changes in SAN Volume Controller V6.4

SAN Volume Controller V6.4 has the following enhancements and changes:

- FCoE Support

  With V6.4, SAN Volume Controller systems with 10 Gbps Ethernet ports now support attachment to next-generation CEE networks using FCoE. This support enables SVC connections to servers for host attachment and to other SAN Volume Controller systems for clustering or for mirroring using Fibre Channel or FCoE interfaces using these networks. The same ports may also be used for iSCSI server connections.

- Non-Disruptive Volume Movement across clustered systems

  V6.4 enhances data mobility, with greater flexibility for nondisruptive volume moves. Previous versions of SVC provided the ability to move volumes nondisruptively between the nodes in an I/O group. Version 6.4 supports moving volumes anywhere within a clustered system without disruption of host access to storage.

- Real-time Compression

  V6.4 is designed to improve storage efficiency by supporting real-time compression for block storage, which is designed to improve efficiency by compressing data by as much as 80%, enabling storage for up to five times as much data in the same physical disk space. Unlike other approaches to compression, IBM Real-time Compression is designed to be used with active primary data such as production databases and email systems, dramatically expanding the range of candidate data that can benefit from compression. IBM Real-time Compression operates as data is written to disk avoiding the need to store uncompressed data while awaiting compression.

- Storwize V7000 clustering

  V6.4 allows for the clustering of multiple Storwise V7000 control enclosures. Like SAN Volume Controller, Storwize V7000 clustering works on the notion of I/O groups. With Storwise V7000, an I/O group is a control enclosure and its associated expansion enclosures. A Storwise V7000 clustered system can consist of 2 - 4 I/O groups.

- Support for direct host attachment

  The Storwize V7000 now supports direct host attachment, while it is not a supported option with SVC.

- Updated thin provisioned volume grain size

  For improved performance and interaction with Easy Tier, the default grain size of a thin provisioned volume has been changed to 256 KB from the previous default of 32 KB.

- Extended support for SCSI-3 persistent reservations

  Additional persistent reserve functions will allow GPFS to use persistent reserves on a Storwize V7000 or SVC system.

- Interoperability

  For an updated list, see “V6.4 Supported Hardware List, Device Driver, Firmware and Recommended Software Levels for SAN Volume Controller” at the following link:

  https://www.ibm.com/support/docview.wss?uid=ssg1S1004111

2.5 Enhancements and changes in SAN Volume Controller V6.3

SAN Volume Controller V6.3 has the following enhancements and changes:

- Enhanced Replication via Global Mirror with Change Volumes (GMCV)

  Enhancements to Global Mirror with the SAN Volume Controller V6.3.0 are designed to provide new options to help administrators balance network bandwidth requirements and recovery point objectives (RPOs) for applications. SVC now supports higher RPO times, providing the option to use a lower-bandwidth link between mirrored sites. This lower-bandwidth remote mirroring uses space-efficient, FlashCopy targets as sources in remote copy relationships to increase the time that is allowed to complete a remote copy data cycle.

- Metro Mirror and Global Mirror Replication between SVC and Storwize V7000 systems

  With both SVC and Storwize V7000 running V6.3.x, the Storwize V7000 can act as an SVC Metro Mirror or Global Mirror partner system. SAN Volume Controller V6.3 introduces a new cluster property called “layer”. Storwize V7000 is either in “replication layer” mode or “storage layer” mode, while SVC is always in “replication layer” mode. Storwize V7000 is in “storage layer” mode by default, and can be switched to “replication layer” using the **`svctask chcluster -layer replication`** command. Once changed to “replication layer” mode, the Storwize V7000 can then be used to create a remote copy relationship with an SVC cluster.

- Automatically shrink thin provisioned volumes

  For thin provisioned volumes, the real capacity starts small and grows as data is written to the volume. In previous versions, if data was deleted real capacity did not automatically shrink, but could be manually shrunk. This might cause problems with FlashCopy (FC) mappings. In SAN Volume Controller V6.3, starting a FlashCopy mapping automatically shrinks used and real capacity to zero for all thin provisioned volumes used in the FC mapping.

- Mirrored Volume Time-out Enhancements

  SAN Volume Controller V6.3 introduces the option of configurable timeout settings for each mirrored volume. The default setting for the “mirror_write_priority” property of the volume is “latency”, and uses a short timeout, prioritizing low host latency. This property can be changed to “redundancy, where a longer timeout is used to prioritize redundancy.

- Support for Round-Robin multipathing to external storage systems

  Before V6.3, all I/O to an external MDisk on a storage system was via a single port on that controller. The selected port changed if that port became unavailable. In V6.3 an I/O is submitted using one path per target port per MDisk per node. Paths are chosen according to port groups presented by the storage system, and the I/O is sent in parallel to all target ports. This enables I/O to an MDisk to progress in a “round robin” fashion, with the following potential benefits:

  – Performance improvement
  – Spreading the I/O across multiple storage system ports
  – Balancing the number of preferred paths per node port both within each port group and across the system as a whole
  – Improved resilience to certain storage system failures
  – Faster detection of path failures

For information about the specific external storage systems that are supported by Round-Robin multipathing in V6.3, see "V6.3 Supported Hardware List, Device Driver, Firmware and Recommended Software Levels for SAN Volume Controller" at the following link:

https://www.ibm.com/support/docview.wss?uid=ssg1S1003907

- Stretched Cluster enhancements

  SVC V6.3.0 introduces the ability to extend the distance between SVC nodes in a Stretched Cluster (Split I/O Group) configuration. While the extended distances depend on application latency restrictions, this function now enables enterprises to access and share a consistent view of data simultaneously across data centers. This function also enables enterprises to relocate data across disk array vendors and tiers, both inside and between data centers at full metro distances.

- Enhanced LDAP authentication support

  V6.3 introduces support for direct authentication to an LDAP server. Authentication via IBM Tivoli Integrated Portal is still supported, but no longer required.

- Support for CLI password authentication

  V6.3 introduces support for CLI authentication using a password. Authentication via SSH key is still supported, but no longer required.

- Storwize V7000 support for additional drive types

  V6.3 introduces support for two new drive types:

  – Storwize V7000 3 TB 3.5" 7.2 K RPM Near-Line SAS drive
  – Storwize V7000 200 GB and 400 GB 2.5" flash drives

- GUI Enhancements

  V6.3 introduces various GUI-related enhancements in the following categories:

  – Usability
    - Per-column grid filtering support
    - New tree table views (MDisks by pool, FC consistgrp, and RC consistgrp)
    - New IBM XIV style status pods
    - Recommended actions and events panels combined
    - Easy Tier icon badge for easy tier pools
  – New cluster features
    - Quorum disk management support (SVC)
    - Native LDAP support
    - Support for Global Mirror with Change volumes
  – Performance monitoring
    - Read and write latency statistics for volumes and MDisk

- Interoperability with new storage controllers, host operating systems, fabric devices, and other hardware

  For an updated list, see "V6.3 Supported Hardware List, Device Driver, Firmware and Recommended Software Levels for SAN Volume Controller" at the following link:

  https://www.ibm.com/support/docview.wss?uid=ssg1S1003907

2.6 Enhancements and changes in SAN Volume Controller V6.2

SAN Volume Controller V6.2 has the following enhancements and changes:

- Support for SAN Volume Controller 2145-CG8

  The new 2145-CG8 engine contains 24 GB of cache and four 8 Gbps FC host bus adapter (HBA) ports for attachment to the SAN. The 2145-CG8 auto-negotiates the fabric speed on a per-port basis and is not restricted to run at the same speed as other node pairs in the clustered system. The 2145-CG8 engine can be added in pairs to an existing system that consists of 64-bit hardware nodes (8F2, 8F4, 8G4, 8A4, CF8, or CG8) up to the maximum of four pairs.

- 10 GbE iSCSI host attachment

  The new 2145-CG8 node comes with the option to add a dual port 10 Gbps Ethernet (GbE) adapter, which can be used for iSCSI host attachment. The 2145-CG8 node also supports the optional use of flash devices (up to four). However, both options cannot coexist on the same SVC node.

- Real-time performance statistics through the management GUI

  Real-time performance statistics provide short-term status information for the system. The statistics are shown as graphs in the management GUI. Historical data is kept for about five minutes. Therefore, you can use Tivoli Storage Productivity Center to capture more detailed performance information, to analyze mid-term and long-term historical data, and to have a complete picture when you develop best-performance solutions.

- Flash drive RAID at levels 0, 1, and 10

  Optional flash drives are not accessible over the SAN. Their usage is done through the creation of RAID arrays. The supported RAID levels are 0, 1, and 10. In a RAID 1 or RAID 10 array, the data is mirrored between flash drives on two nodes in the same I/O group.

- Easy Tier for use with flash drives on 2145-CF8 and 2145-CG8 nodes

  SAN Volume Controller V6.2 restarts support of internal flash drives by allowing Easy Tier to work with internal Subsystem Device Driver (SDD) storage pools.

- Support for a FlashCopy target as a remote copy source

  In SAN Volume Controller V6.2, a FlashCopy target volume can be a source volume in a remote copy relationship.

- Support for the VMware vStorage API for Array Integration (VAAI)

  SAN Volume Controller V6.2 fully supports the VMware VAAI protocols. An improvement that comes with VAAI support is the ability to dramatically offload the I/O processing that is generated by performing a VMware Storage vMotion.

- CLI prefix removal

  The **`svctask`** and **`svcinfo`** command prefixes are no longer necessary when you issue a command. If you have existing scripts that use those prefixes, they continue to function.

- Licensing change for the removal of a physical site boundary

  The licensing for SAN Volume Controller systems (formerly clusters) within the same country and that belong to the same client can be aggregated in a single license.

- FlashCopy license on the main source volumes

  SAN Volume Controller V6.2 changes the way the FlashCopy is licensed so that SAN Volume Controller now counts as the main source in FlashCopy relationships. Previously, if cascaded FlashCopy was set up, multiple source volumes had to be licensed.

- Interoperability with new storage controllers, host operating systems, fabric devices, and other hardware

  For an updated list, see “V6.2 Supported Hardware List, Device Driver, Firmware and Recommended Software Levels for SAN Volume Controller” at the following link:

  https://www.ibm.com/support/docview.wss?uid=ssg1S1003797

- Exceeding entitled virtualization license 45 days from the installation date for migrating data from one system to another

  With the benefit of virtualization, by using SAN Volume Controller, clients can bring new storage systems into their storage environment and quickly and easily migrate data from their existing storage systems to the new storage systems. To facilitate this migration, IBM clients can temporarily (45 days from the date of installation of the SAN Volume Controller) exceed their entitled virtualization license for migrating data from one system to another.

Table 2-1 shows the current and previous usage of one changed common term.

Table 2-1 SVC Version 6.2 terminology mapping table

Term in SAN Volume Controller V6.2	Term in previous versions of SAN Volume Controller	Description
Clustered system or system	Cluster	A collection of nodes that is placed in pairs (I/O groups) for redundancy, which provides a single management interface.

2.7 Enhancements and changes in SAN Volume Controller V6.1

The SAN Volume Controller V6.1 is one of the key milestones in the product development path. It includes the following major hardware and software enhancements and changes that help to simplify the management of your storage infrastructure and drive the operational cost down:

- A newly designed user interface (similar to IBM XIV Storage System)

  The SVC Console has a newly designed GUI that now runs on the SAN Volume Controller and can be accessed from anywhere on the network by using a web browser. The interface includes several enhancements such as greater flexibility of views, display of running command lines, and improved user customization within the GUI. Clients who use Tivoli Storage Productivity Center and IBM Systems Director can take advantage of integration points with the new SVC console.

- SAN Volume Controller and Storwize V7000 interoperability

  The virtualization layer of IBM Storwize V7000 is built upon the IBM SAN Volume Controller technology. SAN Volume Controller V6.1 is the first version that is supported in this environment.

- New licensing for SAN Volume Controller for XIV (5639-SX1)

  Product ID 5639-SX1, IBM SAN Volume Controller for XIV Software V6, is priced by the number of storage devices (also called *modules* or *enclosures*). It eliminates the appearance of double charging for features that are bundled in the XIV software license. Also, you can combine this license with a per TB license to extend the usage of SAN Volume Controller with a mix of back-end storage subsystems.

- Service Assistant

  SAN Volume Controller V6.1 introduces a new method for performing service tasks on the system. In addition to performing service tasks from the front panel, you can service a node through an Ethernet connection by using a web browser or command-line interface (CLI). The web browser runs a new service application that is called the *Service Assistant*. All functions that were previously available through the front panel are now available from the Ethernet connection, with the advantages of an easier to use interface and remote access from the cluster. Furthermore, you can run Service Assistant commands through a USB flash drive for easier serviceability.

- IBM System Storage Easy Tier function added at no charge

  SAN Volume Controller V6.1 delivers IBM System Storage Easy Tier, which is a dynamic data relocation feature that allows host transparent movement of data among two tiers of storage. This feature includes the ability to automatically relocate volume extents with high activity to storage media with higher performance characteristics. Extents with low activity are migrated to storage media with lower performance characteristics. This capability aligns the SAN Volume Controller system with current workload requirements, increasing overall storage performance.

- Temporary withdrawal of support for flash drives on the 2145-CF8 nodes

  At the time of writing, 2145-CF8 nodes that use internal flash drives are unsupported with V6.1.0.x code (fixed in version 6.2).

- Interoperability with new storage controllers, host operating systems, fabric devices, and other hardware

  For an updated list, see "V6.1 Supported Hardware List, Device Driver, Firmware and Recommended Software Levels for SAN Volume Controller" at the following link:

  https://www.ibm.com/support/docview.wss?uid=ssg1S1003697

- Removal of 15-character maximum name length restrictions

  SAN Volume Controller V6.1 supports object names up to 63 characters. Previous levels supported only up to 15 characters.

- SAN Volume Controller code upgrades

  The SVC console code is now removed. Now you need only to update the SAN Volume Controller code. The upgrade from SAN Volume Controller V5.1 requires usage of the former console interface or a command line. After the upgrade is complete, you can remove the existing ICA console application from your SSPC or master console. The new GUI is started through a web browser that points to the SAN Volume Controller IP address.

- SAN Volume Controller to back-end controller I/O change

  SAN Volume Controller V6.1 allows variable block sizes, up to 256 KB against 32 KB supported in the previous versions. This change is handled automatically by the SAN Volume Controller system without requiring any user control.

- Scalability

  The maximum extent size increased four times to 8 GB. With an extent size of 8 GB, the total storage capacity that is manageable for each cluster is 32 PB. The maximum volume size increased to 1 PB. The maximum number of worldwide node names (WWNNs) increased to 1024, allowing up to 1024 back-end storage subsystems to be virtualized.

To coincide with new and existing IBM products and functions, several common terms changed and are incorporated in the SAN Volume Controller information. Table 2-2 shows the current and previous usage of the changed common terms.

Table 2-2 SVC Version 6.1 terminology mapping table

Term in SAN Volume Controller V6.1	Term in previous versions of SAN Volume Controller	Description
Event	Error	A significant occurrence to a task or system. Events can include completion or failure of an operation, a user action, or the change in state of a process.
Host mapping	VDisk-to-host mapping	The process of controlling which hosts have access to specific volumes within a cluster.
Storage pool	Managed disk group	A collection of storage capacity that provides the capacity requirements for a volume.
Thin provisioning (thin-provisioned)	Space efficient	The ability to define a storage unit (full system, storage pool, and volume) with a logical capacity size that is larger than the physical capacity that is assigned to that storage unit.
Volume	Virtual disk (VDisk)	A discrete unit of storage on disk, tape, or other data recording medium that supports some form of identifier and parameter list, such as a volume label or I/O control.

2.8 Enhancements and changes in SAN Volume Controller V5.1

The following major enhancements and changes were introduced in SAN Volume Controller V5.1:

- New capabilities with the 2145-CF8 hardware engine

  SAN Volume Controller offers improved performance capabilities by upgrading to a 64-bit software kernel. With this enhancement, you can take advantage of cache increases, such as 24 GB, that are provided in the new 2145-CF8 hardware engine. SAN Volume Controller V5.1 runs on all SAN Volume Controller 2145 models that use 64-bit hardware, including Models 8F2, 8F4, 8A4, 8G4, and CF8. The 2145-4F2 node (32-bit hardware) is not supported in this version.

  SAN Volume Controller V5.1 also supports optional solid-state drives (flash drives) on the 2145-CF8 node, which provides a new ultra-high-performance storage option. Each 2145-CF8 node supports up to four flash drives with the required serial-attached SCSI (SAS) adapter.

- Multitarget reverse IBM FlashCopy and Storage FlashCopy Manager

  With SAN Volume Controller V5.1, reverse FlashCopy support is available. With reverse FlashCopy, FlashCopy targets can become restore points for the source without breaking the FlashCopy relationship and without waiting for the original copy operation to complete. Reverse FlashCopy supports multiple targets, and therefore, multiple rollback points.

- 1 Gbps iSCSI host attachment

  SAN Volume Controller V5.1 delivers native support of the iSCSI protocol for host attachment. However, all internode and back-end storage communications still flow through the Fibre Channel (FC) adapters.

- I/O group split in SAN Volume Controller across long distances

  With the option to use 8 Gbps Longwave (LW) Small Form Factor Pluggables (SFPs) in the SAN Volume Controller 2145-CF8, SAN Volume Controller V5.1 introduces the ability to split an I/O group in SAN Volume Controller across long distances.

- Remote authentication for users of SVC clusters

  SAN Volume Controller V5.1 provides the Enterprise Single Sign-on client to interact with an LDAP directory server such as IBM Tivoli Directory Server or Microsoft Active Directory.

- Remote copy functions

  The number of cluster partnerships increased from one up to a maximum of three partnerships. That is, a single SVC cluster can have partnerships of up to three clusters at the same time. This change allows the establishment of multiple partnership topologies that include star, triangle, mesh, and daisy chain.

  The maximum number of remote copy relationships increased to 8192.

- Increased maximum virtual disk (VDisk) size to 256 TB

  SAN Volume Controller V5.1 provides greater flexibility in expanding provisioned storage by increasing the allowable size of VDisks from the former 2 TB limit to 256 TB.

- Reclaiming unused disk space by using space-efficient VDisks and VDisk mirroring

  SAN Volume Controller V5.1 enables the reclamation of unused allocated disk space when you convert a fully allocated VDisk to a space-efficient virtual disk by using the VDisk mirroring function.

- New reliability, availability, and serviceability (RAS) functions

  The RAS capabilities in SAN Volume Controller are further enhanced in V5.1. Administrators benefit from better availability and serviceability of SAN Volume Controller through automatic recovery of node metadata, with improved error notification capabilities (across email, syslog, and SMNP). Error notification supports up to six email destination addresses. Also, quorum disk management improved with a set of new commands.

- Optional second management IP address configured on eth1 port

  The existing SVC node hardware has two Ethernet ports. Until SAN Volume Controller V4.3, only one Ethernet port (eth0) was used for cluster configuration. In SAN Volume Controller V5.1, a second, new cluster IP address can be optionally configured on the eth1 port.

- Added interoperability

  Interoperability is now available with new storage controllers, host operating systems, fabric devices, and other hardware. For an updated list, see “V5.1.x - Supported Hardware List, Device Driver and Firmware Levels for SAN Volume Controller” at the following link:

  https://www.ibm.com/support/docview.wss?uid=ssg1S1003553

- Withdrawal of support for 2145-4F2 nodes (32-bit)

  As stated previously, SAN Volume Controller V5.1 supports only SAN Volume Controller 2145 engines that use 64-bit hardware. Therefore, support is withdrawn for 32-bit 2145-4F2 nodes.

- Up to 250 drives, running only on 2145-8A4 nodes, allowed by SAN Volume Controller Entry Edition

  The SAN Volume Controller Entry Edition uses a per-disk-drive charge unit and now can be used for storage configurations of up to 250 disk drives.

3

2145-DH8 and SAN Volume Controller software version 7.3

In conjunction with the release of the 7.3 code, IBM introduced a hardware refresh for the IBM SAN Volume Controller (SVC) platform. This new hardware model 2145-DH8 includes the enhancements outlined in Chapter 2, “Change history” on page 23.

In this chapter, we focus on the details of the specific hardware changes and version 7.3 software enhancements including:

- New SVC platform based on IBM x3650 M4 server with front LCD panel removed.
- Integrated battery pack replacing the external uninterruptible power supply (UPS).
- Dual boot drives.
- Enhanced scalability and flexibility with more PCIe slots.
- Expansion unit 2145-24F for 24 flash drives.
- Cache rearchitecture in version 7.3.
- Improved Real-time Compression engine with hardware assistance.
- Enhanced Easy Tier function with automatic Storage Pool Balancing.
- Modified functionality of FlashCopy leveraging shared cache.

3.1 New hardware 2145-DH8

The SVC storage engine model DH8 and SVC SFF Expansion Enclosure Model 24F deliver increased performance, expanded connectivity, compression acceleration, and additional internal flash storage capacity.

SVC Storage Engine Model DH8, which is based on IBM System x® server technology, consists of one Xeon E5 v2 eight-core 2.6 GHz processor and 32 GB of memory. It includes three 1 Gb Ethernet ports as standard for 1 Gb iSCSI connectivity and supports up to three I/O adapter cards for 8 Gb FC and 10 Gb iSCSI/FCoE connectivity. It also includes two integrated AC power supplies and battery units replacing the uninterruptible power supply feature that was required on the previous generation storage engine models.

The front view of the two-node cluster based on the 2145-DH8 is shown in Figure 3-1.

Figure 3-1 Front view of the 2145-DH8

The 2145-DH8 brings with it several significant changes and enhancements over the previous generation hardware, the 2145-CG8. These changes include:

- A 19-inch rack-mounted 2U device enables greater performance and flexibility.
- Processor upgrade from 6-core to 8-core Ivy Bridge.
- Enhanced I/O connectivity with up to 12 FC ports (8 Gbps) or four 10 GbE iSCSI/FCoE.
- Memory allocation for new cache architecture.
- Two integrated battery units eliminate a need for separate UPS or power switch.
- Dual boot drives with boot data mirrored.
- LCD front panel is replaced by LED system indicators.
- Optional 2145-24F expansion enclosure with space for 24 Small Form Factor flash drives.
- Two slots reserved for hardware-assisted Real-time Compression.
- Technician port for easy initial configuration.

The IBM SAN Volume Controller 2145-DH8 ships with preinstalled V7.3 software. It is not supported to downgrade the software to the version 7.2 or lower. The 2145-DH8 will reject any attempt to install a version lower than 7.3. See the following link for integration with existing clustered systems, compatibility, and interoperability with installed nodes and other system components:

http://www.ibm.com/support/docview.wss?uid=ssg1S1002999

3.1.1 IBM SAN Volume Controller platform

The next generation of the IBM storage virtualization flagship, the IBM SAN Volume Controller, is not just a migration to the newest hardware with certain enhancements in the software. The 2145-DH8, with its incorporated V7.3 software, evolves to a platform that is significantly more scalable, flexible, and powerful, than any previous generation-to-generation upgrade. And still provides the same best-of-breed ease of use.

Scalability

Moving from the traditional 1U platform to a 2U platform based on IBM x3650 M4 server makes more room available for the deployment of more hardware components:

- Up to 3x more I/O slot capability
- Up to 2x more processor resources
- Up to 4x more memory capacity
- Up to 3x more local internal disks
- Up to 192 SVC-owned flash drives through a 12 Gbps SAS attachment

With the migration to the 2U platform and integration of previous external components into a single chassis, no more rack space is required. The 2145-DH8 integrates redundant battery backup system, eliminating the need for external rack-mount UPS, optional power switch, and related cabling.

The system is just as scalable as previous generations of SVC, with configuring the system of up to four dual-node clusters (I/O groups) possible. The IBM SAN Volume Controller model 2145-DH8 seamlessly integrates into the existing infrastructure and enables nondisruptive hardware upgrades from previous generations. The upgrade process and its prerequisites are briefly described in 3.3, “Upgrade considerations” on page 53.

Flexibility

The IBM SAN Volume Controller system flexibility is represented by various combinations of I/O connectivity, reflecting the complexity of your SAN infrastructure. Specifically, the number and types of backend storage devices and requirements for host protocol attachment, dictate the need for the increased variability of I/O cards within an I/O group. The list of available I/O cards (do not confuse with Host Interface Cards, as they are known in the IBM Storwize family of products) and their possible combinations are provided in 3.1.2, “I/O connectivity” on page 42.

The option to install a secondary 8-core Xeon E5 v2 processor in a bundle with an additional 32 GB of memory for cache, increases the throughput for the highest performance and enables advanced features such as IBM Real-time Compression.

Performance

The 2145-DH8 moves system performance forward in the following ways:

- Moving from 6-core to 8-core Ivy Bridge processor pushes raw IOPS twice 70/30 miss IOPS compared to previous SVC.
- Migration to a maximum of three I/O cards and PCIe Gen3 vastly improves system bandwidth by up to three times Read and Write Miss MB/s compared to previous SVC models.
- Hardware Compression Accelerator cards help Real-time Compression deliver across more compressed volumes and faster.
- Expansion enclosures 2145-24F that are optimized for dual-port attached flash drives give more than 8 GBps throughput with 300 K IOPS capability from the flash drive itself.

More details about SVC performance and its considerations are discussed in Chapter 8, “Performance data and statistics gathering” on page 153.

Hardware layout

The 2145-DH8 has adopted the layout and system architecture of the very successful IBM x3650 M4 server with specific hardware modifications that match the needs of the SVC.

In its minimal configuration, this SVC-specific server incorporates a single Intel 2.6 GHz eight-core Xeon E5 v2 Ivy Bridge processor, 32 GB memory for SVC cache, and two riser cards with 3 PCIe Ge3 slots for I/O connectivity each. The second riser card with slots 4-6 becomes active when a secondary processor, an additional 32 GB of memory, and a supplemental fan module is installed in the node.

The front panel unit contains a dual-battery pack with its backplane acting as an uninterruptible power supply to the SVC. Batteries are fully redundant and hot-swappable. An SVC node is able to operate with one healthy battery, however it will be logged and the node becomes degraded, but will still continue I/O operations. The embedded disk shelf includes a pair of mirrored 2.5” 300 GB 10 K RPM SAS boot drives.

The 2145-DH8 hardware layout is presented in Figure 3-2. The mandatory components of basic configuration are marked red; the optional upgrades are highlighted in blue.

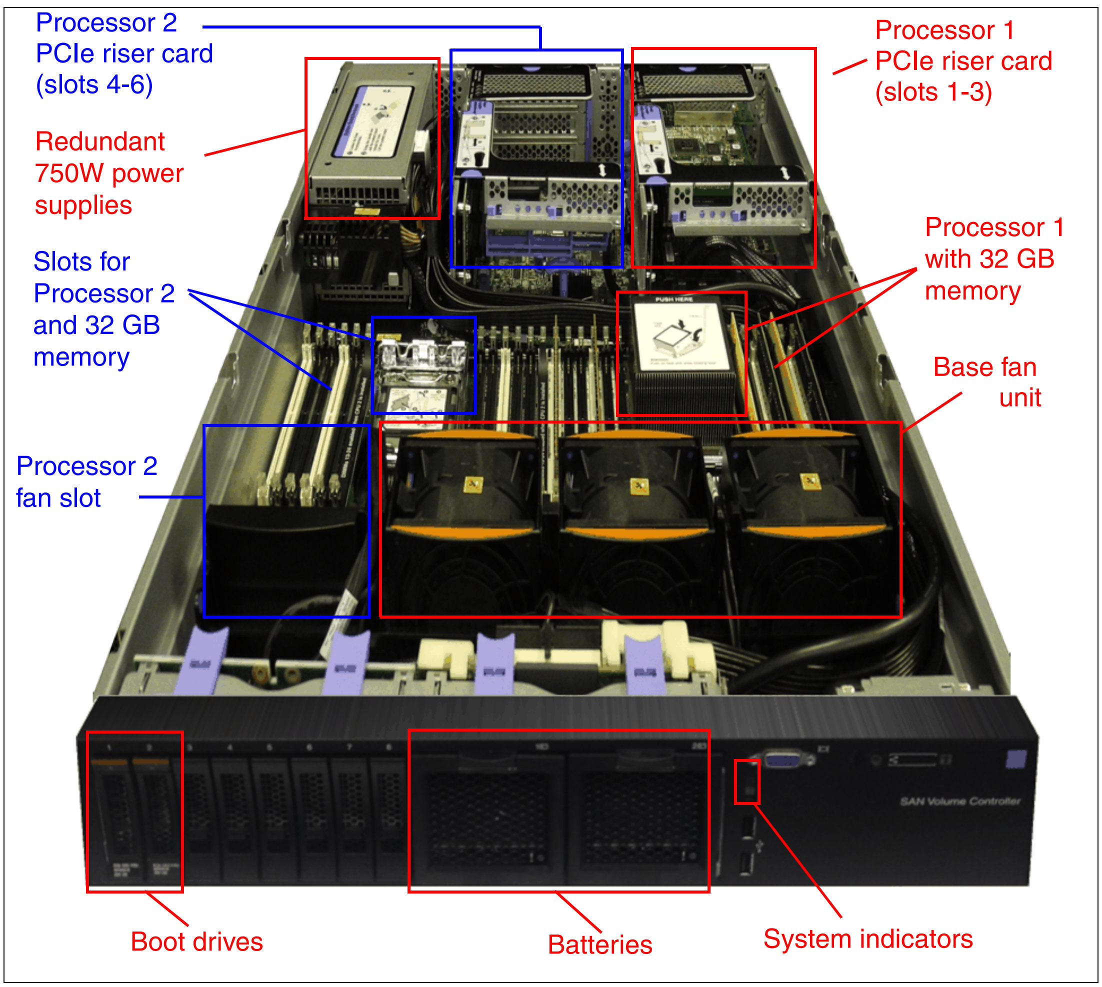


Figure 3-2 The SVC hardware layout

The 2145-DH8 does not include the front LCD panel for local user service interaction and its functionality has been integrated with the Technician port on the rear of the device. The functionality and purpose of the Technician port is described in 3.1.5, “Technician port” on page 46. An overview of the new front panel is shown in Figure 3-3.

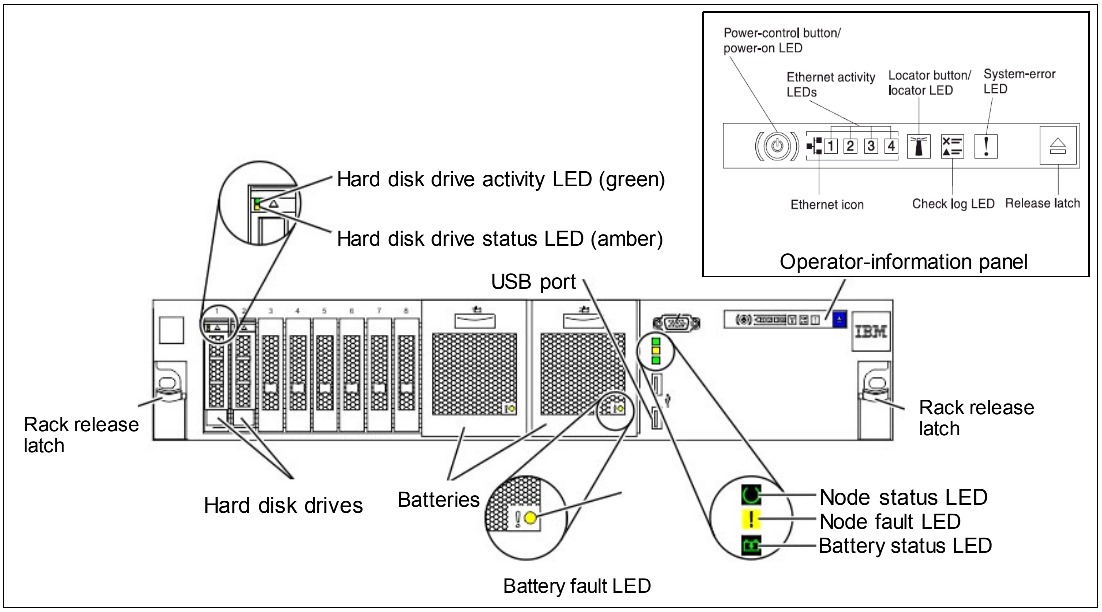


Figure 3-3 Front panel scheme with indicator LEDs

The operator-information panel provides standard IBM System x notifications, while the set of three vertical LEDs next to the batteries is customized to SVC node requirements. The meaning of each LED indicator is as follows:

- Node status LED:
  - Off - node is powered off or SVC software is not running
  - Blink (once a second) - node is in candidate or service state
  - Fast blink - node is currently dumping data to disk
  - On - node is active in a cluster
- Node fault LED:
  - On - node has a fatal error
  - Off - node does not have any fatal error preventing it from running SVC
- Battery status LED:
  - On - two Firehose Dumps (FHDs) are supported
  - Blink - one FHD is supported by system
  - Fast blink - no FHD supported due to insufficient charge, but charging is in progress
  - Off - no FHD supported, which typically indicates a battery fault

Configuration options

The SAN Volume Controller model 2145-DH8 is available in three functional configurations:

- Base configuration:

  The SVC node ships with one processor, and 32 GB of memory — thus it cannot support IBM Real-time Compression (RtC). Up to three I/O cards can be installed in the PCIe riser attached to processor 1. If there is a 2145-24F enclosure attached to the system, the I/O connectivity is limited to two adapters; slot 3 is used for 12 Gbps SAS attachment. A node in this configuration cannot join the I/O group or system with compressed volumes already defined and any attempt to do so will fail.

- Extended I/O connectivity:

  In this configuration, the second processor and additional 32 GB of memory (64 GB in total) is installed into the node in order to enable utilization of slots 4 - 6 in the second PCIe riser card. This node supports an additional I/O card that can be installed in slot 5. It can be either FC or 10 GbE, however only one 10 GbE card is allowed in the node.

- Compression-enabled:

  In order to support compressed volumes, at a minimum one Compression Accelerator card needs to be installed in slot 4 or 6. This also dictates the requirement for the installation of the second processor, giving 64 GB of memory in total (38 GB of it reserved for IBM Real-time Compression cache) to enable the second PCIe riser. An SVC node with a single Compression Accelerator card supports up to 200 compressed volumes, and with two accelerators it scales to 512 compressed volumes.

IBM feature codes

The IBM SAN Volume Controller model 2145-DH8 with SVC software V7.3 introduces new feature codes as shown in Table 3-1. Not all components are mentioned. For the complete list, consult your storage specialist or IBM service representative.

Table 3-1 SVC feature codes

Description	Machine type	Model	Feature code
SVC Storage Engine	2145	DH8	
SVC SFF Expansion Enclosure	2145	24F	
4-port 8Gb FC Card	2145	DH8	AH10
4-port 10Gb Ethernet Card	2145	DH8	AH12
SAS Enclosure Attach Card	2145	DH8	AH13
Compression Accelerator	2145	DH8	AH1A
1st CPU and 32GB RAM	2145	DH8	AH15
2nd CPU and 32GB RAM	2145	DH8	AH1B
200GB 2.5 Inch Flash Drive	2145	24F	AH20
400GB 2.5 Inch Flash Drive	2145	24F	AH21
800GB 2.5 Inch Flash Drive	2145	24F	AH22
1.5m 12Gb SAS Cable (mSAS HD)	2145	24F	ACUB
3m 12Gb SAS Cable (mSAS HD)	2145	24F	ACUC
6m 12Gb SAS Cable (mSAS HD)	2145	24F	ACUD

3.1.2 I/O connectivity

The IBM SAN Volume Controller model 2145-DH8 offers various options of I/O cards for installation and configuration. The I/O connectivity capabilities of each node are based on the type of purchased node configuration as described in "Configuration options" on page 41, namely the number of installed processors and amount of memory.

The 2U rack-mount form factor of the 2145-DH8 allows the SVC node to accommodate up to six PCIe Gen3 cards for I/O connectivity or compression support. The rear view of the 2145-DH8 is shown in Figure 3-4.

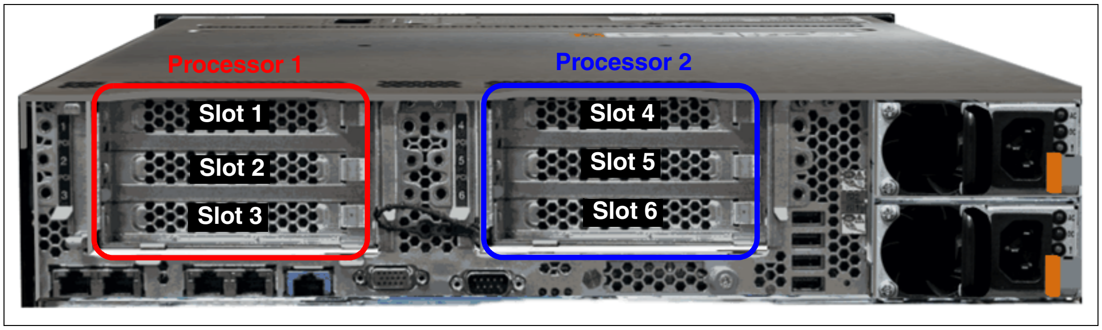


Figure 3-4 Rear view of 2145-DH8

Slots 4-6 are internally attached to Processor 2, and therefore available only in cases where both processors and 64 GB of total memory are installed in the SVC node. The installation options for each type of I/O card are outlined in Table 3-2.

Table 3-2 Layout of I/O card options

Top of Node	
Processor 1 attach	**Processor 2 attach**
Slot 1 - I/O card (FC only)	Slot 4 - Compression Accelerator card
Slot 2 - I/O card	Slot 5 - I/O card
Slot 3 - SAS (for flash expansion)	Slot 6 - Compression Accelerator card

However, there are certain requirements and limitations of I/O card combinations:

- Minimum one 4-port 8 Gbps FC card (feature code AH10) is required in slot 1 (node is shipped with it from manufacturing).
- For maximum FC connectivity (12 ports) the second processor and 64 GB memory needs to be installed in order to activate slots 4 - 6. Third FC card (AH10) is seated in slot 5.
- Maximum quantity of 4-port 10 GbE card (AH12) is one per node either in slot 2 or 5 (slot 5 requires second processor and 64 GB memory). 10 GbE card supports both iSCSI or FCoE connectivity.
- One or two Compression Accelerator cards (AH1A) can be installed in slots 4 or 6 when the secondary processor and 64 GB memory is available in the SVC node.
- 4-port 12 Gbps SAS card (AH13) for 2145-24F expansion enclosure seats in slot 3. Only ports 1 and 3 on the card are used for SAS cabling.

3.1.3 Dual boot drives

The IBM SAN Volume Controller 2145-DH8 has two internal drives and will be able to boot from either drive. Previous SVC nodes had one internal boot drive but had a second boot device as part of the front LCD panel (in the most recent version, the front panel included a USB stick). Unlike the second boot device in the old version, both boot devices in the DH8 contain a full installation of SVC software. To support a larger write cache (and more non-volatile data structures) the SVC stripes the FHD across both drives in the event of AC power loss. This will double the rate at which data can be written to disk.

The SVC 2145-DH8 takes advantage of having two internal drives to provide new functionality:

- The emergency FHD when AC power is lost will stripe data across both internal drives for performance reasons. This allows SVC to support a larger write cache.
- The alternative boot device on previous models did not contain a full software installation. With two internal boot drives, the DH8 is more resilient to drive failures, and in particular, during a concurrent code upgrade other service actions are able to continue operation.

The 2145-DH8 has two batteries and two internal drives. The normal procedure when AC power fails is to stripe the FHD across both drives such that each drive has half of the cache contents. With two fully charged batteries, there is sufficient charge to perform two FHDs in this way. A node will not start I/O until it has sufficient battery charge to perform at least one FHD. Consequently, with two drives and two batteries the node can perform an FHD and then boot and restart I/O operations as soon as AC power is restored. If a second AC failure occurs before the batteries have charged, a second FHD can be performed, but the node will then have to wait for the batteries to charge (to 50%) before resuming I/O processing.

If one battery has failed, as long as the remaining battery is fully charged the node will be able to perform a Firehose Dump striping data across both drives. When AC power is restored, the node must wait for the single battery to fully charge (to 100%) before restarting I/O operations.

Note: Operating with just one battery is an error condition. Using just one battery to perform a Firehose Dump shortens the life of the battery.

If one drive fails, as long as there are two fully charged batteries the node will be able to perform an FHD writing all the data to the healthy drive. This FHD takes twice as long as a normal Firehose Dump and will therefore use up all the battery charge. When AC power is restored, the node must wait for the two batteries to fully charge (to 100%) before restarting I/O operations.

Note: Operating with just one drive is an error condition. This prevents battery maintenance procedures from running.

If one drive and one battery has failed, the node enters service state and will refuse to operate until at least one of the faults is fixed. Example 3-1 shows the status of boot drives.

Example 3-1 Status of boot drives

```
IBM_2145:ITSO_SVC2:superuser>sainfo lsbootdrive
panel_id node_id node_name  can_sync slot_id  booted status actual_drive_sn actual_node_sn
KD8P1BP  2       ITSO_SVCN2 no       1        yes    online 3410A02FFTH5    KD8P1BP
KD8P1BP  2       ITSO_SVCN2 no       2        no     online 24R0A056FTH5    KD8P1BP
```

3.1.4 Compression Accelerator card

Compressed volumes are a special type of volume where data is compressed as it is written to disk, saving additional space. To use the compression function, you must obtain the IBM Real-time Compression license and the hardware level for both nodes within the I/O group must be either SAN Volume Controller 2145-DH8, 2145-CG8, or 2145-CF8 for that I/O group to support compression. SAN Volume Controller 2145-DH8 nodes must have two processors, 64 GB of memory, and at least one *Compression Accelerator* card installed in order to use compression. Enabling compression on SAN Volume Controller 2145-DH8 nodes does not affect non-compressed host to disk I/O performance. Figure 3-5 shows the Compression Accelerator card and its possible placement in the SVC node.

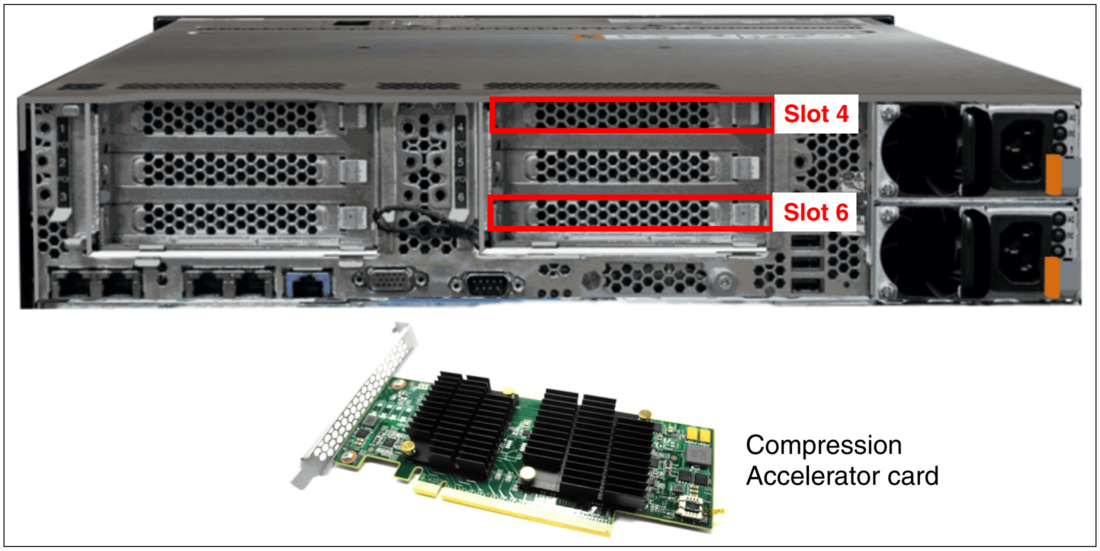


Figure 3-5 Placement of Compression Accelerator cards

The slot 5 connected to the second processor is reserved for an additional I/O card. If there is no I/O card installed in the slot 5, a compression card can be in any slot connected to this second processor. However, we strongly recommend placing Compression Accelerator cards into their dedicated slots 4 and 6.

Up to a total of two Compression Accelerator adapters can be installed, each additional that is installed improves the I/O performance and in particular the maximum bandwidth when using compressed volumes.

Remember: At a minimum, one Compression Accelerator card is compulsory for each node in an I/O group in order to use compression.

For an I/O Group containing a 2145-DH8 with no compression accelerator, an attempt to create a compressed volume fails. The `addnode` command will also fail if trying to add a 2145-DH8 without a compression accelerator to an I/O group or system that already has any compressed volumes.

An I/O group consisting of 2145-DH8 nodes with a single compression accelerator in each supports a maximum of 200 compressed volumes, while a fully equipped I/O group with four compression accelerators supports up to 512 compressed volumes. Table 3-3 summarizes the configuration options within the I/O group and its associated parameters.

Table 3-3 Configuration options and dependencies for each node

Accelerators in I/O group	Required processors	Required memory	RtC cache allocation	Compressed volumes	License needed?
1 per node	2	64 GB	36 GB	max. 200	Yes
2 per node	2	64 GB	36 GB	max. 512	Yes

More details about implementation of IBM Real-time Compression in 2145-DH8 are provided in Chapter 7, “Real-time Compression and the SVC 2145-DH8 Node” on page 139.

3.1.5 Technician port

The purpose and key benefit of the IBM SAN Volume Controller Technician port is to simplify and ease the initial basic configuration of the SVC system by the local administrator or by service personnel. It eliminates the need for the LCD front panel as presented on all previous models. The Technician port is marked with a **T** (Ethernet port 4) as depicted in Figure 3-6.

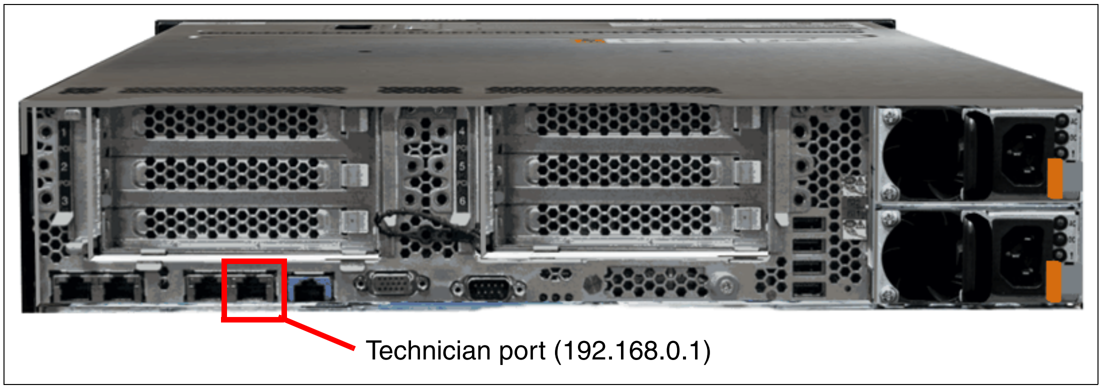


Figure 3-6 Location of Technician port

To initialize a new system, you must connect a personal computer to the Technician port on the rear of a node canister and run the initialization tool. This port runs a Dynamic Host Configuration Protocol (DHCP) server in order to facilitate service/maintenance out of the box in lieu of the front panel. Make sure that your computer has DHCP enabled, otherwise manually set the IP address to `192.168.0.2`, network mask to `255.255.255.0`, and gateway to `192.168.0.1`.

Attention: Never connect the Technician port to a switch. If a switch is detected, the Technician port connection might shut down, causing an SVC node error 746 in the log.

As soon as the system is prepared and you are connected to the Technician port you get directed to the new Init tool, where the initial Service IP needs to be configured. Service IP is not associated with the Technician port, but will continue to be assigned to Ethernet port 1.

If the node has Candidate status when you open the web browser, the initialization tool is displayed. Otherwise, the service assistant interface is displayed. Only the Technician port has a Password Reset option from the Service Assistant available, as shown in Figure 3-7.

Figure 3-7 Resetting default superuser password

The whole procedure of the initial configuration is described step-by-step in Chapter 6, “SAN Volume Controller initial configuration” on page 113.

3.1.6 Expansion enclosure

Expansion enclosures 2145-24F are rack-mount hardware that contains several components of the system: canisters, drives, and power supplies. Enclosures can be used to extend the capacity of the system and are only supported on IBM SAN Volume Controller 2145-DH8 nodes. For other models of the system, you must use external storage systems to provide capacity for data. The 2145-24F supports flash drives of capacity 200, 400, or 800 GB. The front and real view of the enclosure is shown in Figure 3-8.

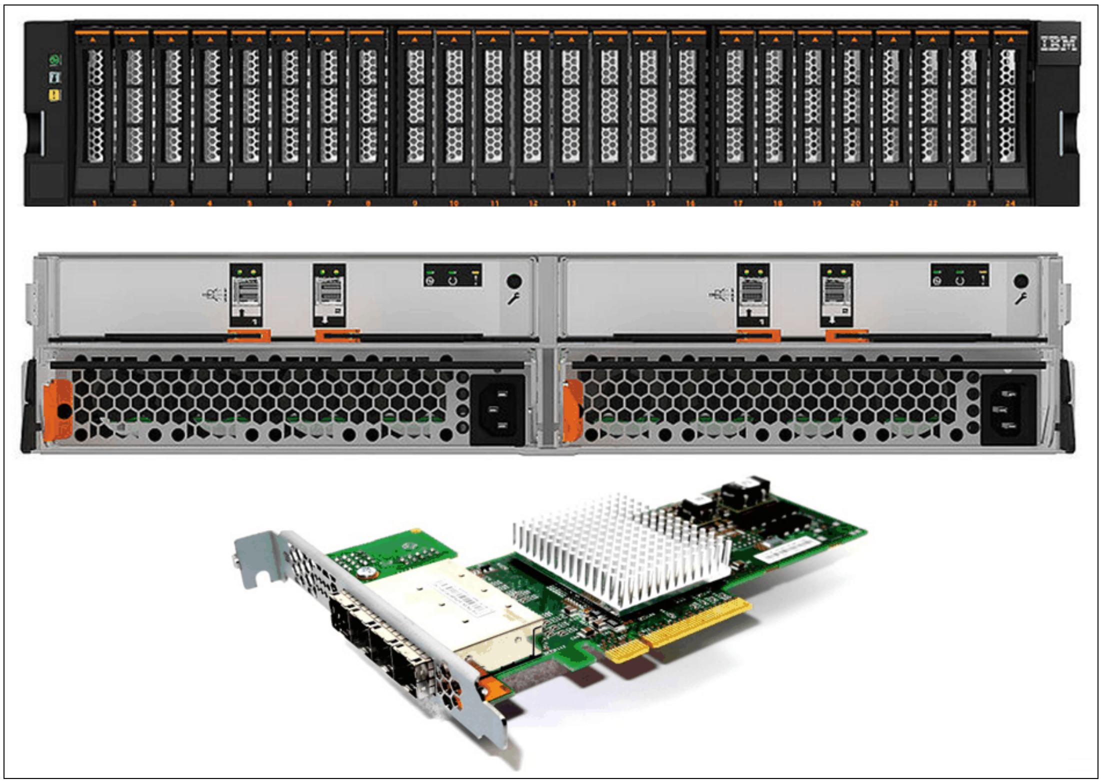


Figure 3-8 The expansion enclosure 2145-24F and 12 Gbps SAS card

The optional 2U expansion enclosure 2145-24F accommodates up to 24 flash drives (for tier 0 when Easy Tier function is enabled). The RAID array can be built across both expansions attached to each node in the I/O group. The 12 Gbps SAS four-port adapter needs to be installed in each SVC node.

The system supports up to four I/O groups with a total of eight expansion enclosures per system. No more than one expansion enclosure can be chained to either of the two SAS ports on the node (SAS ports 1 and 3). This configuration creates two separate chains of one enclosure each. Also, no cable can be connected between a port on a left canister and a port on a right canister of the expansion enclosures.

The recommended deployment of devices in the simplest environments is illustrated in Figure 3-9 on page 48. It is a good practice to place expansion units 2145-24F between their parent SVC nodes in a rack. It helps to manage simplicity and transparency of the system layout in the rack.

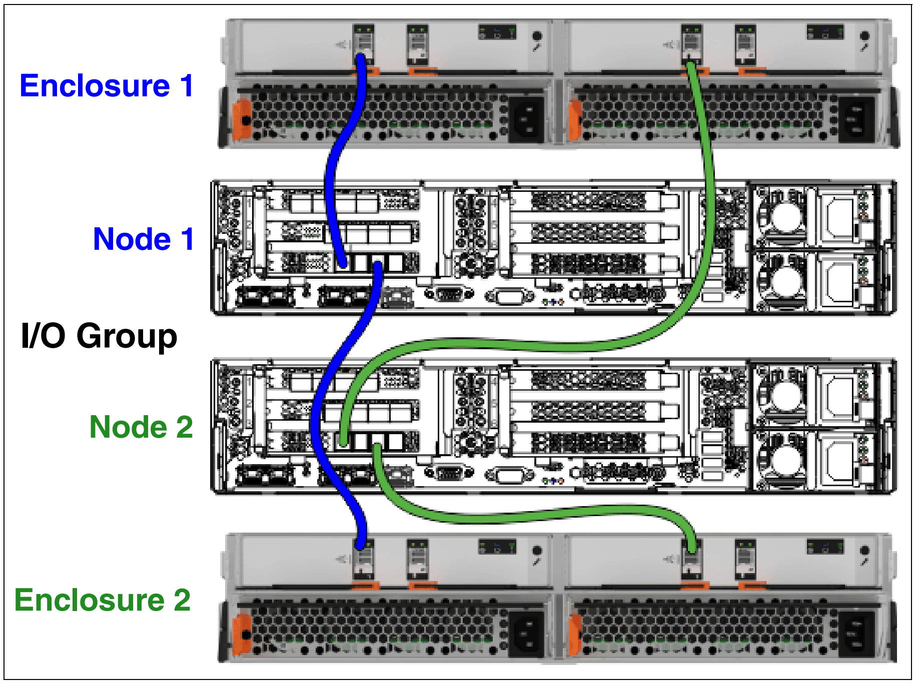


Figure 3-9 Attaching expansion enclosures to SVC nodes

Only ports 1 and 3 of the 12 Gbps SAS card can be used to attach expansion enclosures of flash drives. Expansion enclosures 2145-24F are physically identical to the IBM Storwize V7000 Gen2 (2076-524) expansion enclosures (2076-24F), but have a different product feature code. These enclosures are not interchangeable.

The recommended deployment of SVC devices in the most complex environments (fully equipped systems with eight nodes in multiple racks) differs from the single I/O group installation. This is due to the cabling requirements, serviceability, and maintainability of the system in the enterprise-class data centers. For details, refer to 4.2.1, “Single rack or dual rack configuration” on page 60.

3.1.7 Battery backup system

In order to retain space requirements for the IBM SAN Volume Controller node and its associated components within two rack units and to simplify the cabling and rack layout, the 2145-DH8 node integrates the UPSs inside the node chassis. The SVC has two hot-swappable batteries in the front of the node with the battery backplane at the back of battery drawers. There is no longer any need for the external UPSs for each SVC node. See Figure 3-10 for details.

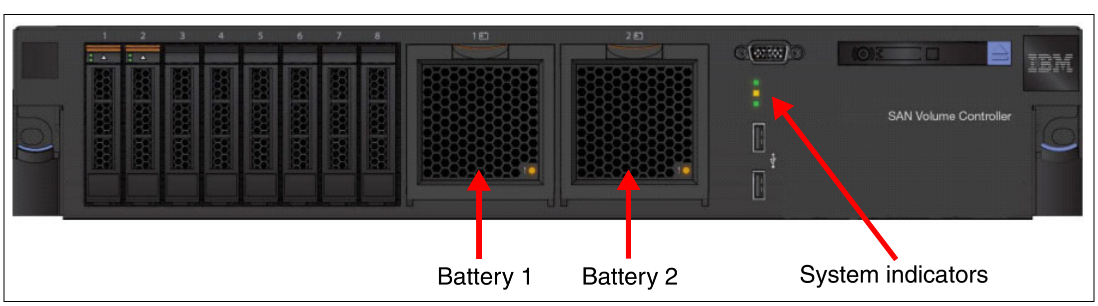


Figure 3-10 Position of batteries in 2145-DH8

The concept of the 2145-DH8 battery unit differs in several significant ways from the batteries used in the previous models of IBM SAN Volume Controller:

- Dual batteries per SVC node.
- They are hot-swappable.
- Designed as redundant within a node.
- Batteries incorporate a test load capability.
- Each battery has its own fault LED indicator.

The 2145-DH8 is designed for two batteries, but continues to operate on a single battery. To achieve maximum redundancy and to get the full life rating of the cells, the system needs to run with both batteries. Running with a single battery results in almost a full discharge and places a higher discharge current on the cells, which leads to a reduced capacity after a number of cycles. Running with just one battery is a degraded state and a node error event is logged to ensure the missing or failed battery is replaced.

An SVC 2145-DH8 node is able to continue operation with one failed battery, although after an AC power failure the node might have to wait for the battery to charge before resuming host I/O operation.

The operational status of batteries and their Vital Product Data are available from the IBM SAN Volume Controller CLI using command **sainfo lsservicestatus**, as shown in Example 3-2.

Example 3-2 Checking the battery status from CLI

```
IBM_2145:ITSO_SVC2:superuser>sainfo lsservicestatus
Battery_count 2
Battery_id 1
Battery_status active
Battery_FRU_part 00AR260
Battery_part_identity 11S00AR056YM30BG43J0C8
Battery_fault_led off
Battery_charging_status idle
Battery_cycle_count 3
Battery_power_on_hours 298
Battery_last_recondition 140512140814
Battery_id 2
Battery_status active
Battery_FRU_part 00AR260
Battery_part_identity 11S00AR056YM30BG43J0BA
Battery_fault_led off
Battery_charging_status idle
Battery_cycle_count 2
Battery_power_on_hours 298
Battery_last_recondition
```

3.2 Software enhancements

In this section, we introduce software enhancements and changes in IBM SAN Volume Controller V7.3. They include:

- Cache rearchitecture
- Easy Tier version 3
- IBM Real-time Compression
- FlashCopy improvement

3.2.1 Cache

SVC 7.3 introduces a major upgrade to the cache code, and in association the 2145-DH8 provides an additional cache capacity upgrade. A base SVC node configuration includes 28 GB of memory for cache. Adding the second processor and cache upgrade for Real-time Compression (RtC) takes a single node to a total of 56 GB of cache. A single I/O Group with support for RtC contains 112 GB of cache, while an 8 node SVC system with a maximum cache configuration contains a total of 448 GB of cache.

In each I/O Group with RtC enabled, 72 GB of Cache is allocated to RtC, 24 GB of cache is allocated to write cache, and 28 GB allocated to read cache.

As in previous versions of SVC, the write cache is partitioned to ensure that in the event of a back-end storage pool under-performing that no impact is introduced to other storage pools managed by SVC. The write cache partitioning is shown in Table 3-4.

Table 3-4 Write cache partitioning

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

No single partition can occupy more than its upper limit of cache capacity with write data. These limits are upper limits, and they are the points at which the SAN Volume Controller cache will start to limit incoming I/O rates for volumes that are created from within that storage pool. If a particular partition reaches this upper limit, the net result is the same as a global cache resource that is full. That is, the host writes will be serviced on a one-out-one-in basis because the cache destage writes to the back-end disks.

Only writes that are targeted at the full partition are limited. All I/O activities destined for other (non-limited) storage pools continue as normal. The read I/O requests for the limited partition also continue normally. However, because the SAN Volume Controller is destaging write data at a rate that is obviously greater than the controller can sustain (otherwise the partition does not reach the upper limit), read response times are also likely affected, but crucially they will continue.

As part of the cache redesign, the cache has been separated from a single layer cache, into a dual layer cache. This flexible plug-and-play style algorithm allows for additional enhancements in the future while delivering SVC like layer 2 cache for advanced functionality.

Figure 3-11 shows the separation of the cache layers.

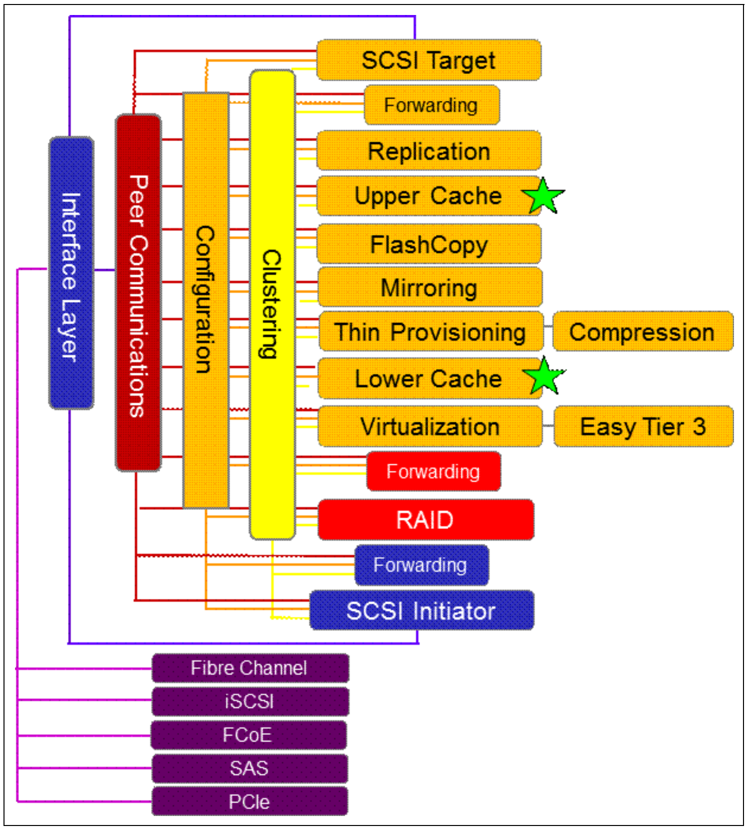


Figure 3-11 Dual layer cache

These changes in the cache architecture have the following benefits:

- Cache enhancements:
  - Performance boost - improved under colliding I/O scenarios
  - Torn Write Protection (atomic writes) up to 16 KB any alignment or 32 KB for 32 KB alignment
  - Improved cache partitioning with destage feedback based on MDisk rather than pool
  - Improved I/O throughput and caching statistics
  - Ability to enable read cache only
  - Ability to change preferred node without changing I/O Group
  - Provide statistics for IBM Tivoli Storage Productivity Center
- Compression enhancements:
  - Main cache stores compressed data (more effective cache capacity)
  - Full stride writes for compressed volumes
  - Support larger number of compressed volumes
- FlashCopy enhancements:
  - Near instant prepare (versus minutes) same for Global Mirror with Change Volumes
  - Multiple snapshots of golden image now share cache data (instead of N copies)
  - Full stride write for FlashCopy volumes no matter what grain size
- Global Mirror with Change Volumes enhancements:
  - Almost zero impact; I/O pause when triggering next change volume (due to near instant prepare)

  - Less impact to source volumes while prepare, as prepare bound by normal destage, not a forced flush
- Stretched System enhancements:
  - Reduced link bandwidth requirements (up to 50% reduced write bandwidth)
- Volume Mirroring enhancements:
  - Support for different tiers in the mirror/not slowed to slowest tier on writes
  - Full stride write for VDM no matter what grain size
- Thin Provisioning enhancements:
  - Full stride writes for Thin volumes no matter what grain size

Additional details are discussed in Chapter 4, “Planning and configuration” on page 57.

3.2.2 Easy Tier

Easy Tier is a built-in dynamic data relocation feature that allows host-transparent movement of data among the storage system resources. This feature significantly improves configuration flexibility and performance tuning and planning.

The *first generation of Easy Tier* introduced automated storage performance management by efficiently boosting Enterprise-class performance (Tier 1), or Capacity Class (Tier 2) with flash drives (SSDs - Tier 0) and automating storage tiering from Tier 1 or 2 drives to Tier 0, thus optimizing flash deployments with minimal costs.

The *second generation of Easy Tier* was only delivered on the IBM System Storage DS8000® product.

The *third generation of Easy Tier* introduces further enhancements that provided automated storage performance and storage economics management across all three drive tiers (Flash, Enterprise, and Near-line storage tiers), which allows you to consolidate and efficiently manage more workloads on a single SVC system. It also introduces support for *Storage Pool Balancing* in homogeneous pools.

More details about Easy Tier version 3 are explained in Chapter 5, “SAN Volume Controller Easy Tier” on page 105.

3.2.3 IBM Real-time Compression

Compressed volumes are a special type of volume where data is compressed as it is written to disk, saving additional space. To use the compression function, you must obtain the IBM Real-time Compression license, and the hardware level for both nodes within the I/O group must be either SAN Volume Controller 2145-DH8, 2145-CG8, or 2145-CF8 for that I/O group to support compression. SAN Volume Controller 2145-DH8 nodes must have two processors, 64 GB of memory, and at least one *Compression Accelerator* card installed in order to use compression.

The Compression Accelerator cards and their placement in the SVC node are introduced in 3.1.4, “Compression Accelerator card” on page 44. The concept of hardware assistance when using IBM Real-time Compression in SVC is described in Chapter 7, “Real-time Compression and the SVC 2145-DH8 Node” on page 139.

3.2.4 FlashCopy

The IBM SAN Volume Controller V7.3 and its cache rearchitecture improves the processing of FlashCopy operations. Its positive impact is introduced in 3.2.1, “Cache” on page 50.

3.3 Upgrade considerations

The 2145-DH8 nodes can be integrated within the existing IBM SAN Volume Controller clustered systems with only a few additional steps in regard of the new worldwide names (WWNs) structure. The nodes can be intermixed in pairs in the existing SVC systems.

Important: When installing 2145-DH8 nodes into the existing SVC environment with compressed volumes, all DH8 nodes must have available the second processor, 64 GB memory, and at least one Compression Accelerator card.

Mixing node types in a system results in volume performance characteristics that depend on the node type in the volume’s I/O Group. The nondisruptive clustered system upgrade process can be used to replace older engines with new 2145-DH8 engines.

The prerequisites for the nondisruptive upgrade include:

- All nodes that are configured in the system are present and online.
- All errors in the system event log are addressed and marked as fixed.
- There are no volumes, managed disks (MDisks), or external storage systems with a status of degraded or offline.
- The replacement node is not powered on and not connected to the SAN (or not zoned).
- You have backed up the system configuration and saved the `svc.config.backup.xml` file.
- The replacement node must be able to operate at the Fibre Channel or Ethernet connection speed of the node it is replacing.
- If the node being replaced contains flash drives, transfer all flash drives and SAS adapters to the new node if it supports the drives. To prevent losing access to the data, if the new node does not support the existing flash drives, transfer the data from them before replacing the node. Note that flash drives from 2145-CG8 and older nodes cannot be transferred to the 2145-DH8 or to the 2145-24F expansion enclosure.
- If the node being replaced is an IBM SAN Volume Controller 2145-CG8, CF8, or 8A4 and the replacement node is a 2145-DH8, the replacement node must have a four port Fibre Channel card in slot one. If the node being replaced has a second I/O card in addition to the required FC card, the replacement node must have the same card in slot two.

Important: If you plan to reuse the node that you are replacing, ensure that the worldwide node name (WWNN) of the node is set to a unique number on your SAN. If you do not ensure that the WWNN is unique, the WWNN and WWPN are duplicated in the SAN environment and could cause problems.

One of the important considerations when upgrading the system to DH8 nodes or when just installing an additional I/O group based on DH8 nodes, is the use of WWPN Range. The SVC V7.3 platform has too many ports to use the WWPN Range provided by the pre-existing WWNN seed. As such, the DH8 has a different WWNN seed. However, as a consequence of this, the DH8 contains the code function to handle the hardware upgrade from existing nodes in the data centers to the DH8. As a consequence, the DH8 allows the user to set the WWNN seed for the WWPN End Points. As the WWNN seed for old SVCs only supports 59 ports, this does not work on a DH8, which can have up to 16 FC Ports at a 6 x WWPNs per port. However, only one of those is the visible endpoint so the proposal is to use the user provided WWNN as the seed for the visible I/O End points, and use the originally programmed DH8 WWNN seed (up to 128 ports) for the virtual switch port WWPNs.

The procedure of upgrade is nondisruptive because changes to your SAN environment are not required. The replacement (new) node uses the same WWNN as the node that you are replacing. An alternative to this procedure is to replace nodes disruptively either by moving volumes to a new I/O group or by rezoning the SAN. The disruptive procedures, however, will require additional work on the hosts.

3.3.1 Worldwide names scheme

The IBM SAN Volume Controller 2145-DH8 uses the new *80c* product ID, so IBM has the opportunity to define a new scheme to generate WWNs. Public WWNs take the form:

```
500507680c <slot number> <port number> xxxx
```

With four bits for slot number and four for port number, giving 16 public names per slot, and 16 bits for the serial number. The public port names for 2145-DH8 slot 1 under the new scheme would look like that outlined in Table 3-5.

Table 3-5 WWPN scheme

Port	WWPN
1	500507680c**11**xxxx
2	500507680c**12**xxxx
3	500507680c**13**xxxx
.....	
14	500507680c**1e**xxxx
15	500507680c**1f**xxxx
16	500507680c**10**xxxx

Other slots follow the same pattern but with their own slot number in place of the **1**. Private names for a slot are first taken from its pool of unused public names, and any extra needed can be generated by adding **8** to the slot number (wrapping around to 0 for slot 8) and starting again from port 1. This means the 17th port name for slot 1 would be `500507680c91xxxx`, the 18th `500507680c92xxxx`, and so on up to the 28th, which would be `500507680c9cxxxx`. This allows SVC to support up to eight slots with seven addresses per physical port (assuming four ports per card), which is enough to cover FC direct-attach with a spare address per port.

There are two classes of WWNs used in 2145-DH8 nodes:

- **Public names**, which need to be migrated during hardware upgrade:
  - Used by hosts to identify storage
  - Used by backend controllers for LUN masking
  - Needed for fabric zoning
- **Private names**, which can change during hardware upgrade:
  - Used for the various “fake switch” components for FC direct-attach
  - Used for SAS initiator in 2145-24F expansion unit

Ideally, SVC would like to be able to generate all publicly addressable WWNs for any possible 2145-DH8 configuration from an old *801* seed. This means that users who upgraded an older node to a DH8 and then upgraded their DH8 to a newer future model would only have to give the newer system the original 801 seed from their old node, rather than both the 801 and the 80c.

There are seven bits to work with in the 801 address space (bits 17-23, bit 16 was given to manufacturing), which gives us 119 possible WWNs when iSCSI reserved addresses and the node name are taken out. The 2145-DH8 is currently limited to four I/O cards but IBM might eventually allow users the option of using all six slots for I/O, which would give us a maximum of (119/6) = 19 WWNs per slot.

The current “worst case” number of public WWNs needed for a single slot is 16 (four ports, each supporting up to four FCFs). However, as the most common configurations will likely only need a maximum of four public names per slot SVC provides sensible addresses for those ports. Additional ports beyond the first four can be generated by adding two to the second digit. The full port naming scheme for a node with six cards each supporting up to four FCFs per port is illustrated in Figure 3-12.

Port	WWPN					
	Slot 1	Slot 2	Slot 3	Slot 4	Slot 5	Slot 6
1	500507680140xxxx	500507680150xxxx	500507680190xxxx	500507680118xxxx	500507680158xxxx	500507680198xxxx
2	500507680130xxxx	500507680160xxxx	5005076801a0xxxx	500507680128xxxx	500507680168xxxx	5005076801a8xxxx
3	500507680110xxxx	500507680170xxxx	5005076801b0xxxx	500507680138xxxx	500507680178xxxx	5005076801b8xxxx
4	500507680120xxxx	500507680180xxxx	5005076801c0xxxx	500507680148xxxx	500507680188xxxx	5005076801c8xxxx
5	500507680142xxxx	500507680152xxxx	500507680192xxxx	50050768011axxxx	50050768015axxxx	50050768019axxxx
6	500507680132xxxx	500507680162xxxx	5005076801a2xxxx	50050768012axxxx	50050768016axxxx	5005076801aaxxxx
7	500507680112xxxx	500507680172xxxx	5005076801b2xxxx	50050768013axxxx	50050768017axxxx	5005076801baxxxx
8	500507680122xxxx	500507680182xxxx	5005076801c2xxxx	50050768014axxxx	50050768018axxxx	5005076801caxxxx
9	500507680144xxxx	500507680154xxxx	500507680194xxxx	50050768011cxxxx	50050768015cxxxx	50050768019cxxxx
10	500507680134xxxx	500507680164xxxx	5005076801a4xxxx	50050768012cxxxx	50050768016cxxxx	5005076801acxxxx
11	500507680114xxxx	500507680174xxxx	5005076801b4xxxx	50050768013cxxxx	50050768017cxxxx	5005076801bcxxxx
12	500507680124xxxx	500507680184xxxx	5005076801c4xxxx	50050768014cxxxx	50050768018cxxxx	5005076801ccxxxx
13	500507680146xxxx	500507680156xxxx	500507680196xxxx	50050768011exxxx	50050768015exxxx	50050768019exxxx
14	500507680136xxxx	500507680166xxxx	5005076801a6xxxx	50050768012exxxx	50050768016exxxx	5005076801aexxxx
15	500507680116xxxx	500507680176xxxx	5005076801b6xxxx	50050768013exxxx	50050768017exxxx	5005076801bexxxx
16	500507680126xxxx	500507680186xxxx	5005076801c6xxxx	50050768014exxxx	50050768018exxxx	5005076801cexxxx

Figure 3-12 Mapping of WWPNs to ports

Note that the first slot maintains the 4,3,1,2 order from SVC nodes so that ports retain the same order counting from left to right.

The whole upgrade or replacement procedure in regards of WWN requirements is described in the IBM SAN Volume Controller Knowledge Center, available at the following link:

http://www.ibm.com/support/knowledgecenter/STPVGU/landing/SVC_welcome.html?lang=en

4

Planning and configuration

In this chapter, we describe the steps that are required when you plan the installation of the IBM SAN Volume Controller (SVC), utilizing the new 2145-DH8 nodes in your environment. We look at the implications for your storage network and also discuss performance considerations.

4.1 General planning rules

Important: At the time of writing, the statements that we make are correct, but they might change over time. Always verify any statements that have been made with the SVC supported hardware list, device driver, firmware, and recommended software levels at:

http://www.ibm.com/support/docview.wss?uid=ssg1S1003658

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

Note: Check the pre-sale Technical and Delivery Assessment (TDA) document at this website:

https://www.ibm.com/partnerworld/wps/servlet/mem/ContentHandler/salib_SA572/lc=en_ALL_ZZ

A pre-sale TDA should be conducted before submitting a final proposal to a client and must be conducted before placing an order to ensure the configuration is correct and the solution being proposed is valid. The preinstall System Assurance Planning Review (SAPR) package includes various files that are used in preparation for an SVC preinstall TDA. A preinstall TDA should be conducted shortly after the order is placed and before the equipment arrives at the client location to ensure the client's site is ready for the delivery and the responsibilities are documented regarding the client's and IBM or Business Partner roles in the implementation.

Tip: For comprehensive information about the topics that are discussed here, see *IBM System Storage SAN Volume Controller: Planning Guide*, GA32-0551.

We also go into much more depth about these topics in *SAN Volume Controller Best Practices and Performance Guidelines*, SG24-7521, which is available at:

http://www.redbooks.ibm.com/abstracts/sg247521.html?Open

Follow these steps when planning for the SVC:

1. Collect and document the number of hosts (application servers) to attach to the SVC, the traffic profile activity (read or write, sequential, or random), and the performance requirements (I/O per second (IOPS)).
2. Collect and document the storage requirements and capacities:
   - The total backend storage existing in the environment to be provisioned on SVC.
   - The total backend new storage to be provisioned on the SVC.
   - The required storage capacity for local mirror copy (volume mirroring).
   - The required storage capacity for point-in-time copy (FlashCopy).
   - The required storage capacity for remote copy (Metro Mirror and Global Mirror).
   - The required storage capacity for compressed volumes.
   - Per host: Storage capacity, the host logical unit number (LUN) quantity, and sizes.
   - The required virtual storage capacity that is used as a fully managed volume and used as a thin-provisioned volume.

3. Define the local and remote SAN fabrics and clustered systems if a remote copy or a secondary site is needed.
4. Define the number of clustered systems and the number of pairs of nodes (between one and four) for each system. Each pair of nodes (an I/O Group) is the container for the volume. The number of necessary I/O Groups depends on the overall performance requirements.
5. Design the SAN according to the requirement for high availability and best performance. Consider the total number of ports and the bandwidth that is needed between the host and the SVC, the SVC and the disk subsystem, between the SVC nodes, and for the inter-switch link (ISL) between the local and remote fabric.

Note: Check and carefully count the required ports for extended links. Especially in a Stretched System environment, you might need many longwave gigabit interface converters (GBICs).

6. Design the iSCSI network according to the requirements for high availability and best performance. Consider the total number of ports and bandwidth that is needed between the host and the SVC.
7. Determine the SVC service IP address.
8. Determine the IP addresses for the SVC system and for the host that connects through iSCSI.
9. Determine the IP addresses for IP replication.
10. Define a naming convention for the SVC nodes, host, and storage subsystem.
11. Define the managed disks (MDisks) in the disk subsystem.
12. Define the storage pools. The storage pools depend on the disk subsystem in place and the data migration requirements.
13. Plan the logical configuration of the volume within the I/O Groups and the storage pools to optimize the I/O load between the hosts and the SVC.
14. Plan for the physical location of the equipment in the rack.

SVC planning can be categorized into two types:

- Physical planning
- Logical planning

We describe these planning types in more detail in the following sections.

4.2 Physical planning

This section of the document is intended to provide guidance on the cabinet elevation layouts to use for physically installing your SVCs in racks.

One of the first factors to consider is whether you are building a brand new SVC cluster with only 2145-DH8(s) in or if you are adding the 2145-DH8 to an existing cluster having older model SVC nodes in it. A second factor is, if a brand new SVC cluster, you need to determine if you will be racking your SVCs in a single cabinet layout or a dual cabinet layout. Additionally, when using the optional 2145-24F flash arrays as part of your SVC cluster implementation, the distance you can separate the 2145-DH8 nodes in the I/O Group away from their shared 2145-24F flash array is limited by the maximum length of the 6-meter SAS cable used to attach the array to the SVC units.

- You must consider the maximum power rating of the rack; *do not exceed it*. For more information about the power requirements, see the following IBM Knowledge Center page:

  http://pic.dhe.ibm.com/infocenter/svc/ic/index.jsp?topic=%2Fcom.ibm.storage.svc.console.doc%2Fsvc_preparephysicalenvcg8_1a1733.html

4.2.1 Single rack or dual rack configuration

The SVC system must be installed in pairs to provide high availability. Each pair makes up an I/O Group. An SVC cluster can contain up to 4 I/O Groups, or a total of 8 nodes. Each node requires two IEC-C13 power cable connections to connect to the 750w power supplies.

Figure 4-1 shows a 2145-DH8 node with two 750W power supplies.

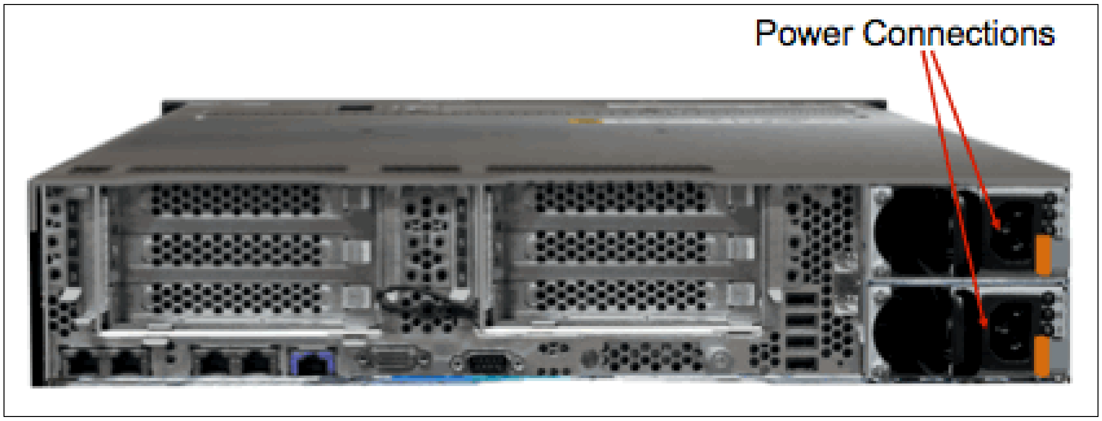


Figure 4-1 2145-DH8 power connectors

Choosing whether to install all eight nodes of an SVC cluster in a single cabinet or to split them across two separate cabinets (A Nodes in one, and B Nodes in another) requires consideration of the overall storage environment you are implementing, as well as the facilities resiliency of the data centers you are installing your storage environment into. These guidelines are for a system installed in a single machine room.

You can also further increase availability by splitting the system between rooms (Stretched System), but that requires careful design of the data system as a whole, including host configuration, storage system configuration, fabric design, and SVC configuration. For more details about SVC Stretched System, see Chapter 11, “SAN Volume Controller Stretched System” on page 345.

Upstream redundancy of the power to your cabinet (power circuit panels and on-floor PDUs) and within cabinet power redundancy (dual power strips or in-cabinet PDUs) as well as upstream high availability structures (UPS, generators, and so on) will feed into your decision. Many data centers today are at an Uptime Tier 3 or higher level so power redundancy concerns that would require a dual cabinet SVC implementation are no longer an issue.

Fire Protection Systems Type (such as overhead wet pipe sprinkler systems), however, should be taken into consideration. In association to these items you should also consider physical separation and location of other key storage environment components.

If you are implementing a single SVC cluster attached to a single back-end storage device or string (XIV frame or DS8000 string), separating your SVC cluster across two cabinets is not any more advantageous to your overall storage environment than installing it in a single cabinet in case of a facilities problem such as breakage of an overhead wet-pipe sprinkler head.

If you have your fabric switching pair that all your front-end devices (servers, and so on) and your back-end devices (SVCs, disk stores, VTLs, and so on) access installed together in the same cabinet, again separating your SVC cluster across two cabinets is not any more advantageous to your overall storage environment than installing it in a single cabinet in case of a facilities problem such as breakage of an overhead wet-pipe sprinkler head.

If you are implementing your entire storage environment with multiple redundant devices physically separated across multiple cabinets and strings you need to ensure sufficient enough physical distance to ensure that your redundant components are in different fire protection zones or power sourced zones. Otherwise, your end to end storage environment can be compromised in case of a zonal facilities failure.

If the data center you are moving into has the proper power redundancy attributes and your storage environment design strategy does not have fully redundant components placed at sufficient distances apart, a single cabinet implementation will save you the costs associated with the second cabinet.

If the data center does not have a robust enough power redundancy infrastructure or your storage environment design strategy does not have fully redundant components placed at sufficient distances apart, the investment in a dual cabinet implementation is certainly justified in furthering the level of high availability/redundancy of your overall storage environment.

Another consideration would be that if you anticipate that you will be adding another SVC cluster to your storage environment in the future by implementing a dual cabinet approach from the start, and reserving remaining space in each cabinet for nodes from the second cluster, you accomplish both objectives.

Figure 4-2 shows a single cabinet implementation of a new 2145-DH8 SVC node implementation that does not involve the optional 2145-24F SSD arrays. The SVC nodes can be installed starting in any RU location in the cabinet depending upon other existing or planned equipment but must be racked in the groupings as shown. Also, note that space reservations have been assumed at the top of the cabinet for structured data wiring, which is recommended also as a data center best practice.

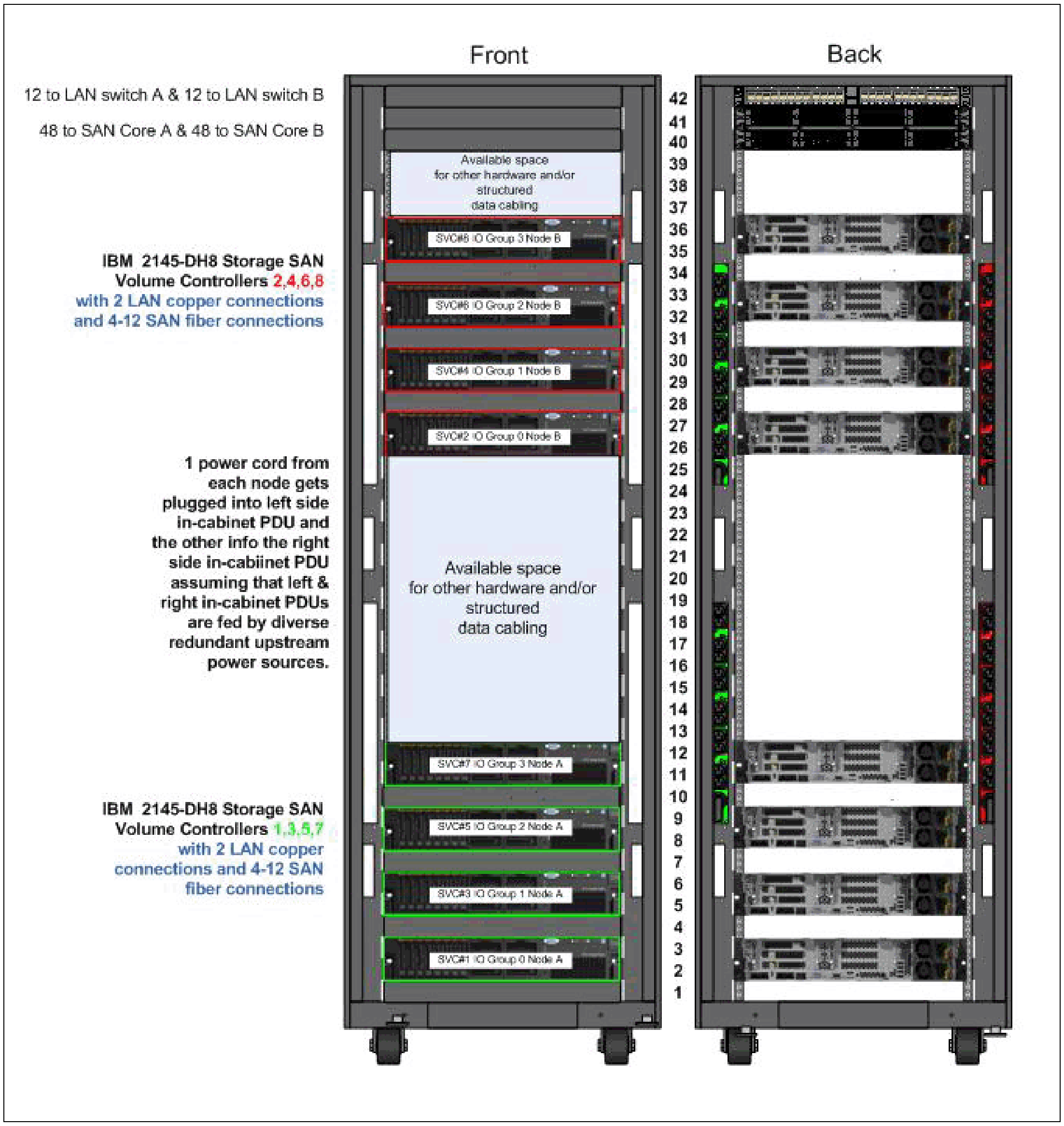


Figure 4-2 Single cabinet implementation of 2145-DH8 SVC cluster

Figure 4-3 shows a single cabinet implementation of a new 2145-DH8 SVC node that does include a first set of the optional 2145-24F flash arrays. In this depiction, the 2145-24F flash arrays have been consolidated in the middle of the rack to provide some separation between the odd and even-numbered 2145-DH8 SVC nodes. Also of note is the reservation left at the top of the rack for an optional second set of 2145-24F arrays should they be needed.

Also, you must ensure that you run the SAS cables within the cable management arms in the back of the 2145-DH8 SVC nodes. You will want to make sure you use feature code ACUC: 3 m 12 Gb SAS Cable (mSAS HD to mSAS HD) to attach the 2145-24F to 2145-DH8. Also, note that space reservations have been assumed at the top of the cabinet for structured data wiring, which is recommended also as a data center best practice.

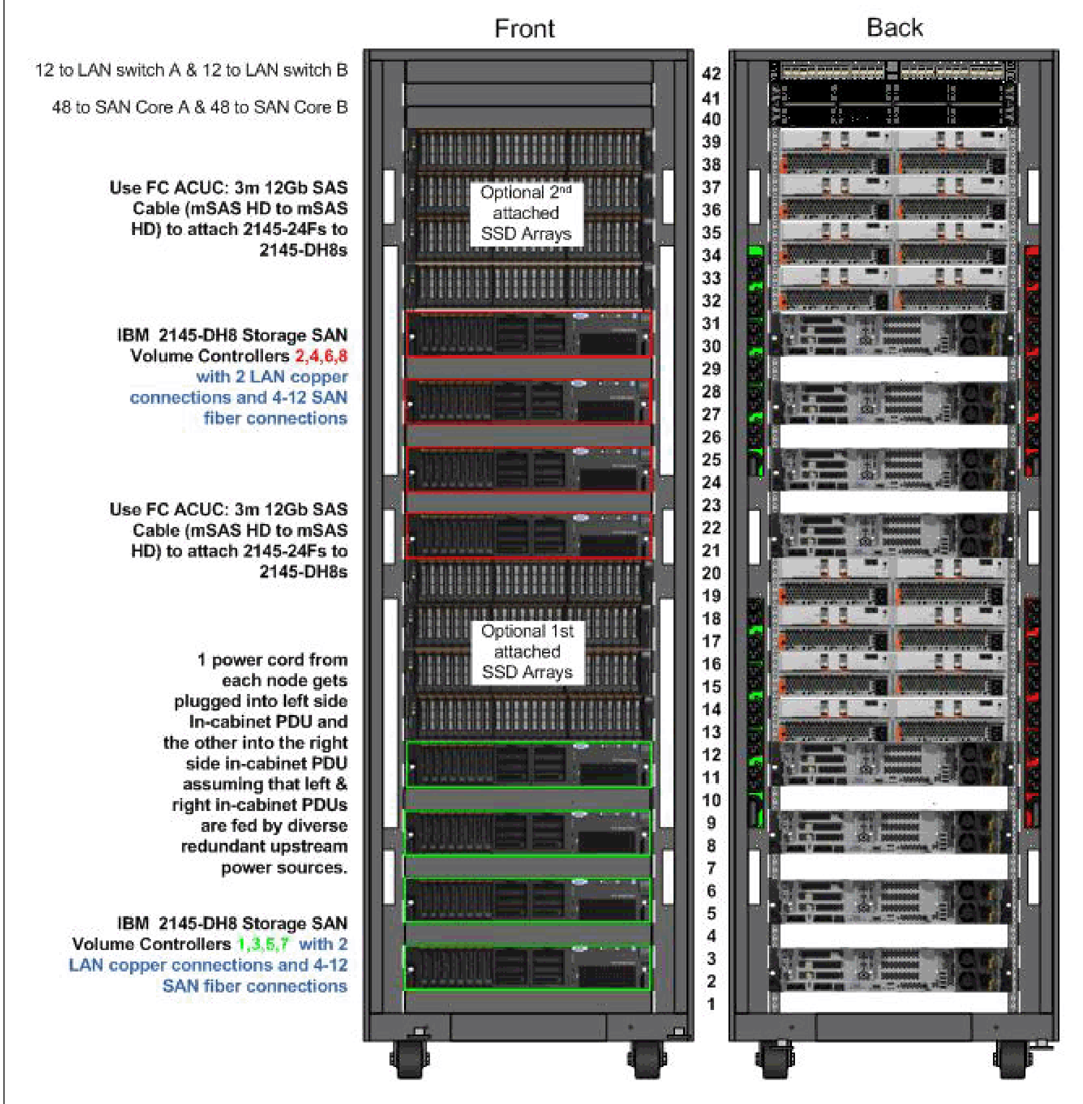


Figure 4-3 Single cabinet implementation without expansion enclosures

Figure 4-4 shows a single cabinet implementation of 2145-DH8 SVC node that does include a first set of the optional 2145-24F flash arrays. In this depiction, the 2145-24F enclosures have been intermingled between their corresponding 2145-DH8 SVC I/O Group node pairs.

This iteration is less desirable than the consolidated iteration as it requires that the first DH8 node be installed starting in 1U in the cabinet and will leave 2U gaps in the cabinet if the optional second 2145-24F group is never acquired.

Also, you must ensure that you run the SAS cables within the cable management arms in the back of the 2145-DH8 SVC nodes. You will want to make sure you use feature code ACUC: 3 m 12 Gb SAS Cable (mSAS HD to mSAS HD) to attach the 2145-24F to 2145-DH8. Also, note that space reservations have been assumed at the top of the cabinet for structured data wiring, which is recommended also as a data center best practice.

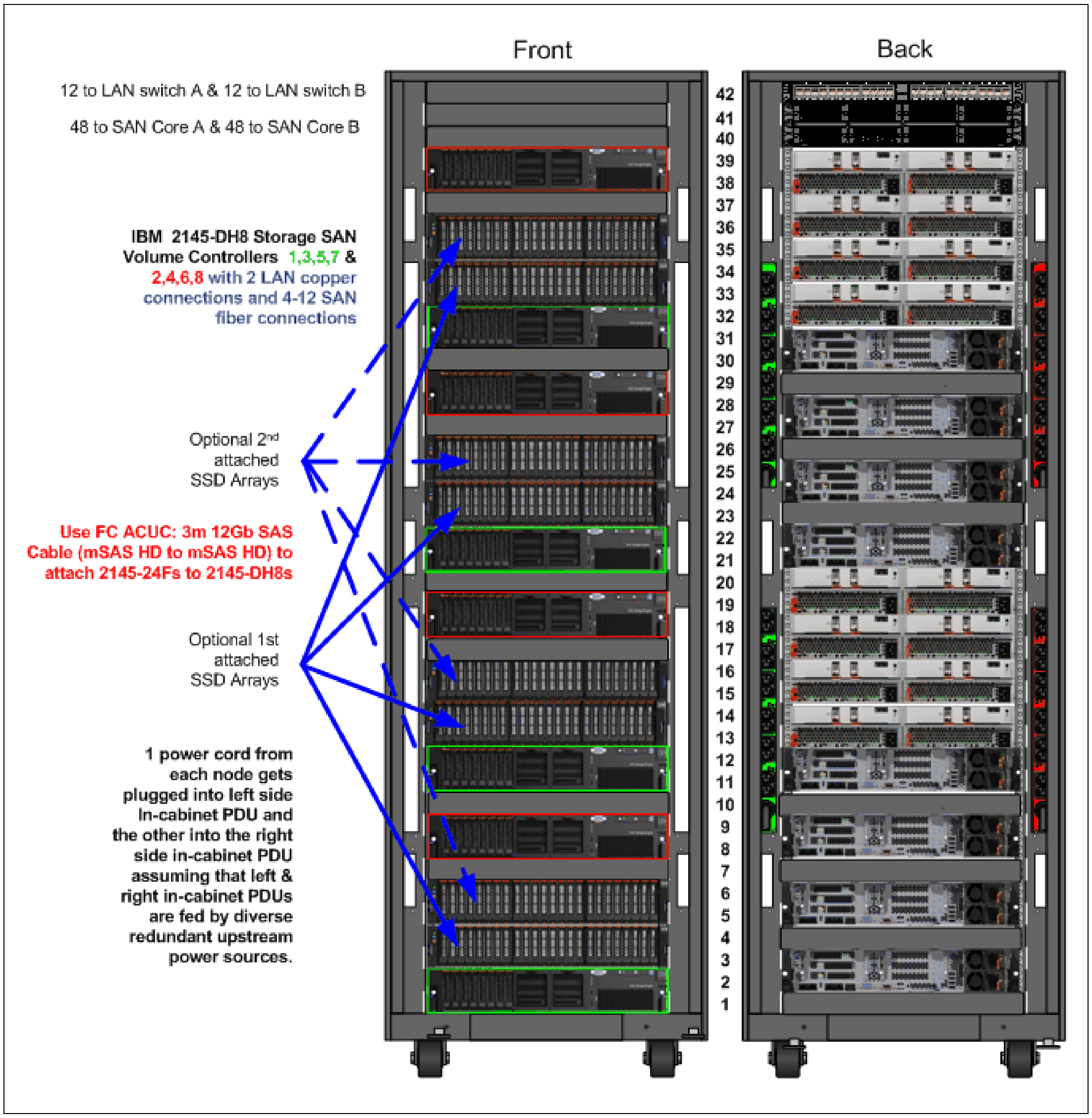


Figure 4-4 Single rack with 8 node 2145-DH8 cluster including the optional 2145-24F intermixed with the 2145-DH8 nodes

Figure 4-5 shows a single cabinet implementation of a new 2145-DH8 SVC node that does include a first set of the optional 2145-24F flash arrays.

In this depiction, the 2145-24F flash arrays have been intermingled between their corresponding 2145-DH8 SVC I/O Group Node pairs. Space for a second set of 2145-24F flash drives is allocated at the top of the rack instead of being intermingled.

This iteration is also less desirable than the consolidated iteration as it requires that the first DH8 node be installed starting in 1U in the cabinet in order to leave sufficient space at the top of the cabinet if the optional second 2145-24F group is acquired.

Also, you must ensure that you run the SAS cables within the cable management arms in the back of the 2145-DH8 SVC nodes. Make sure you use feature code ACUC: 3 m 12 Gb SAS Cable (mSAS HD to mSAS HD) to attach the 2145-24F to 2145-DH8. Also, note that space reservations have been assumed at the top of the cabinet for structured data wiring, which is recommended also as a data center best practice.

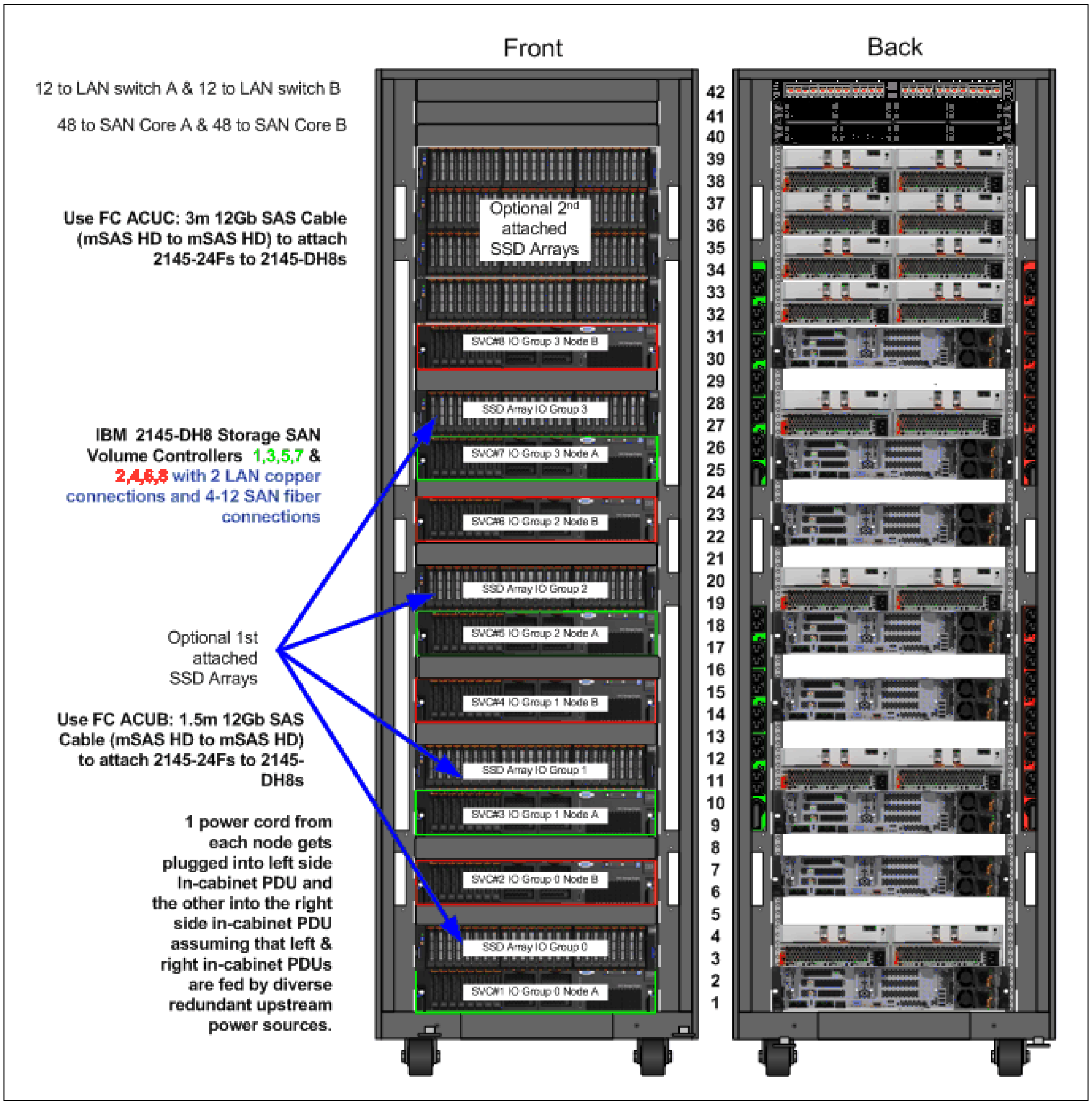


Figure 4-5 Shows a single cabinet implementation of a new 2145-DH8 SVC node with four intermingled 2145-24F expansion shelves and four clustered expansion shelves

Figure 4-6 shows a single cabinet implementation of a mixed cluster using new 2145-DH8 SVC nodes being added to an existing SVC cluster that already has 2145-CG8. As it is assumed that the earliest models of the SVC nodes did not make use of flash arrays the option 2145-24F's are assumed not to be used with these 2145-DH8 nodes added to the existing cluster. Also note that for consistency, space reservations have been assumed at the top of the cabinet for structured data wiring, which is recommended also as a data center best practice.

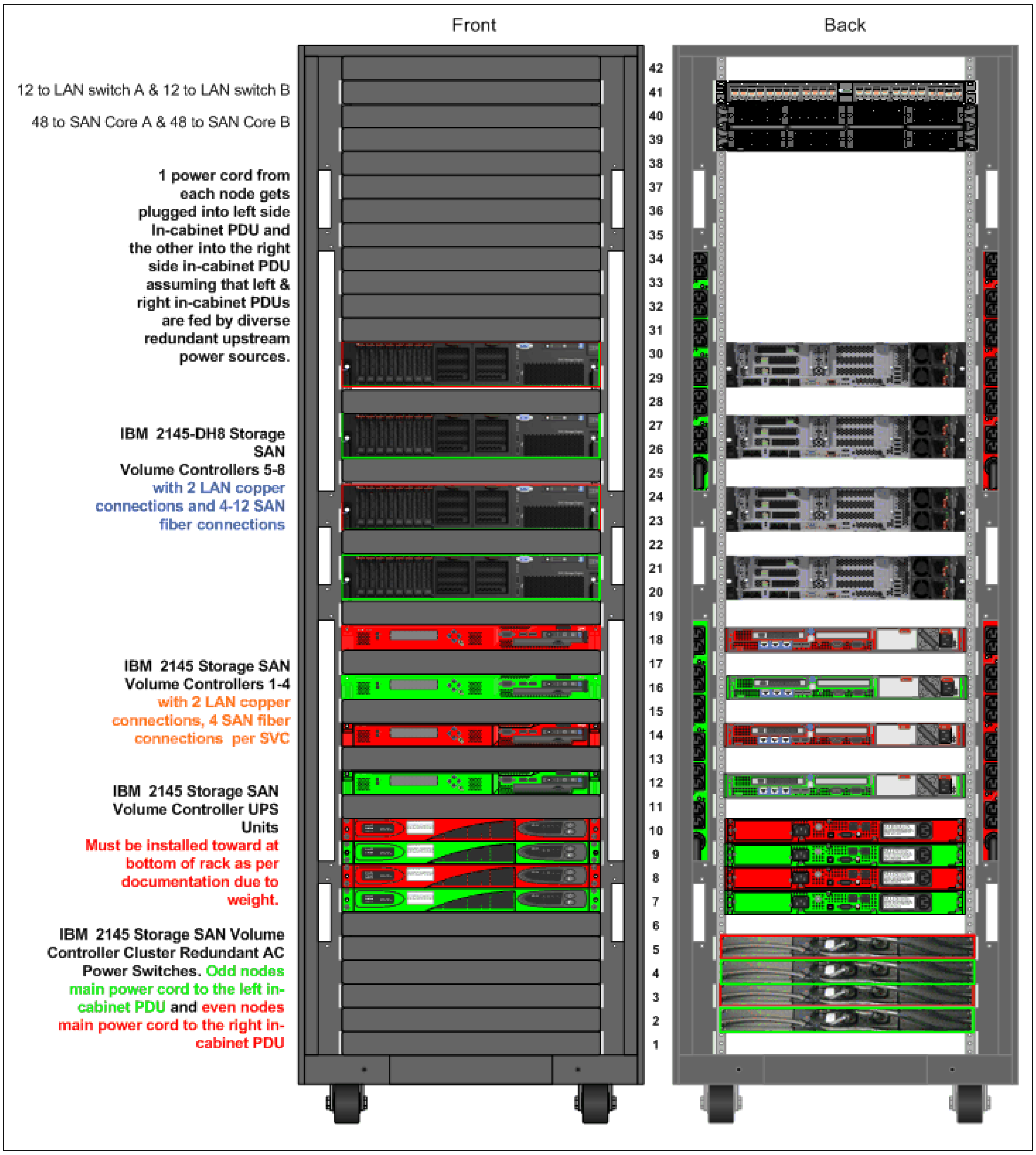


Figure 4-6 Single cabinet implementation of a mixed cluster using both 2145-DH8 and 2145-CG8 nodes

Figure 4-7 shows the front view of a dual cabinet implementation of a new 2145-DH8 SVC node implementation that does not include the optional 2145-24F flash arrays. The SVC nodes can be installed starting in any RU location in the cabinets depending upon other existing or planned equipment, but must be racked in the groupings as shown. Also, note that space reservations have been assumed at the top of the cabinet for structured data wiring, which is recommended also as a data center best practice.

Caution: Cabinets can be located further apart as needed but if there is any chance of the optional 2145-24F flash arrays being added at a later time, these cabinets need to be close enough together for the 6-meter SAS cables to be able to reach.

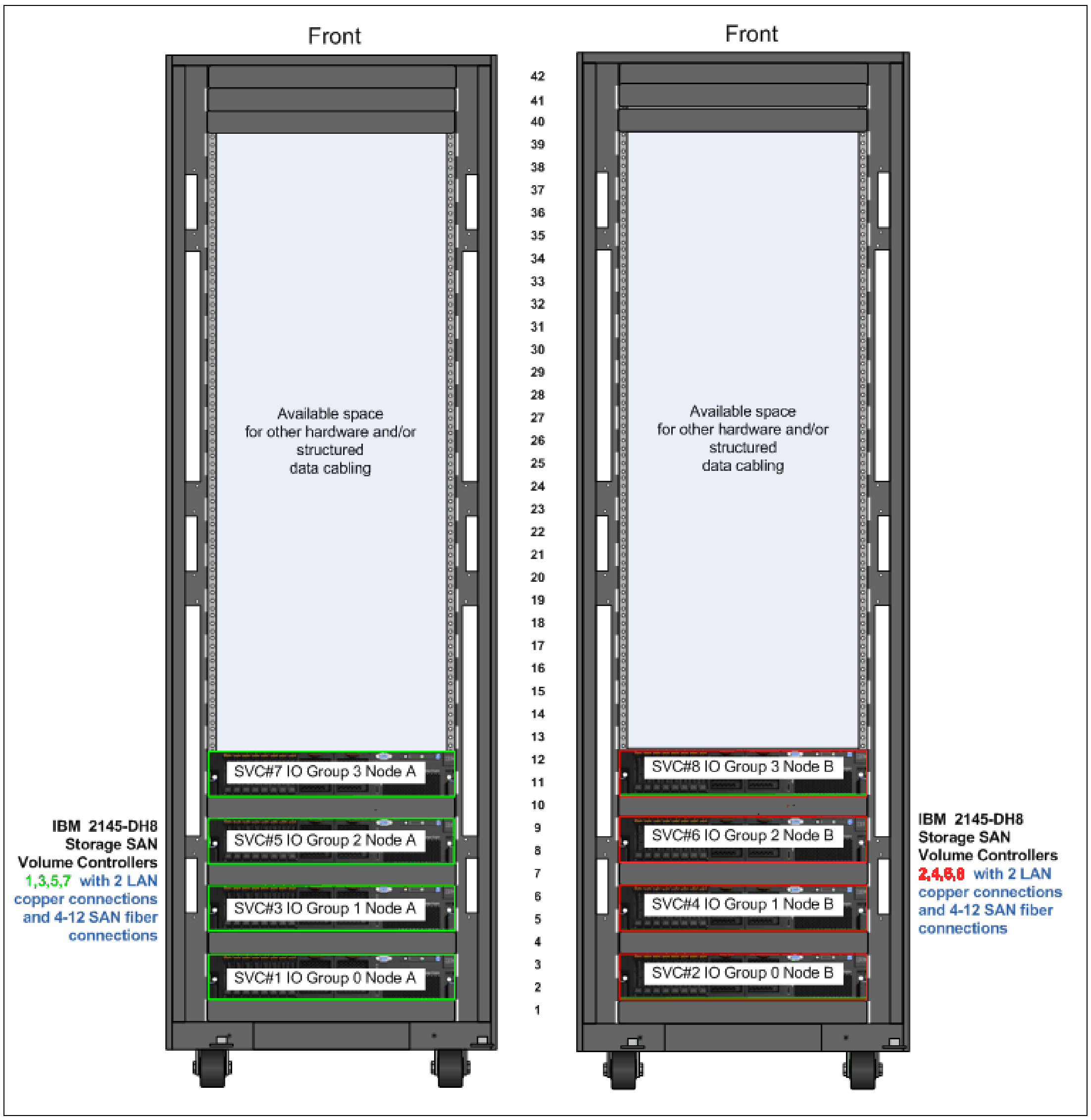


Figure 4-7 Front view dual cabinet implementation of a new 2145-DH8 SVC Cluster

Figure 4-8 shows the rear view of a dual cabinet implementation of a new 2145-DH8 SVC node implementation that does not include the optional 2145-24F flash arrays.

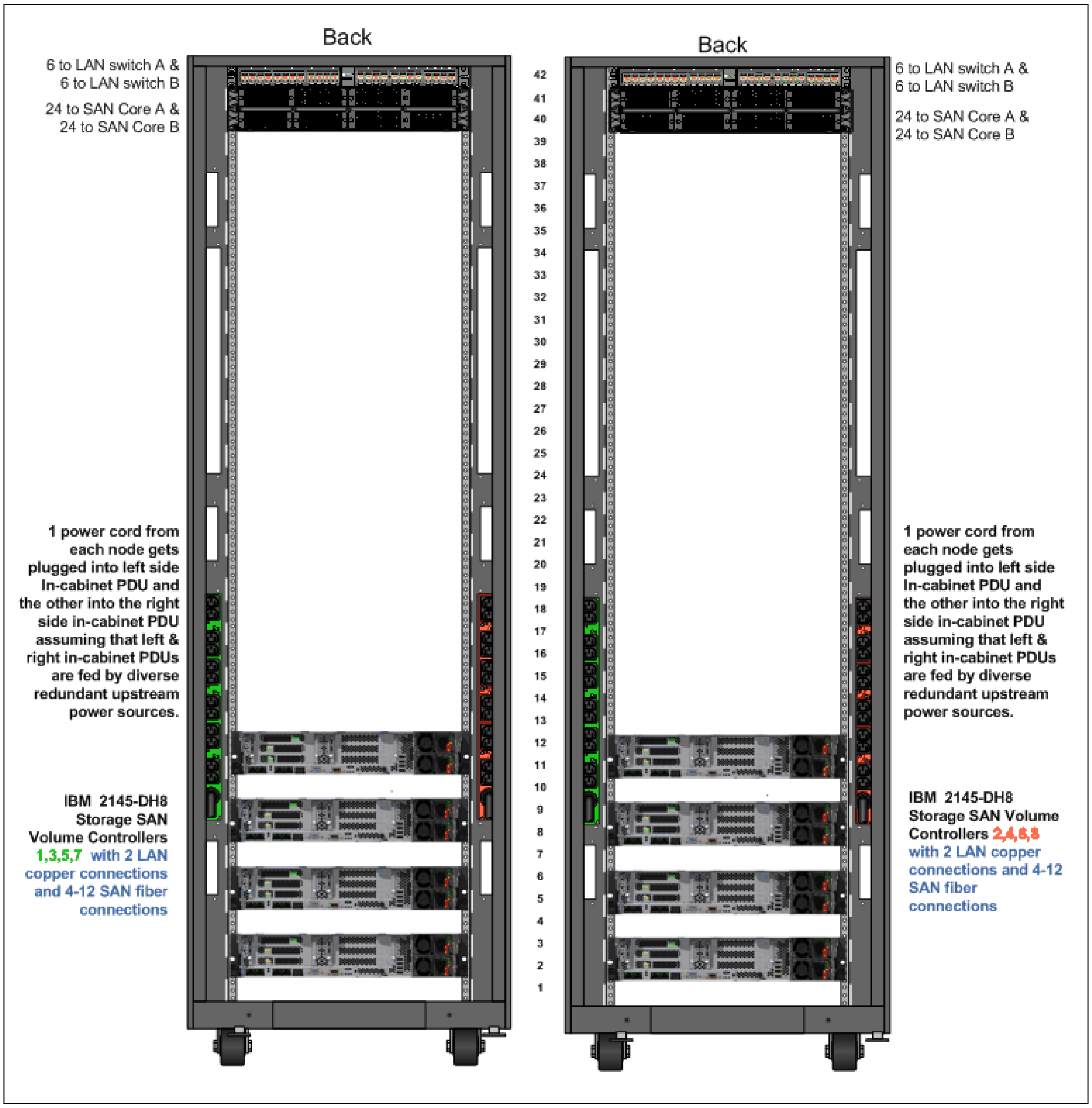


Figure 4-8 Rear view dual cabinet implementation of a new 2145-DH8 SVC Cluster

Figure 4-9 shows a dual cabinet implementation of a new 2145-DH8 SVC node with the optional 2145-24F flash arrays implemented.

Since you need to interconnect the 2145-DH8s to their respective enclosures across a pair of cabinets, the location where you rack the nodes and the arrays is very important.

The very first consideration is whether the cabling between the cabinets will be below the cabinets (and raised floor) or above the cabinets. The templates show a below cabinet/floor implementation. If they need to go above the cabinets, reverse the location of the 2145-24Fs to be at the top of the cabinet, followed by the 2145-DH8 nodes grouped just below them.

Also, note that space reservations have been assumed at the top of the cabinet for structured data wiring, which is recommended also as a data center best practice. Also, you must ensure that you run the SAS cables within the cable management arms in the back of the 2145-DH8 SVC nodes. Therefore, you will want to make sure that you use feature code ACUC: 3 m 12 Gb SAS Cable (mSAS HD to mSAS HD) to attach the 2145-24Fs to 2145-DH8s in the same cabinet, and feature code ACUD: 6 m 12 Gb SAS Cable (mSAS HD to mSAS HD) to attach the 2145-24Fs to 2145-DH8s in the opposite cabinet.

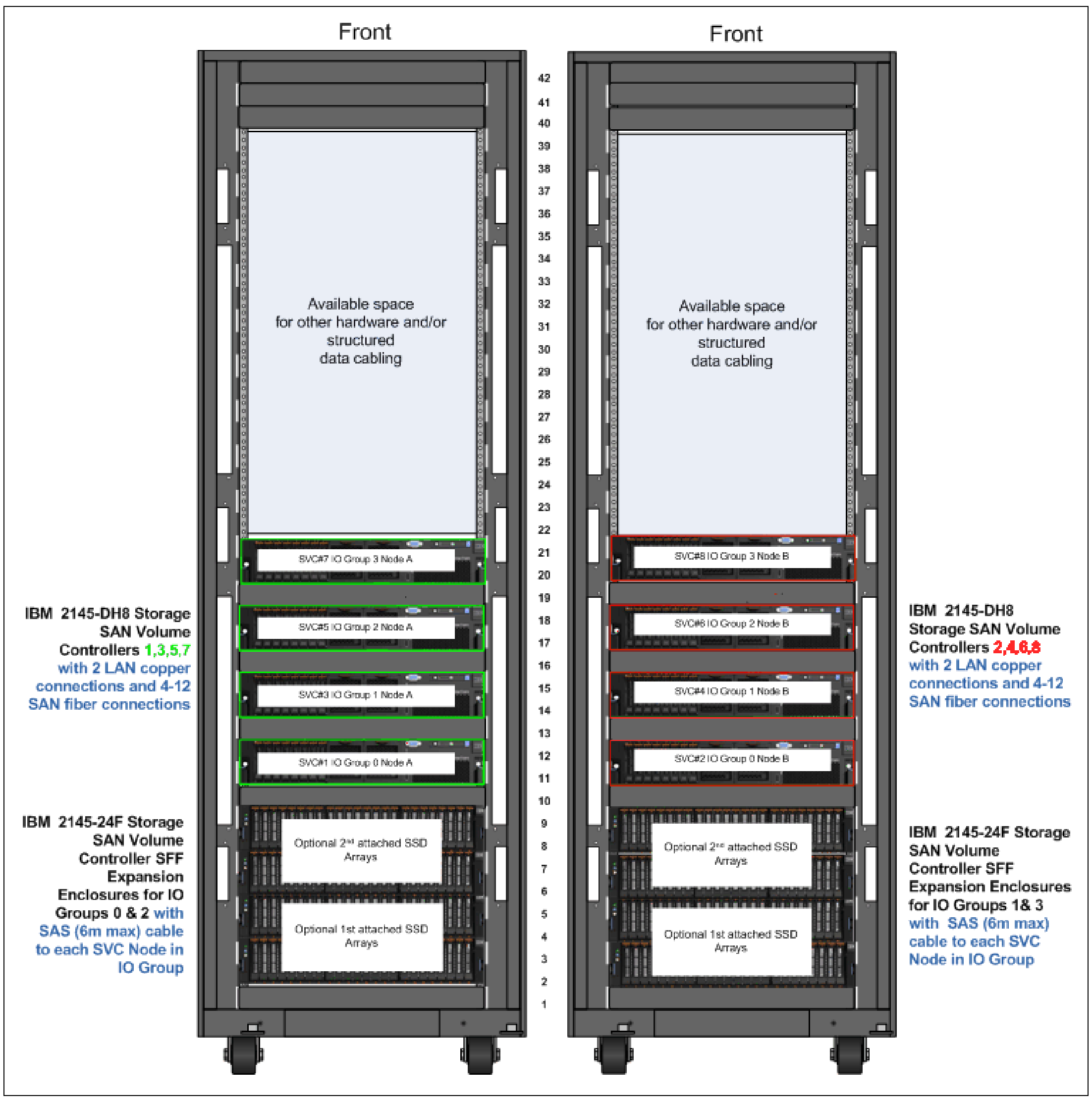


Figure 4-9 Dual cabinet implementation of a new 2145-DH8 SVC node implementation with the optional 2145-24F SSD arrays included

Figure 4-10 shows the rear view of a dual cabinet implementation of a new 2145-DH8 SVC node implementation with the optional 2145-24F flash arrays implemented.

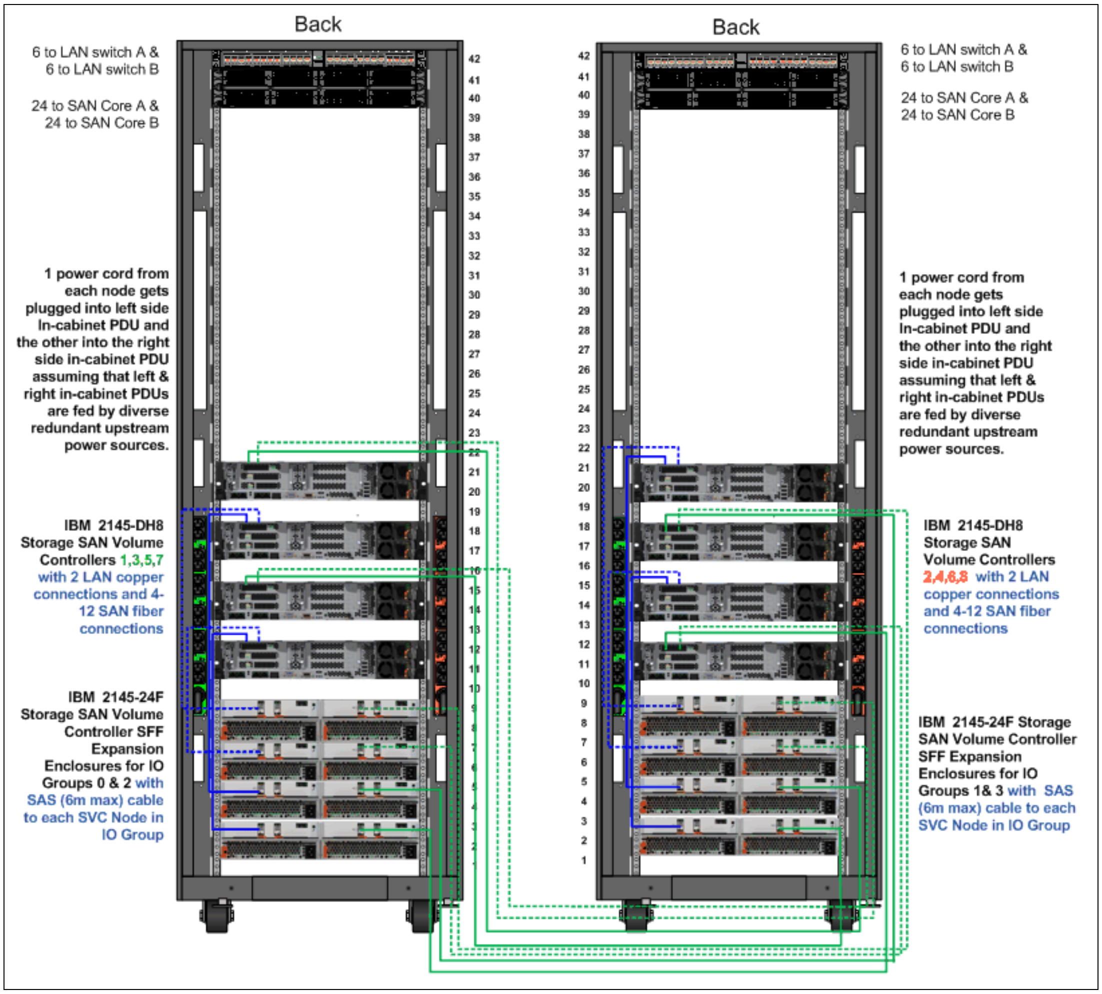


Figure 4-10 Rear view dual cabinet implementation of a new 2145-DH8 SVC node implementation with the optional 2145-24F SSD arrays installed

4.2.2 Cable connections

Create a cable connection table or similar documentation to track all of the connections that are required for the setup:

- Nodes
- Ethernet
- iSCSI/FCoE connections
- FC ports

Figure 4-11 shows the rear view of a 2145-DH8 node with the six PCIe adapter slots identified.

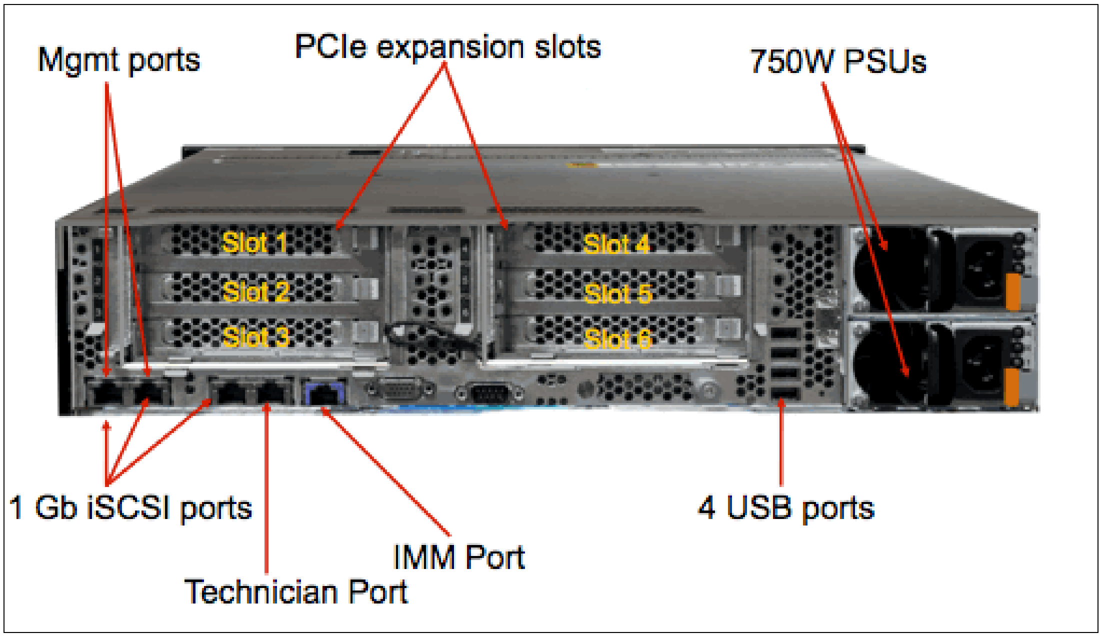


Figure 4-11 2145-DH8 PCIe Expansion Slots

Each 2145-DH8 node can support up to six PCIe expansion I/O cards, as identified in Table 4-1, to provide a range of connectivity and capacity expansion options.

Table 4-1 Layout of Expansion Card options for 2145-DH8 Node

Top of Node	
Processor 1 Attach	Processor 2 Attach
PCIe Slot 1 - I/O Card (FC only)	PCIe Slot 4 - Compression Accelerator Card
PCIe Slot 2 - I/O Card	PCIe Slot 5 - I/O Card
PCIe Slot 3 - SAS (For Flash Expansion)	PCIe Slot 6 - Compression Accelerator Card

There are four orderable I/O adapter options:

- Feature code AH10 - 4 Port 8 Gbps FC Card
  - Ships one 4 Port 8 Gbps FC Card with 4 Shortwave Transceivers.
  - Maximum feature quantity is 3.
- Feature code AH12 - 4 Port 10 GbE Ethernet Card
  - Ships one 4 Port 10 GbE Ethernet Card with 4 SFP+ Transceivers.
  - Maximum feature quantity is 1.
- Feature code AH1A - Compression Acceleration Card
  - Ships one Compression Acceleration Card.
  - Maximum feature quantity is 2.
- Feature code AH13 - SAS Expansion Enclosure Attach Card
  - Ships one 4 port 12 Gbps SAS Adapter.
  - Required for 2145-24F SAS Expansion Enclosure attachment.

A sample cable connection table can be downloaded using the following steps:

1. Go to: `http://www.ibm.com/storage/support/2145`.

2. Under Product support content, click “Plan and install documentation”.
3. In the search box, type and search for “IBM System Storage SAN Volume Controller Planning Information”; then click the search result.
4. Click the wanted link to download it to your computer.

4.3 Logical planning

For logical planning, we cover these topics:

- Management IP addressing plan
- SAN zoning and SAN connections
- iSCSI IP addressing plan
- IP Replication
- Backend storage subsystem configuration
- SVC system configuration
- Stretched cluster system configuration
- Storage pool configuration
- Volume configuration
- Host mapping (LUN masking)
- Advanced Copy Services functions
- SAN boot support
- Data migration from non-virtualized storage subsystems
- SVC configuration backup procedure

4.3.1 Management IP addressing plan

Starting with SVC 6.1, the system management is performed through an embedded GUI running on the nodes. A separate console, such as the traditional SVC Hardware Management Console (HMC) or IBM System Storage Productivity Center (SSPC), is no longer required to access the management interface. To access the management GUI, you direct a web browser to the system management IP address.

The SVC 2145-DH8 node introduces a new feature called a *Technician port*. Ethernet port 4 is allocated as the Technician service port, and is marked with a T. All initial configuration for each node is performed via the Technician port. The port broadcasts a DHCP service so that a notebook or computer is automatically assigned an IP address on connection to the port.

After the cluster configuration has been completed, the Technician port automatically routes the connected user directly to the service GUI.

Note: The default IP address for the Technician port on a 2145-DH8 Node is 192.168.0.1. If the Technician port is connected to a switch, it is disabled and an error is logged.

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

Note: For increased redundancy, an optional second Ethernet connection is supported for each SVC node. This cable is connected to Ethernet port 2.

To ensure system failover operations, Ethernet port 1 on all nodes must be connected to the same set of subnets. If used, Ethernet port 2 on all nodes must also be connected to the same set of subnets. However, the subnets for Ethernet port 1 do not have to be the same as Ethernet port 2.

Each SVC cluster has a Cluster Management IP address as well as a Service IP address for each node in the cluster. See Example 4-1 for details.

Example 4-1 Management IP address sample

```
management IP add. 10.11.12.120
node 1 service IP add. 10.11.12.121
node 2 service IP add. 10.11.12.122
node 3 service IP add. 10.11.12.123
Node 4 service IP add. 10.11.12.124
```

Each node in an SVC clustered system needs to have at least one Ethernet connection.

Figure 4-12 shows the IP configuration possibilities.

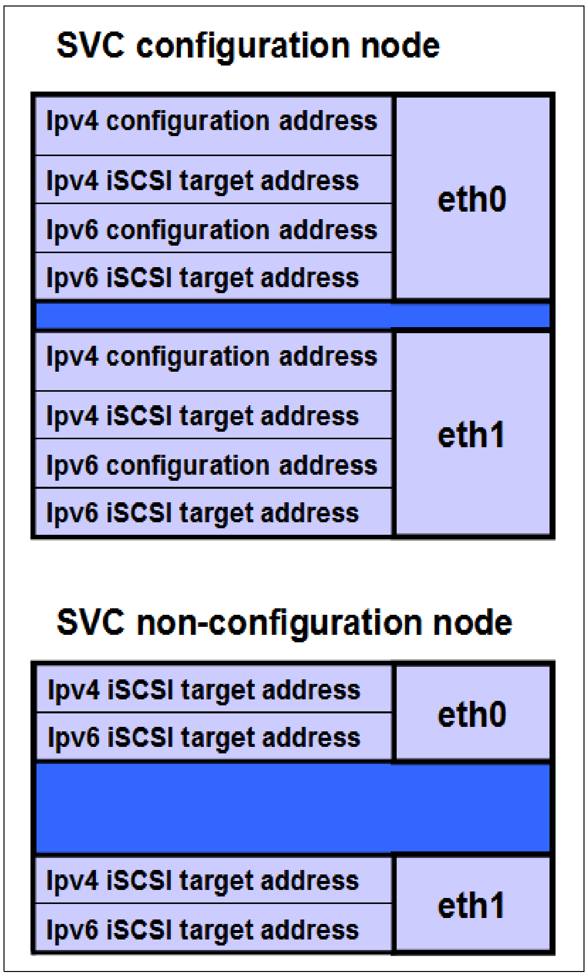


Figure 4-12 IP configuration possibilities

Support for iSCSI provides one additional IPv4 and one additional IPv6 address for each Ethernet port on every node. These IP addresses are independent of the clustered system configuration IP addresses.

When accessing the SVC through the GUI or Secure Shell (SSH), choose one of the available IP addresses to which to connect. No automatic failover capability is available. If one network is down, use an IP address on the alternate network. Clients might be able to use the intelligence in domain name servers (DNSs) to provide partial fail over.

4.3.2 SAN zoning and SAN connections

SAN storage systems using the SVC can be configured with two, or up to eight, SVC nodes, arranged in an SVC clustered system. These SVC nodes are attached to the SAN fabric, along with disk subsystems and host systems. The SAN fabric is zoned to allow the SVCs to “see” each other’s nodes and the disk subsystems, and for the hosts to “see” the SVCs. The hosts are not able to directly see or operate LUNs on the disk subsystems that are assigned to the SVC system. The SVC nodes within an SVC system must be able to see each other and all of the storage that is assigned to the SVC system.

The zoning capabilities of the SAN switch are used to create three distinct zones. SVC 7.3 supports 2 GBps, 4 GBps, or 8 GBps FC fabric, depending on the hardware platform and on the switch where the SVC is connected. In an environment where you have a fabric with multiple-speed switches, the preferred practice is to connect the SVC and the disk subsystem to the switch operating at the highest speed.

All SVC nodes in the SVC clustered system are connected to the same SANs, and they present volumes to the hosts. These volumes are created from storage pools that are composed of MDisks presented by the disk subsystems.

The fabric must have three distinct zones:

- *SVC cluster system zones*: Create up to two zones per fabric, and include a single port per node, which is designated for intracluster traffic. No more than four ports per node should be allocated to intracluster traffic.
- *Host zones*: Create an SVC host zone for each server accessing storage from the SVC system.
- *Storage zone*: Create one SVC storage zone for each storage subsystem that is virtualized by the SVC.

Port designation recommendations

The port to local node communication is used for mirroring write cache as well as metadata exchange between nodes and is critical to the stable operation of the cluster. The DH8 nodes with their 8-port and 12-port configurations provide an opportunity to isolate the port to local node traffic from other cluster traffic on dedicated ports thereby providing a level of protection against misbehaving devices and workloads that could compromise the performance of the shared ports.

Additionally, there is benefit in isolating remote replication traffic on dedicated ports as well to ensure that problems impacting the cluster-to-cluster interconnect do not adversely impact ports on the primary cluster and thereby impact the performance of workloads running on the primary cluster.

IBM recommends the following port designations for isolating both port to local and port to remote node traffic as shown in Table 4-2 on page 75.

Table 4-2 Port designation recommendations for isolating traffic

Port	SAN	4-port nodes	8-port nodes	12-port nodes	12-port nodes, write Data Rate > 3 GB/sec per IO Group
C1P1	A	Host/Storage/ Inter-node	Host/Storage	Host/Storage	Inter-node
C1P2	B	Host/Storage/ Inter-node	Host/Storage	Host/Storage	Inter-node
C1P3	A	Host/Storage/ Inter-node	Host/Storage	Host/Storage	Host/Storage
C1P4	B	Host/Storage/ Inter-node	Host/Storage	Host/Storage	Host/Storage
C2P1	A		Inter-node	Inter-node	Inter-node
C2P2	B		Inter-node	Inter-node	Inter-node
C2P3	A		Replication or Host/Storage	Host/Storage	Host/Storage
C2P4	B		Replication or Host/Storage	Host/Storage	Host/Storage
C5P1	A			Host/Storage	Host/Storage
C5P2	B			Host/Storage	Host/Storage
C5P3	A			Replication or Host/Storage	Replication or Host/Storage
C5P4	B			Replication or Host/Storage	Replication or Host/Storage
localfcportmask		1111	110000	110000	110011

Table 4-2 notes:

SAN column assumes an odd/even SAN port configuration. Modifications must be made if other SAN connection schemes are used.

Care needs to be taken when zoning so that inter-node ports are *not* used for Host/Storage in the 8-port and 12-port configurations.

These options represent optimal configurations based on port assignment to function. Using the same port assignment but different physical locations will not have any significant performance impact in most client environments.

This recommendation provides the wanted traffic isolation while also simplifying migration from existing configurations with only 4 ports, or even later migrating from 8-port or 12-port configurations to configurations with additional ports. More complicated port mapping configurations that spread the port traffic across the adapters are supported and can be considered, but these approaches do not appreciably increase availability of the solution since the mean time between failures (MTBF) of the adapter is not significantly less than that of the non-redundant node components.

Note that while it is true that alternate port mappings that spread traffic across HBAs may allow adapters to come back online following a failure, they will not prevent a node from going offline temporarily to reboot and attempt to isolate the failed adapter and then rejoin the cluster. Our recommendation takes all these considerations into account with a view that the greater complexity may lead to migration challenges in the future and the simpler approach is best.

4.3.3 iSCSI IP addressing plan

Since version 6.3, SVC supports host access through iSCSI (as an alternative to FC), and the following considerations apply:

- SVC uses the built-in Ethernet ports for iSCSI traffic. If the optional 10 Gbps Ethernet feature is installed, you can connect host systems through the two 10 Gbps Ethernet ports per node.
- All node types, which can run SVC 6.1 or later, can use the iSCSI feature.
- SVC supports the Challenge Handshake Authentication Protocol (CHAP) authentication methods for iSCSI.
- iSCSI IP addresses can fail over to the partner node in the I/O Group if a node fails. This design reduces the need for multipathing support in the iSCSI host.
- iSCSI IP addresses can be configured for one or more nodes.
- iSCSI Simple Name Server (iSNS) addresses can be configured in the SVC.
- The iSCSI qualified name (IQN) for a SVC node is `iqn.1986-03.com.ibm:2145.<cluster_name>.<node_name>`. Because the IQN contains the clustered system name and the node name, it is important not to change these names after iSCSI is deployed.
- Each node can be given an iSCSI alias, as an alternative to the IQN.
- The IQN of the host to an SVC host object is added in the same way that you add FC WWPNs.
- Host objects can have both WWPNs and IQNs.
- Standard iSCSI host connection procedures can be used to discover and configure SVC as an iSCSI target.

Next, we explain several ways in which you can configure SVC 6.1 or later.

Figure 4-13 on page 77 shows the use of IPv4 management and iSCSI addresses in the same subnet.

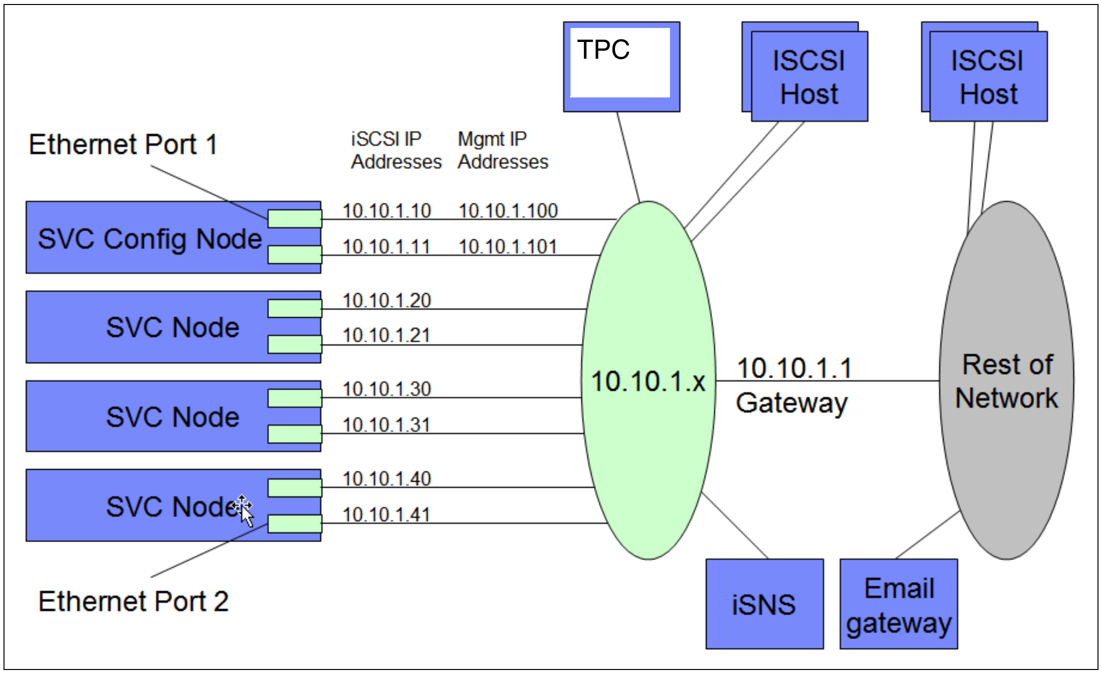


Figure 4-13 Use of IPv4 addresses

You can set up the equivalent configuration with only IPv6 addresses.

Figure 4-14 shows the use of IPv4 management and iSCSI addresses in two separate subnets.

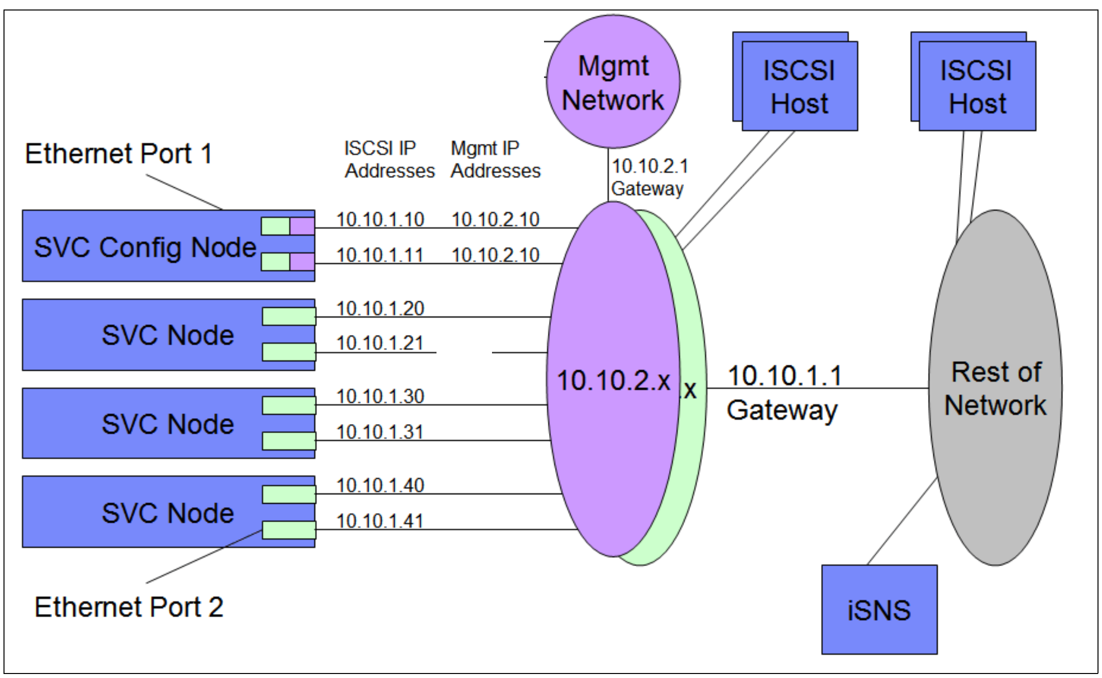


Figure 4-14 IPv4 address plan with two subnets

Figure 4-15 shows the use of redundant networks.

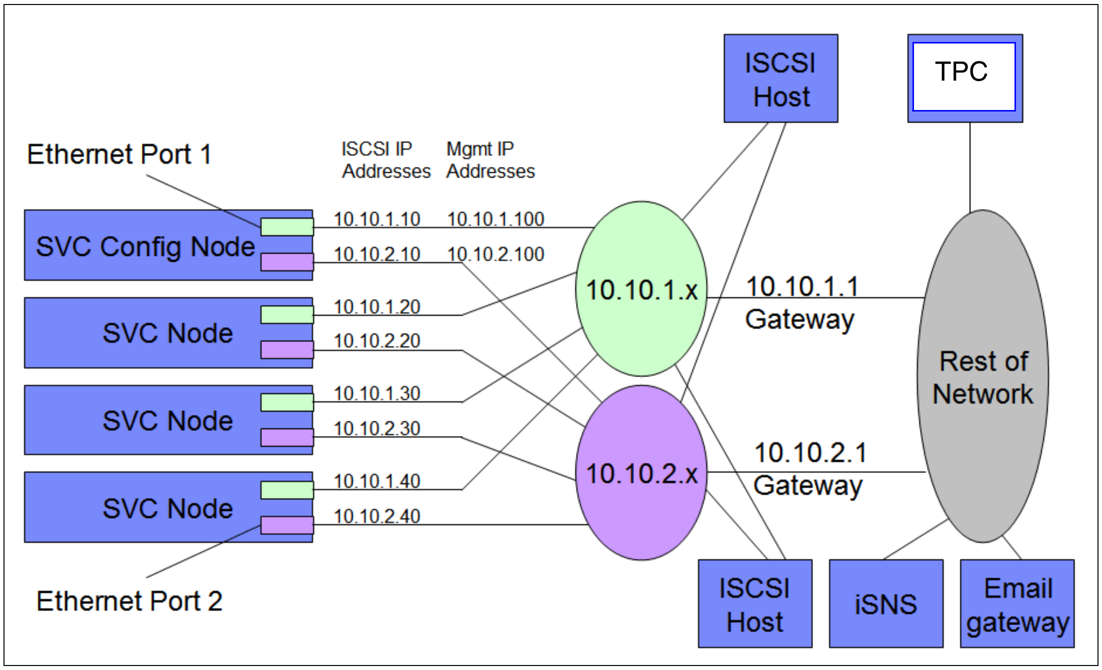


Figure 4-15 Redundant networks

Figure 4-16 shows the use of a redundant network and a third subnet for management.

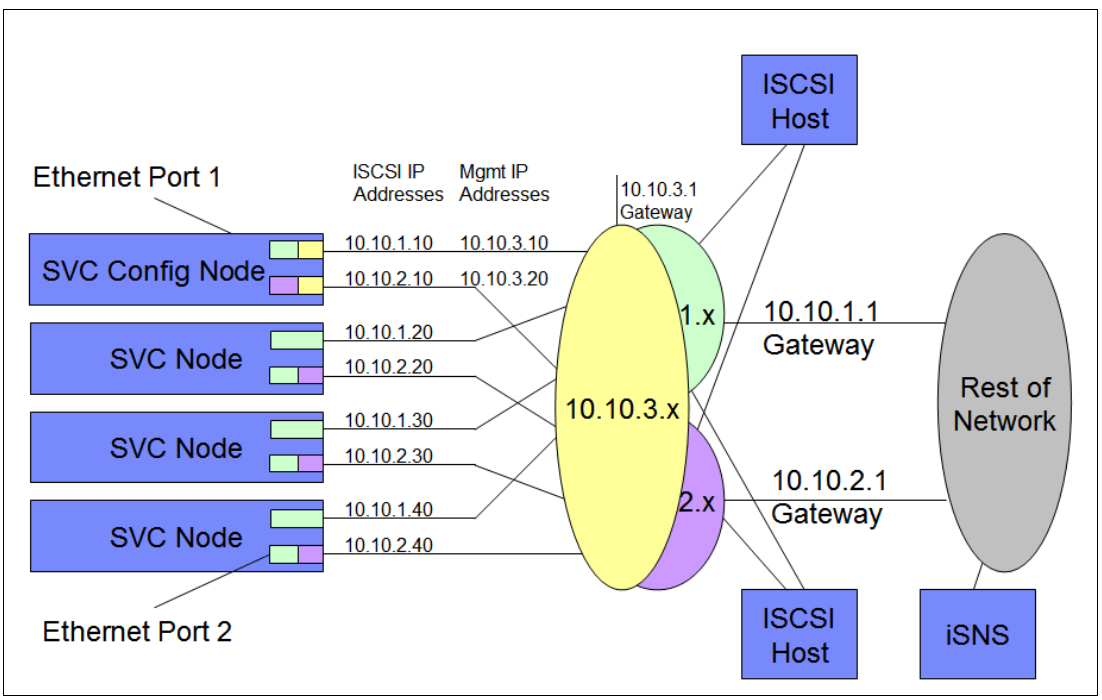


Figure 4-16 Redundant network with third subnet for management

Figure 4-17 shows the use of a redundant network for both iSCSI data and management.

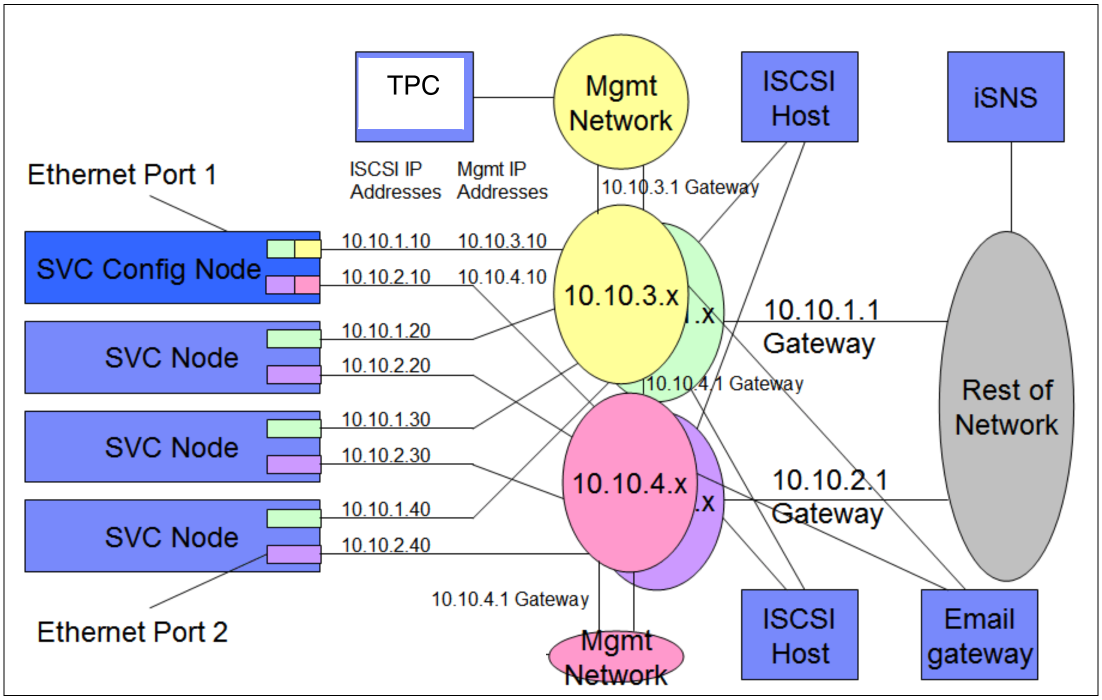


Figure 4-17 Redundant network for iSCSI and management

Be aware of these considerations:

- All of the examples are valid for IPv4 and IPv6 addresses.
- It is valid to use IPv4 addresses on one port and IPv6 addresses on the other port.
- It is valid to have separate subnet configurations for IPv4 and IPv6 addresses.

4.3.4 IP replication

One of the most important new functions introduced in version 7.2 of the Storwize family is IP replication, which enables the use of lower-cost Ethernet connections for remote mirroring. The capability is available as an option (Metro or Global Mirror) on all Storwize family systems. The new function is transparent to servers and applications in the same way that traditional FC-based mirroring is. All remote mirroring modes (Metro Mirror, Global Mirror, and Global Mirror with Changed Volumes) are supported. Configuration of the system is straightforward: Storwize family systems can normally find each other on the network and can be selected from the GUI. IP replication includes Bridgeworks SANSlide network optimization technology and is available at no additional charge. Remote mirror is a chargeable option but the price does not change with IP replication. Existing remote mirror users have access to the new function at no additional charge.

Note: Full details of how to set up and configure IP Replication are available in the SVC and Storwize family Native IP Replication publication at the following Redbooks publication site:

http://www.redbooks.ibm.com/redpieces/abstracts/redp5103.html

4.3.5 Backend storage subsystem configuration

Backend storage subsystem configuration planning must be applied to all storage controllers that are attached to the SVC.

See the following website for a list of currently supported storage subsystems:

http://www.ibm.com/support/docview.wss?uid=ssg1S1003658

Apply the following general guidelines for backend storage subsystem configuration planning:

- In the SAN, storage controllers that are used by the SVC clustered system must be connected through SAN switches. Direct connection between the SVC and the storage controller is not supported.
- Multiple connections are allowed from the redundant controllers in the disk subsystem to improve data bandwidth performance. It is not mandatory to have a connection from each redundant controller in the disk subsystem to each counterpart SAN, but it is a preferred practice. Therefore, canister A in a Storwize V3700 subsystem can be connected to SAN A only, or to SAN A and SAN B. And, canister B in the same Storwize V3700 subsystem can be connected to SAN B only, or to SAN B and SAN A.
- Stretched System configurations are supported by certain rules and configuration guidelines. See Chapter 11, “SAN Volume Controller Stretched System” on page 345 for more information.
- All SVC nodes in an SVC clustered system must be able to see the same set of ports from each storage subsystem controller. Violating this guideline causes the paths to become degraded. This degradation can occur as a result of applying inappropriate zoning and LUN masking. This guideline has important implications for a disk subsystem, such as DS3000, Storwize V3700, Storwize V5000, or Storwize V7000, which imposes exclusivity rules as to which HBA WWPNs a storage partition can be mapped.

MDisks within storage pools:

- SVC 6.1 and later provide for better load distribution across paths within storage pools.

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

  This condition is more likely to occur with a smaller number of MDisks contained in the storage pool. Starting with SVC 6.1, the code contains logic that considers MDisks within storage pools. Therefore, the code more effectively distributes their active paths that are based on the storage controller ports that are available.
- The `Detect Mdisk` commands must be run following the creation or modification (add or remove MDisk) of storage pools for paths to be redistributed.

If you do not have a storage subsystem that supports the SVC round-robin algorithm, make the number of MDisks per storage pool a multiple of the number of storage ports that are available. This approach ensures sufficient bandwidth to the storage controller and an even balance across storage controller ports.

In general, configure disk subsystems as though no SVC exists. However, we suggest the following specific guidelines:

- Disk drives:
  - Exercise caution with large disk drives so that you do not have too few spindles to handle the load.
  - RAID 5 is suggested for most workloads.
- Array sizes:
  - An array size of 8+P or 4+P is suggested for the IBM DS4000® and DS5000™ families, if possible.
  - Use the DS4000 segment size of 128 KB or larger to help the sequential performance.
  - Upgrade to EXP810 drawers, if possible.
  - Create LUN sizes that are equal to the RAID array and rank size. If the array size is greater than 2 TB and the disk subsystem does not support MDisks larger than 2 TB, create the minimum number of LUNs of equal size.
  - An array size of 7+P is suggested for the V3700, V5000, and V7000 Storwize families.
  - When adding more disks to a subsystem, consider adding the new MDisks to existing storage pools versus creating additional small storage pools.
- Maximum of 1024 WWNNs per cluster:
  - EMC DMX/SYMM, all HDS, and SUN/HP HDS clones use one WWNN per port. Each WWNN appears as a separate controller to the SVC.
  - IBM, EMC CLARiiON, and HP use one WWNN per subsystem. Each WWNN appears as a single controller with multiple ports/WWPNs, for a maximum of 16 ports/WWPNs per WWNN.
- DS8000 using four of, or eight of, the 4-port HA cards:
  - Use ports 1 and 3 or ports 2 and 4 on each card (it does not matter for 8 Gb cards).
  - This setup provides 8 or 16 ports for SVC use.
  - Use eight ports minimum, up to 40 ranks.
  - Use 16 ports for 40 or more ranks. Sixteen is the maximum number of ports.
- DS4000/DS5000 – EMC CLARiiON/CX:
  - Both systems have the preferred controller architecture, and SVC supports this configuration.
  - Use a minimum of four ports, and preferably eight or more ports, up to a maximum of 16 ports, so that more ports equate to more concurrent I/O that is driven by the SVC.
  - Support is available for mapping controller A ports to Fabric A and controller B ports to Fabric B or cross-connecting ports to both fabrics from both controllers. The cross-connecting approach is preferred to avoid auto volume transfer (AVT)/Trespass occurring if a fabric or all paths to a fabric fail.
- DS3400 subsystems:
  - Use a minimum of four ports.
- Storwize family:
  - Use a minimum of four ports, and preferably eight ports.

- IBM XIV requirements and restrictions:
  - The use of XIV extended functions, including snaps, thin provisioning, synchronous replication (native copy services), and LUN expansion of LUNs presented to the SVC is not supported.
  - A maximum of 511 LUNs from one XIV system can be mapped to an SVC clustered system.
- Full 15 module XIV recommendations – 161 usable TB:
  - Use two interface host ports from each of the six interface modules.
  - Use ports 1 and 3 from each interface module and zone these 12 ports with all SVC node ports.
  - Create 48 LUNs of equal size, each of which is a multiple of 17 GB. This creates approximately 1632 GB if you are using the entire full frame XIV with the SVC.
  - Map LUNs to the SVC as 48 MDisks, and add all of them to the single XIV storage pool so that the SVC drives the I/O to four MDisks and LUNs for each of the 12 XIV FC ports. This design provides a good queue depth on the SVC to drive XIV adequately.
- Six module XIV recommendations - 55 TB usable:
  - Use two interface host ports from each of the two active interface modules.
  - Use ports 1 and 3 from interface modules 4 and 5. (Interface module 6 is inactive). Also, zone these four ports with all SVC node ports.
  - Create 16 LUNs of equal size, each of which is a multiple of 17 GB. This creates approximately 1632 GB if you are using the entire XIV with the SVC.
  - Map the LUNs to the SVC as 16 MDisks, and add all of them to the single XIV storage pool, so that the SVC drives I/O to four MDisks and LUNs per each of the four XIV FC ports. This design provides a good queue depth on the SVC to drive the XIV adequately.
- Nine module XIV recommendations - 87 usable TB:
  - Use two interface host ports from each of the four active interface modules.
  - Use ports 1 and 3 from interface modules 4, 5, 7, and 8. (Interface modules 6 and 9 are inactive). Also, zone these eight ports with all of the SVC node ports.
  - Create 26 LUNs of equal size, each of which is a multiple of 17 GB. This creates approximately 1632 GB approximately if you are using the entire XIV with the SVC.
  - Map the LUNs to the SVC as 26 MDisks, and map all of them to the single XIV storage pool, so that the SVC drives I/O to three MDisks and LUNs on each of the six ports and four MDisks and LUNs on the other two XIV FC ports. This design provides a useful queue depth on SVC to drive XIV adequately.
- Configure XIV host connectivity for the SVC clustered system:
  - Create one host definition on XIV, and include all SVC node WWPNs.
  - You can create clustered system host definitions (one per I/O Group), but the preceding method is easier.
  - Map all LUNs to all SVC node WWPNs.

4.3.6 Real-time Compression

The 2145-DH8 introduces additional hardware dedicated to the improvement of the Real-time Compression functionality within SVC. It is a mandatory requirement that each node has the second processor and cache upgrade, as well as a minimum of one Compression Acceleration card, for the I/O Group to support compressed volumes.

Consideration should be taken when sizing the number of Compression Acceleration cards per node. If your active data workload is greater than 8 TB, you should consider deploying both Compression Acceleration cards per node. With a single Compression Acceleration card in each node, the existing recommendation on the number of compressed volumes able to be managed per I/O group remains the same at 200 volumes. However, with the addition of the second Compression Acceleration card in each node (a total of four cards per I/O group), the total number of managed compressed volumes increases to 512.

> **Note:** Active Data Workload is typically 5 - 8% of the total managed capacity. In a single I/O Group, 8 TB of active data equates to approximately 160 TB managed. In an eight-node cluster, this equates to 32 TB of active data (8 TB per I/O Group).

4.3.7 EasyTier version 3

With the release of SVC 7.3, Easy Tier has been enhanced to support several new features:

- Easy Tier with three tiers in a pool

  Nearline (NL-SAS), Enterprise (SAS), and Flash (SSD or Flash).

- Easy Tier puts hot extents on faster storage, and cold extents on slower storage.
- Easy Tier with any two tiers in a pool

  So Nearline + Enterprise as well as (anything) + Flash.

- Drive and storage system sensitivity
  - Easy Tier understands exactly what type of drive and RAID level, and approximately what class of storage system is being used.
  - So it knows how much performance to expect from an MDisk and avoids overloading.
- Major enhancements to the STAT tool to support the above and add more metrics.
  - STAT tool outputs three sets of data
  - Detailed logging can be uploaded to Disk Magic to validate skew curves

Figure 4-18 on page 84 shows the basic layout of how Easy Tier works. With SVC, the user must manually define flash disk MDisk or Nearline MDisks; all MDisks are classed as Enterprise by default.

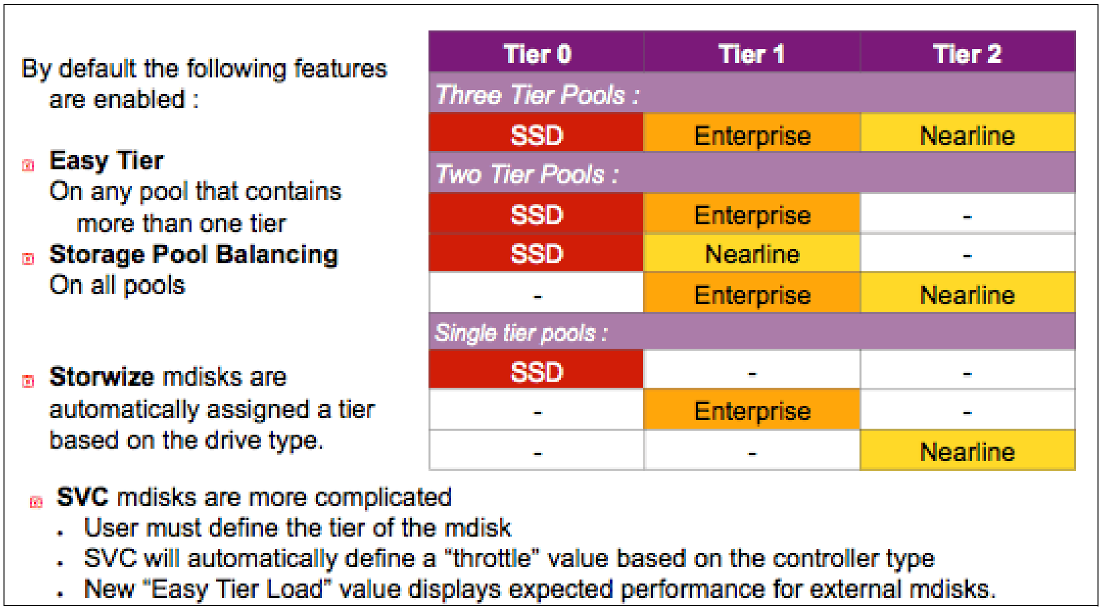


Figure 4-18 Identifies the three tiers of disk that are accessible by Easy Tier

4.3.8 SAN Volume Controller clustered system configuration

To ensure high availability in SVC installations, consider the following guidelines when you design a SAN with the SVC:

- All nodes in a clustered system must be in the same LAN segment because the nodes in the clustered system must be able to assume the same clustered system or service IP address. Make sure that the network configuration allows any of the nodes to use these IP addresses. If you plan to use the second Ethernet port on each node, it is possible to have two LAN segments. However, port 1 of every node must be in one LAN segment, and port 2 of every node must be in the other LAN segment.
- To maintain application uptime in the unlikely event of an individual SVC node failing, SVC nodes are always deployed in pairs (I/O Groups). If a node fails or is removed from the configuration, the remaining node operates in a degraded mode, but it is still a valid configuration. The remaining node operates in write-through mode, meaning that the data is written directly to the disk subsystem (the cache is disabled for the write).
- The FC SAN connections between the SVC node and the switches are optical fiber. These connections can run at either 2 GBps, 4 GBps, or 8 GBps, depending on your SVC and switch hardware.
- The SVC node ports must be connected to the FC fabric only. Direct connections between the SVC and the host, or the disk subsystem, are unsupported.
- Two SVC clustered systems cannot have access to the same LUNs within a disk subsystem. Configuring zoning so that two SVC clustered systems have access to the same LUNs (MDisks) can, and will likely, result in data corruption.
- The two nodes within an I/O Group can be co-located (within the same set of racks) or can be in separate racks and separate rooms. See 4.3.9, "Stretched System configuration" on page 85 for more information about this topic.
- The SVC uses three MDisks as quorum disks for the clustered system. A preferred practice for redundancy is to have each quorum disk in a separate storage subsystem, where possible. The current locations of the quorum disks can be displayed using the **`lsquorum`** command and relocated using the **`chquorum`** command.

4.3.9 Stretched System configuration

You can implement a stretched system configuration (historically referred to as a *Split I/O Group,* or *Stretched Cluster*) as a high-availability option.

The SVC Stretched System configuration provides a continuous availability platform, whereby host access is maintained in the event of the loss of any single failure domain. This availability is accomplished through the inherent active/active architecture of SVC along with the use of volume mirroring. During a failure, the SVC nodes and associated mirror copy of the data remain online and available to service all host I/O.

The existing SVC Stretched System configuration has two locations, each with one node from each I/O Group pair. The quorum disks are usually held in a third location.

See Chapter 11, “SAN Volume Controller Stretched System” on page 345 for more information about Stretched System configurations.

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

MDisks in the SVC are LUNs assigned from the underlying disk subsystems to the SVC and can be either managed or unmanaged. A managed MDisk is an MDisk that is assigned to a storage pool:

- A storage pool is a collection of MDisks. An MDisk can only be contained within a single storage pool.
- SVC supports up to 128 storage pools.
- The number of volumes that can be allocated from a storage pool is unlimited; however, an I/O Group is limited to 2048, and the clustered system limit is 8192.
- Volumes are associated with a single storage pool, except in cases where a volume is being migrated or mirrored between storage pools.

The SVC supports extent sizes of 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, and 8192 MB. Support for extent sizes 4096 and 8192 was added in SVC 6.1. The extent size is a property of the storage pool and is set when the storage pool is created. All MDisks in the storage pool have the same extent size, and all volumes that are allocated from the storage pool have the same extent size. The extent size of a storage pool cannot be changed. If another extent size is wanted, the storage pool must be deleted and a new storage pool configured.

Table 4-3 lists all of the extent sizes that are available in an SVC.

Table 4-3 Extent size and maximum clustered system capacities

Extent size	Maximum clustered system capacity
16 MB	64 TB
32 MB	128 TB
64 MB	256 TB
128 MB	512 TB
256 MB	1 PB
512 MB	2 PB
1,024 MB	4 PB
2,048 MB	8 PB
4,096 MB	16 PB
8,192 MB	32 PB

Consider the following information about storage pools:

- Maximum clustered system capacity is related to the extent size:
  - Sixteen MB extent = 64 TB and doubles for each increment in extent size; for example, 32 MB = 128 TB. We strongly advise a minimum 128/256 MB. The IBM submissions to the Storage Performance Council (SPC) benchmarks used a 256 MB extent.
  - Pick the extent size, and use that size for all storage pools.
  - You cannot migrate volumes between storage pools with separate extent sizes. However, you can use volume mirroring to create copies between storage pools with separate extent sizes.
- Storage pool reliability, availability, and serviceability (RAS) considerations:
  - It might make sense to create multiple storage pools if you ensure that a host only gets its volumes built from one of the storage pools. If the storage pool goes offline, it affects only a subset of all the hosts using the SVC. However, creating multiple storage pools can cause a high number of storage pools, approaching the SVC limits.
  - If you do not isolate hosts to storage pools, create one large storage pool. Creating one large storage pool assumes that the physical disks are all the same size, speed, and RAID level.
  - The storage pool goes offline if an MDisk is unavailable, even if the MDisk has no data on it. Do not put MDisks into a storage pool until they are needed.
  - Create at least one separate storage pool for all the image mode volumes.
  - Make sure that the LUNs that are given to the SVC have all host-persistent reserves removed.

- Storage pool performance considerations

  It might make sense to create multiple storage pools if you are attempting to isolate workloads to separate disk spindles. Storage pools with too few MDisks cause an MDisk overload, so it is better to have more spindle counts in a storage pool to meet workload requirements.

- The storage pool and SVC cache relationship

  The SVC employs cache partitioning to limit the potentially negative effect that a poorly performing storage controller can have on the clustered system. The partition allocation size is defined based on the number of configured storage pools. This design protects against individual controller overloading and failures from consuming write cache and degrading performance of the other storage pools in the clustered system. We discuss more details in 4.4.3, “Cache” on page 100.

  Table 4-4 Limit of the cache data

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

  Consider the rule that no single partition can occupy more than its upper limit of cache capacity with write data. These limits are upper limits, and they are the points at which the SVC cache will start to limit incoming I/O rates for volumes that are created from the storage pool. If a particular partition reaches this upper limit, the net result is the same as a global cache resource that is full. That is, the host writes will be serviced on a one-out-one-in basis because the cache destages writes to the backend disks.

  However, only writes that are targeted at the full partition are limited. All I/O destined for other (non-limited) storage pools continues as normal. The read I/O requests for the limited partition also continue normally. However, because the SVC is destaging write data at a rate that is obviously greater than the controller can sustain (otherwise, the partition does not reach the upper limit), read response times are also likely affected.

4.3.10 Volume configuration

An individual volume is a member of one storage pool and one I/O Group. When creating a volume, you first identify the wanted performance, availability, and cost requirements for that volume, and then select the storage pool accordingly:

- The storage pool defines which MDisks provided by the disk subsystem make up the volume.
- The I/O Group (two nodes make an I/O Group) defines which SVC nodes provide I/O access to the volume.

Important: There is no fixed relationship between I/O Groups and storage pools.

Perform volume allocation based on the following considerations:

- Optimize performance between the hosts and the SVC by attempting to distribute volumes evenly across available I/O Groups and nodes within the clustered system.

- Reach the level of performance, reliability, and capacity that you require by using the storage pool that corresponds to your needs (you can access any storage pool from any node). That is, choose the storage pool that fulfills the demands for your volumes regarding performance, reliability, and capacity.
- I/O Group considerations:
  - When you create a volume, it is associated with one node of an I/O Group. By default, every time that you create a new volume, it is associated with the next node using a round-robin algorithm. You can specify a *preferred access node*, which is the node through which you send I/O to the volume instead of using the round-robin algorithm. A volume is defined for an I/O Group.
  - Even if you have eight paths for each volume, all I/O traffic flows only toward one node (the preferred node). Therefore, only four paths are used by the IBM Subsystem Device Driver (SDD). The other four paths are used only in the case of a failure of the preferred node or when concurrent code upgrade is running.
- Creating image mode volumes:
  - Use image mode volumes when an MDisk already has data on it, from a non-virtualized disk subsystem. When an image mode volume is created, it directly corresponds to the MDisk from which it is created. Therefore, volume logical block address (LBA) x = MDisk LBA x. The capacity of image mode volumes defaults to the capacity of the supplied MDisk.
  - When you create an image mode disk, the MDisk must have a mode of *unmanaged* and therefore does not belong to any storage pool. A capacity of 0 is not allowed. Image mode volumes can be created in sizes with a minimum granularity of 512 bytes, and they must be at least one block (512 bytes) in size.
- Creating managed mode volumes with sequential or striped policy

  When creating a managed mode volume with sequential or striped policy, you must use a number of MDisks containing extents that are free and of a size that is equal to or greater than the size of the volume that you want to create. There might be sufficient extents available on the MDisk, but a contiguous block large enough to satisfy the request might not be available.
- Thin-provisioned volume considerations:
  - When creating the thin-provisioned volume, you need to understand the utilization patterns by the applications or group users accessing this volume. You must consider items such as the actual size of the data, the rate of creation of new data, modifying or deleting existing data.
  - Two operating modes for thin-provisioned volumes are available:
    - *Autoexpand volumes* allocate storage from a storage pool on demand with minimal required user intervention. However, a misbehaving application can cause a volume to expand until it has consumed all of the storage in a storage pool.
    - *Non-autoexpand volumes* have a fixed amount of assigned storage. In this case, the user must monitor the volume and assign additional capacity when required. A misbehaving application can only cause the volume that it uses to fill up.
  - Depending on the initial size for the real capacity, the grain size and a warning level can be set. If a volume goes offline, either through a lack of available physical storage for autoexpand, or because a volume that is marked as non-expand had not been expanded in time, a danger exists of data being left in the cache until storage is made available. This situation is not a data integrity or data loss issue, but you must not rely on the SVC cache as a backup storage mechanism.

Important:

- Keep a warning level on the used capacity so that it provides adequate time to respond and provision more physical capacity.
- Warnings must not be ignored by an administrator.
- Use the autoexpand feature of the thin-provisioned volumes.

  – When you create a thin-provisioned volume, you can choose the grain size for allocating space in 32 KB, 64 KB, 128 KB, or 256 KB chunks. The grain size that you select affects the maximum virtual capacity for the thin-provisioned volume. The default grain size is 256 KB, and is the strongly recommended option. If you select 32 KB for the grain size, the volume size cannot exceed 260,000 GB. The grain size cannot be changed after the thin-provisioned volume is created. Generally, smaller grain sizes save space but require more metadata access, which could adversely affect performance. If you are not going to use the thin-provisioned volume as a FlashCopy source or target volume, use 256 KB to maximize performance. If you are going to use the thin-provisioned volume as a FlashCopy source or target volume, specify the same grain size for the volume and for the FlashCopy function.

  – Thin-provisioned volumes require more I/Os because of directory accesses. For truly random workloads with 70% read and 30% write, a thin-provisioned volume requires approximately one directory I/O for every user I/O.

  – The directory is two-way write-back-cached (just like the SVC fastwrite cache), so certain applications perform better.

  – Thin-provisioned volumes require more processor processing, so the performance per I/O Group can also be reduced.

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

- Volume mirroring guidelines:

  – Create or identify two separate storage pools to allocate space for your mirrored volume.

  – Allocate the storage pools containing the mirrors from separate storage controllers.

  – If possible, use a storage pool with MDisks that share the same characteristics. Otherwise, the volume performance can be affected by the poorer performing MDisk.

4.3.11 Host mapping (LUN masking)

For the host and application servers, the following guidelines apply:

- Each SVC node presents a volume to the SAN through four ports. Because two nodes are used in normal operations to provide redundant paths to the same storage, a host with two HBAs can see multiple paths to each LUN that is presented by the SVC. Use zoning to limit the pathing from a minimum of two paths to the maximum that is available of eight paths, depending on the kind of high availability and performance that you want to have in your configuration.

  It is best to use zoning to limit the pathing to four paths. The hosts must run a multipathing device driver to limit the pathing back to a single device. The multipathing driver supported and delivered by SVC is the SDD. Native multipath I/O (MPIO) drivers on selected hosts are supported. For operating system-specific information about MPIO support, see this website:

http://www.ibm.com/systems/storage/software/virtualization/svc/interop.html

You can find the actual version of the Subsystem Device Driver Device Specific Module (SDDDSM) for IBM products at the following link:

http://www.ibm.com/support/docview.wss?uid=ssg1S4000350

- The number of paths to a volume from a host to the nodes in the I/O Group that owns the volume must not exceed eight, even if eight is not the maximum number of paths supported by the multipath driver (SDD supports up to 32). To restrict the number of paths to a host volume, the fabrics must be zoned so that each host FC port is zoned to no more than two ports from each SVC node in the I/O Group that owns the volume.

Multipathing:

The following list suggests the number of paths per volume (*n*+1 redundancy):

- With 2 HBA ports, zone the HBA ports to the SVC ports 1:2 for a total of 4 paths.
- With 4 HBA ports, zone the HBA ports to the SVC ports 1:1 for a total of 4 paths.

Optional (*n*+2 redundancy):

With 4 HBA ports, zone the HBA ports to the SVC ports 1:2 for a total of 8 paths.

We use the term *HBA port* to describe the *SCSI initiator*. We use the term *SVC port* to describe the *SCSI target*.

The maximum number of host paths per volume must not exceed eight.

- If a host has multiple HBA ports, each port must be zoned to a separate set of SVC ports to maximize high availability and performance.
- To configure greater than 256 hosts, you must configure the host to I/O Group mappings on the SVC. Each I/O Group can contain a maximum of 256 hosts, so it is possible to create 1024 host objects on an eight-node SVC clustered system. Volumes can only be mapped to a host that is associated with the I/O Group to which the volume belongs.
- Port masking

  You can use a *port mask* to control the node target ports that a host can access, which satisfies two requirements:

  - As part of a security policy to limit the set of WWPNs that are able to obtain access to any volumes through a given SVC port
  - As part of a scheme to limit the number of logins with mapped volumes visible to a host multipathing driver, such as SDD, and thus limit the number of host objects configured without resorting to switch zoning
- The port mask is an optional parameter of the **`mkhost`** and **`chhost`** commands. The port mask is four binary bits. Valid mask values range from 0000 (no ports enabled) to 1111 (all ports enabled). For example, a mask of 0011 enables port 1 and port 2. The default value is 1111 (all ports enabled).
- The SVC supports connection to the Cisco MDS family and Brocade family. See the following website for the latest support information:

  http://www.ibm.com/systems/storage/software/virtualization/svc/interop.html

4.3.12 Advanced Copy Services

The SVC offers these Advanced Copy Services:

- FlashCopy
- Metro Mirror
- Global Mirror

Layers: SVC 6.3 introduces a new property for the clustered system that is called *layer*. This property is used when a copy services partnership exists between an SVC and an IBM Storwize V7000. There are two layers: *replication* and *storage*. All SVC clustered systems are *replication* layers and cannot be changed. By default, the IBM Storwize V7000 is a *storage* layer and must be changed with the CLI command **`chsystem`** before you use it to make any copy services partnership with the SVC. Figure 4-19 shows an example for replication and storage layer.

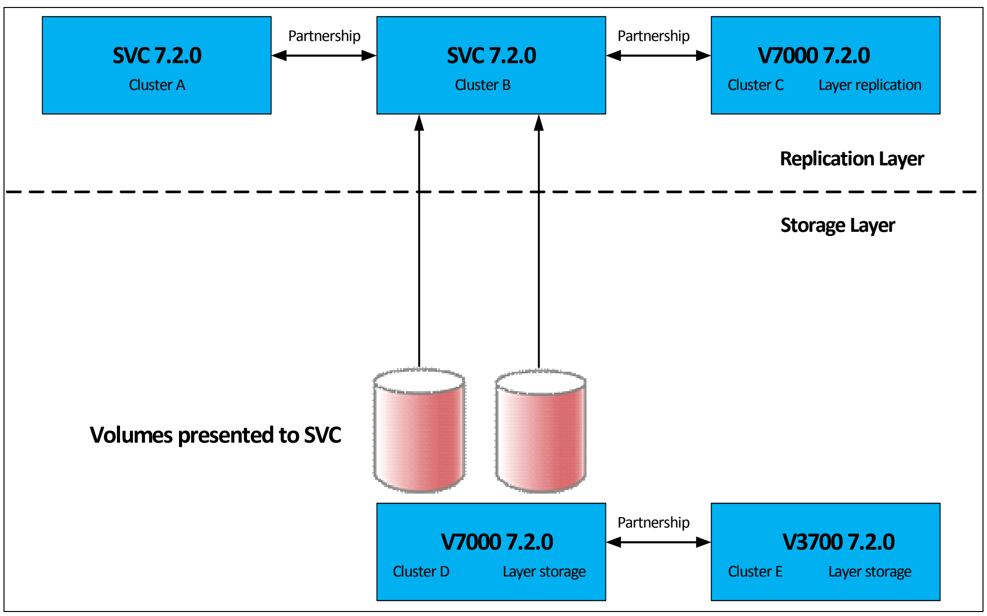


Figure 4-19 Replication Layer and Storage Layer

SVC Advanced Copy Services must apply the following guidelines.

FlashCopy guidelines

Consider these FlashCopy guidelines:

- Identify each application that must have a FlashCopy function implemented for its volume.
- FlashCopy is a relationship between volumes. Those volumes can belong to separate storage pools and separate storage subsystems.
- You can use FlashCopy for backup purposes by interacting with the Tivoli Storage Manager Agent, or for cloning a particular environment.
- Define which FlashCopy best fits your requirements: No copy, Full copy, Thin-Provisioned, or Incremental.

- Define which FlashCopy rate best fits your requirement in terms of the performance and the amount of time to complete the FlashCopy. Table 4-5 shows the relationship of the background copy rate value to the attempted number of grains to be split per second.
- Define the grain size that you want to use. A *grain* is the unit of data that is represented by a single bit in the FlashCopy bitmap table. Larger grain sizes can cause a longer FlashCopy elapsed time and a higher space usage in the FlashCopy target volume. Smaller grain sizes can have the opposite effect. Remember that the data structure and the source data location can modify those effects.

  In an actual environment, check the results of your FlashCopy procedure in terms of the data that is copied at every run and in terms of elapsed time, comparing them to the new SVC FlashCopy results. Eventually, adapt the grain/second and the copy rate parameter to fit your environment's requirements.

Table 4-5 Grain splits per second

User percentage	Data copied per second	256 KB grain per second	64 KB grain per second
1 - 10	128 KB	0.5	2
11 - 20	256 KB	1	4
21 - 30	512 KB	2	8
31 - 40	1 MB	4	16
41 - 50	2 MB	8	32
51 - 60	4 MB	16	64
61 - 70	8 MB	32	128
71 - 80	16 Mb	64	256
81 - 90	32 MB	128	512
91 - 100	64 MB	256	1024

Metro Mirror and Global Mirror guidelines

SVC supports both intracluster and intercluster Metro Mirror and Global Mirror. From the intracluster point of view, any single clustered system is a reasonable candidate for a Metro Mirror or Global Mirror operation. Intercluster operation, however, needs at least two clustered systems that are separated by a number of moderately high-bandwidth links.

Figure 4-20 on page 93 shows a schematic of Metro Mirror connections.

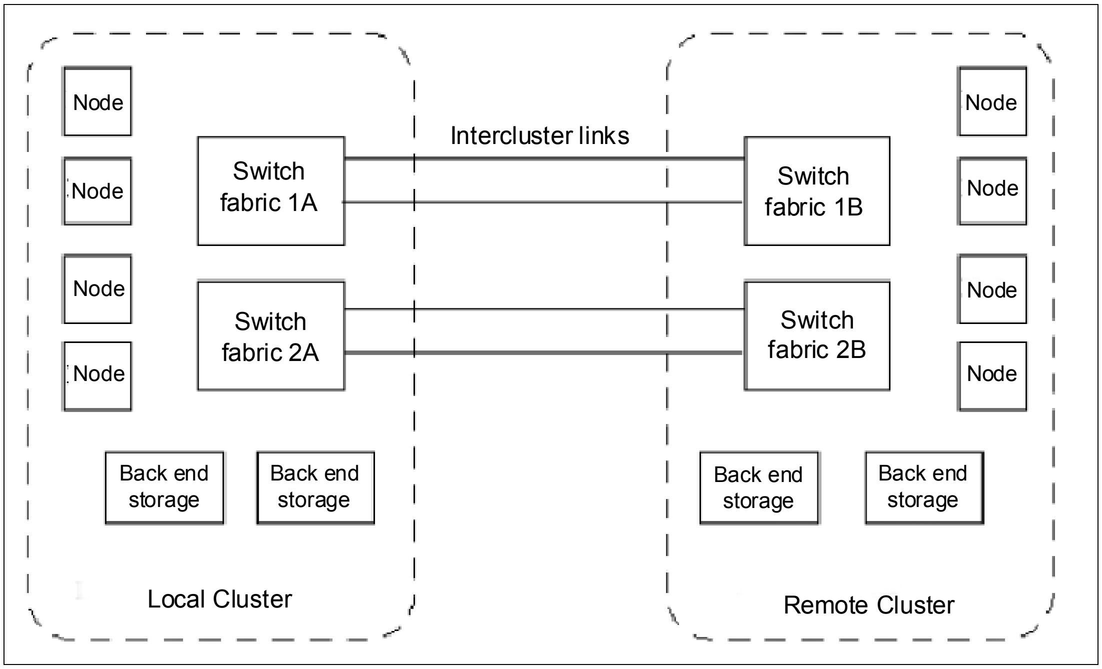


Figure 4-20 Metro Mirror connections

Figure 4-20 contains two redundant fabrics. Part of each fabric exists at the local clustered system and at the remote clustered system. No direct connection exists between the two fabrics.

Technologies for extending the distance between two SVC clustered systems can be broadly divided into two categories:

- FC extenders
- SAN multiprotocol routers

Due to the more complex interactions involved, IBM explicitly tests products of this class for interoperability with the SVC. You can obtain the current list of supported SAN routers in the supported hardware list on the SVC support website:

http://www.ibm.com/storage/support/2145

IBM has tested a number of FC extenders and SAN router technologies with the SVC. You must plan, install, and test FC extenders and SAN router technologies with the SVC so that the following requirements are met:

- The round-trip latency between sites must not exceed 80 ms (40 ms one way). For Global Mirror, this limit allows a distance between the primary and secondary sites of up to 8000 km (4970.96 miles) using a planning assumption of 100 km (62.13 miles) per 1 ms of round-trip link latency.
- The latency of long-distance links depends on the technology that is used to implement them. A point-to-point dark fiber-based link will typically provide a round-trip latency of 1ms per 100 km (62.13 miles) or better. Other technologies provide longer round-trip latencies, which affect the maximum supported distance.
- The configuration must be tested with the expected peak workloads.

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

  Figure 4-21 shows the amount of heartbeat traffic, in megabits per second, that is generated by various sizes of clustered systems.

SVC inter-cluster heartbeat traffic (Megabits per second):

		Cluster 2			
		2 nodes	**4 nodes**	**6 nodes**	**8 nodes**
Cluster 1	**2 nodes**	2.6	4.0	5.4	6.7
	4 nodes	4.0	5.5	7.1	8.6
	6 nodes	5.4	7.1	8.8	10.5
	8 nodes	6.7	8.6	10.5	12.4

Figure 4-21 Amount of heartbeat traffic

- These numbers represent the total traffic between the two clustered systems when no I/O is taking place to mirrored volumes. Half of the data is sent by one clustered system, and half of the data is sent by the other clustered system. The traffic is divided evenly over all available intercluster links. Therefore, if you have two redundant links, half of this traffic is sent over each link during fault-free operation.
- The bandwidth between sites must, at the least, be sized to meet the peak workload requirements, in addition to maintaining the maximum latency that has been specified previously. You must evaluate the peak workload requirement by considering the average write workload over a period of one minute or less, plus the required synchronization copy bandwidth.

  With no active synchronization copies and no write I/O disks in Metro Mirror or Global Mirror relationships, the SVC protocols operate with the bandwidth that is indicated in Figure 4-21. However, you can only determine the true bandwidth that is required for the link by considering the peak write bandwidth to volumes participating in Metro Mirror or Global Mirror relationships and adding it to the peak synchronization copy bandwidth.
- If the link between the sites is configured with redundancy so that it can tolerate single failures, you must size the link so that the bandwidth and latency statements continue to be true even during single failure conditions.
- The configuration is tested to simulate the failure of the primary site (to test the recovery capabilities and procedures), including eventual fail back to the primary site from the secondary.
- The configuration must be tested to confirm that any fail over mechanisms in the intercluster links interoperate satisfactorily with the SVC.
- The FC extender must be treated as a normal link.
- The bandwidth and latency measurements must be made by, or on behalf of, the client. They are *not* part of the standard installation of the SVC by IBM. Make these measurements during installation, and record the measurements. Testing must be repeated following any significant changes to the equipment that provides the intercluster link.

Global Mirror guidelines

Consider these guidelines:

- When using SVC Global Mirror, all components in the SAN must be capable of sustaining the workload that is generated by application hosts and the Global Mirror background copy workload. Otherwise, Global Mirror can automatically stop your relationships to protect your application hosts from increased response times. Therefore, it is important to configure each component correctly.
- Use a SAN performance monitoring tool, such as IBM Tivoli Storage Productivity Center, which allows you to continuously monitor the SAN components for error conditions and performance problems. This tool helps you detect potential issues before they affect your disaster recovery solution.
- The long-distance link between the two clustered systems must be provisioned to allow for the peak application write workload to the Global Mirror source volumes, plus the client-defined level of background copy.
- The peak application write workload ideally must be determined by analyzing the SVC performance statistics.
- Statistics must be gathered over a typical application I/O workload cycle, which might be days, weeks, or months, depending on the environment on which the SVC is used. These statistics must be used to find the peak write workload that the link must be able to support.
- Characteristics of the link can change with use; for example, latency can increase as the link is used to carry an increased bandwidth. The user must be aware of the link's behavior in such situations and ensure that the link remains within the specified limits. If the characteristics are not known, testing must be performed to gain confidence of the link's suitability.
- Users of Global Mirror must consider how to optimize the performance of the long-distance link, which depends on the technology that is used to implement the link. For example, when transmitting FC traffic over an IP link, it can be desirable to enable jumbo frames to improve efficiency.
- Using Global Mirror and Metro Mirror between the same two clustered systems is supported.
- Using Global Mirror and Metro Mirror between the SVC clustered system and IBM Storwize systems with a minimum code level of 6.3 is supported.
- It is supported for cache-disabled volumes to participate in a Global Mirror relationship; however, it not a preferred practice to do so.
- The **`gmlinktolerance`** parameter of the remote copy partnership must be set to an appropriate value. The default value is 300 seconds (five minutes), which is appropriate for most clients.
- During SAN maintenance, the user must choose to reduce the application I/O workload during the maintenance (so that the degraded SAN components are capable of the new workload); disable the **`gmlinktolerance`** feature; increase the **`gmlinktolerance`** value (meaning that application hosts might see extended response times from Global Mirror volumes); or stop the Global Mirror relationships.

  If the **`gmlinktolerance`** value is increased for maintenance lasting *x* minutes, it must only be reset to the normal value *x* minutes after the end of the maintenance activity.

  If **`gmlinktolerance`** is disabled during the maintenance, it must be re-enabled after the maintenance is complete.

- Global Mirror volumes must have their preferred nodes evenly distributed between the nodes of the clustered systems. Each volume within an I/O Group has a preferred node property that can be used to balance the I/O load between nodes in that group.

  Figure 4-22 shows the correct relationship between volumes in a Metro Mirror or Global Mirror solution.

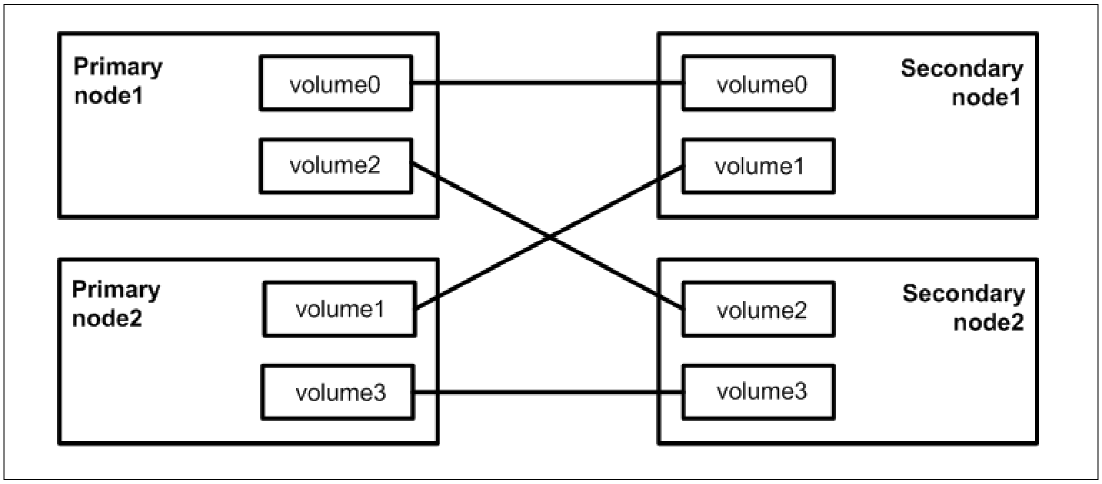


Figure 4-22 Correct volume relationship

- The capabilities of the storage controllers at the secondary clustered system must be provisioned to allow for the peak application workload to the Global Mirror volumes, plus the client-defined level of background copy, plus any other I/O being performed at the secondary site. The performance of applications at the primary clustered system can be limited by the performance of the backend storage controllers at the secondary clustered system to maximize the amount of I/O that applications can perform to Global Mirror volumes.
- It is necessary to perform a complete review before using Serial Advanced Technology Attachment (SATA) for Metro Mirror or Global Mirror secondary volumes. Using a slower disk subsystem for the secondary volumes for high-performance primary volumes can mean that the SVC cache might not be able to buffer all the writes, and flushing cache writes to SATA might slow I/O at the production site.
- Storage controllers must be configured to support the Global Mirror workload that is required of them. You can dedicate storage controllers to only Global Mirror volumes; configure the controller to guarantee sufficient quality of service (QoS) for the disks being used by Global Mirror; or ensure that physical disks are not shared between Global Mirror volumes and other I/O (for example, by not splitting an individual RAID array).
- MDisks within a Global Mirror storage pool must be similar in their characteristics, for example, RAID level, physical disk count, and disk speed. This requirement is true of all storage pools, but it is particularly important to maintain performance when using Global Mirror.
- When a consistent relationship is stopped, for example, by a persistent I/O error on the intercluster link, the relationship enters the consistent_stopped state. I/O at the primary site continues, but the updates are not mirrored to the secondary site. Restarting the relationship begins the process of synchronizing new data to the secondary disk. While this synchronization is in progress, the relationship is in the inconsistent_copying state. Therefore, the Global Mirror secondary volume is not in a usable state until the copy has completed and the relationship has returned to a Consistent state. For this reason, it is highly advisable to create a FlashCopy of the secondary volume before restarting the relationship. When started, the FlashCopy provides a consistent copy of the data, even

while the Global Mirror relationship is copying. If the Global Mirror relationship does not reach the Synchronized state (if, for example, the intercluster link experiences further persistent I/O errors), the FlashCopy target can be used at the secondary site for disaster recovery purposes.

- If you plan to use a Fibre Channel over IP (FCIP) intercluster link, it is extremely important to design and size the pipe correctly.

  Example 4-2 shows a best-guess bandwidth sizing formula.

Example 4-2 WAN link calculation example

```
Amount of write data within 24 hours times 4 to allow for peaks
Translate into MB/s to determine WAN link needed
Example:
250 GB a day
250 GB * 4 = 1 TB
24 hours * 3600 secs/hr. = 86400 secs
1,000,000,000,000/ 86400 = approximately 12 MB/s,
Which means OC3 or higher is needed (155 Mbps or higher)
```

- If compression is available on routers or WAN communication devices, smaller pipelines might be adequate. Note that workload is probably not evenly spread across 24 hours. If there are extended periods of high data change rates, consider suspending Global Mirror during that time frame.
- If the network bandwidth is too small to handle the traffic, the application write I/O response times might be elongated. For the SVC, Global Mirror must support short-term “Peak Write” bandwidth requirements.
- You must also consider the initial sync and resync workload. The Global Mirror partnership’s background copy rate must be set to a value that is appropriate to the link and secondary backend storage. The more bandwidth that you give to the sync and resync operation, the less workload can be delivered by the SVC for the regular data traffic.
- Do not propose Global Mirror if the data change rate will exceed the communication bandwidth or if the round-trip latency exceeds 80 - 120 ms. A greater than 80 ms round-trip latency requires SCORE/RPQ submission.

4.3.13 SAN boot support

The SVC supports SAN boot or startup for AIX, Microsoft Windows Server, and other operating systems. SAN boot support can change from time to time, so check the following websites regularly:

http://www.ibm.com/systems/storage/software/virtualization/svc/interop.html
http://www.ibm.com/systems/support/storage/ssic/interoperability.wss

4.3.14 Data migration from a non-virtualized storage subsystem

Data migration is an extremely important part of an SVC implementation. Therefore, you must accurately prepare a data migration plan. You might need to migrate your data for one of these reasons:

- To redistribute workload within a clustered system across the disk subsystem
- To move workload onto newly installed storage
- To move workload off old or failing storage, ahead of decommissioning it
- To move workload to rebalance a changed workload

- To migrate data from an older disk subsystem to SVC-managed storage
- To migrate data from one disk subsystem to another disk subsystem

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

We can define data migration as belonging to three groups:

- Based on operating system Logical Volume Manager (LVM) or commands
- Based on special data migration software
- Based on the SVC data migration feature

With data migration, apply the following guidelines:

- Choose which data migration method best fits your operating system platform, your kind of data, and your SLA.
- Check the interoperability matrix for the storage subsystem to which your data is being migrated:

  http://www.ibm.com/systems/storage/software/virtualization/svc/interop.html
- Choose where you want to place your data after migration in terms of the storage pools that relate to a specific storage subsystem tier.
- Check whether enough free space or extents are available in the target storage pool.
- Decide if your data is critical and must be protected by a volume mirroring option or if it must be replicated in a remote site for disaster recovery.
- Prepare offline all of the zone and LUN masking and host mappings that you might need, to minimize downtime during the migration.
- Prepare a detailed operation plan so that you do not overlook anything at data migration time.
- Run a data backup before you start any data migration. Data backup must be part of the regular data management process.
- You might want to use the SVC as a data mover to migrate data from a non-virtualized storage subsystem to another non-virtualized storage subsystem. In this case, you might have to add additional checks that relate to the specific storage subsystem to which you want to migrate. Be careful using slower disk subsystems for the secondary volumes for high-performance primary volumes because the SVC cache might not be able to buffer all the writes and flushing cache writes to SATA might slow I/O at the production site.

4.3.15 SAN Volume Controller configuration backup procedure

Save the configuration externally when changes, such as adding new nodes and disk subsystems, have been performed on the clustered system. Saving the configuration is a crucial part of SVC management, and various methods can be applied to back up your SVC configuration. The preferred practice is to implement an automatic configuration backup by applying the configuration backup command. We describe this command for the CLI in Chapter 9, "SAN Volume Controller operations using the command-line interface" on page 165, and we describe the GUI operation in Chapter 10, "SAN Volume Controller operations using the GUI" on page 327.

4.4 Performance considerations

Although storage virtualization with the SVC improves flexibility and provides simpler management of a storage infrastructure, it can also provide a substantial performance advantage for various workloads. The SVC caching capability and its ability to stripe volumes across multiple disk arrays are the reasons why performance improvement is significant when implemented with midrange disk subsystems. This technology is often only provided with high-end enterprise disk subsystems.

> **Tip:** Technically, almost all storage controllers provide both striping (RAID 5 or RAID 10) and a form of caching. The real benefit is the degree to which you can stripe the data across all MDisks in a storage pool and therefore have the maximum number of active spindles at one time. The caching is secondary. The SVC provides additional caching to the caching that midrange controllers provide (usually a couple of GB), whereas enterprise systems have much larger caches.

To ensure the wanted performance and capacity of your storage infrastructure, undertake a performance and capacity analysis to reveal the business requirements of your storage environment. When this analysis is done, you can use the guidelines in this chapter to design a solution that meets the business requirements.

When discussing performance for a system, always identify the bottleneck and, therefore, the limiting factor of a given system. You must also consider the component for whose workload you identify a limiting factor. The component might not be the same component that is identified as the limiting factor for other workloads.

When designing a storage infrastructure with SVC or implementing SVC in an existing storage infrastructure, you must consider the performance and capacity of the SAN, the SVC, the disk subsystems, and the known or expected workload.

The SVC is designed to handle large quantities of multiple paths from the backend storage.

In most cases, the SVC can improve performance, especially on mid-sized to low-end disk subsystems, older disk subsystems with slow controllers, or uncached disk systems, for these reasons:

- The SVC can stripe across disk arrays, and it can stripe across the entire set of supported physical disk resources.
- Each SVC 2145-DH8 node has 32 GB of base cache and 64 GB when the second processor and cache upgrade are added for Real-time Compression, providing a total of 64 GB/128 GB per I/O Group. (An 8 node cluster contains 512 GB base cache/10240 GB with RtC.)

The SVC is capable of providing automated performance optimization of hot spots by using flash drives and Easy Tier.

4.4.1 SAN

The current available SVC models are:

- 2145-CG8
- 2145-DH8

Both of these models connect to 2 GBps, 4 GBps, 8 GBps, and 16 GBps switches. From a performance point of view, connecting the SVC to 8 GBps or 16 GBps switches is better.

Correct zoning on the SAN switch brings security and performance together. Implement a dual HBA approach at the host to access the SVC.

4.4.2 Disk subsystems

Each MDisk presented to SVC should consist of a single RAID group, or MDisk, of a single type of drive, from the underlying storage controller.

Advanced features, such as Disk Tiering, should be disabled on the underlying storage controller as they will skew the results of the performance of the MDisk expected by SVC.

Storwize family controllers should not use MDisk pooling, but should present a single MDisk, as a single pool, as a single volume, as the Storage Pool Balancing feature will affect the way the MDisk behaves to SVC.

4.4.3 Cache

The SVC clustered system is scalable up to eight nodes, and the performance is nearly linear when adding more nodes into an SVC clustered system.

The large cache and advanced cache management algorithms within SVC allow it to improve on the performance of many types of underlying disk technologies. The SVC's capability to manage, in the background, the destaging operations that are incurred by writes (in addition to still supporting full data integrity) assists with SVC's capability in achieving very good database performance.

There are a number of changes to how SVC uses its cache in the 7.3 code level. The cache is separated into two layers, an upper cache, and a lower cache.

Figure 4-23 on page 101 shows the separation of the upper and lower cache.

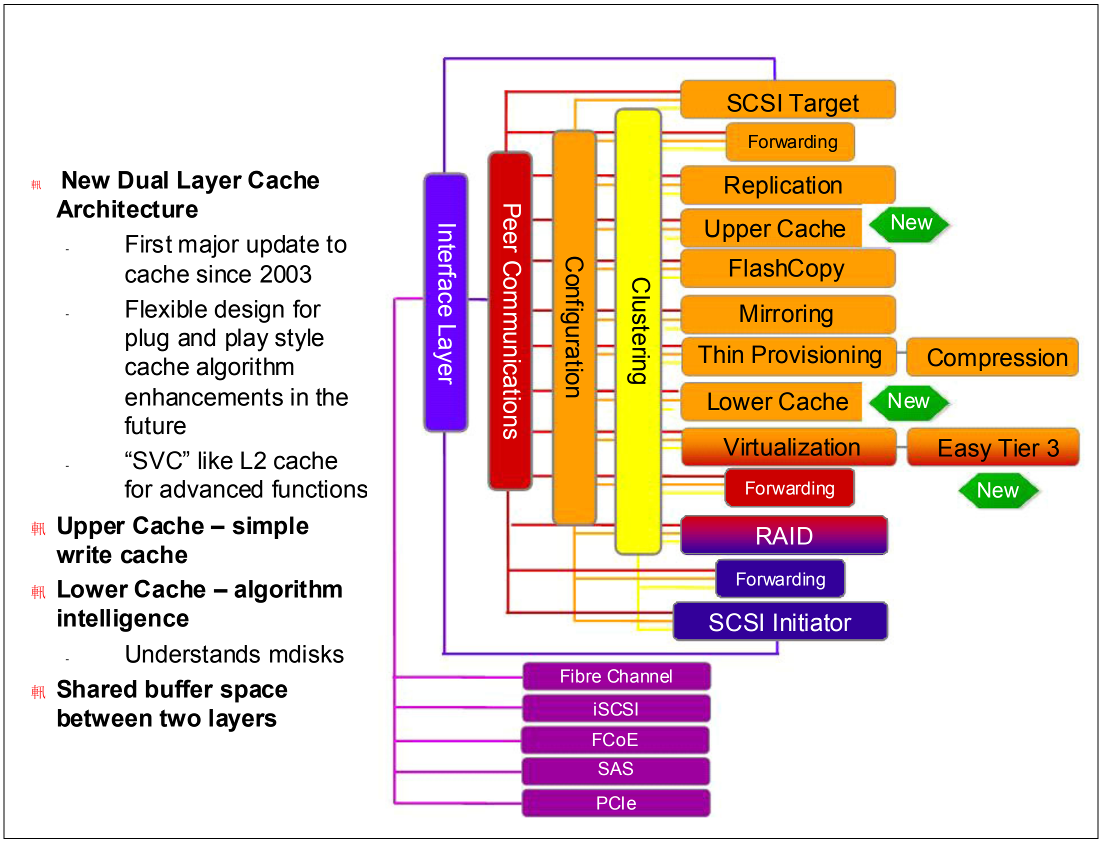


Figure 4-23 Separation of upper and lower cache

The upper cache delivers the following functionality allowing SVC to streamline data write performance:

- Provides fast write response times to the host by being as high up in the I/O stack as possible
- Provides partitioning

The lower cache delivers the following additional functionality:

- Ensure write cache between two nodes is in sync
- Cache partitioning to ensure that a slow back end cannot consume the entire cache
- A destage algorithm that adapts to the amount of data and the back-end performance
- Provide read caching and prefetching

Combined together, the two levels of cache also deliver the following functionality:

- Pin data when LUN goes offline
- Enhanced statistics for TPC while maintaining compatibility with an earlier version
- Trace for debugging
- Reporting medium errors
- Correctly resync cache and provide the atomic write functionality
- Ensure that other partitions continue operation where one partition becomes 100% full of pinned data
- Support Fast-write (2 and 1-way), flush-through, and write-through
- Integrate with T3 recovery procedures

- Support 2-way operation
- Support none, read-only, and read/write as user exposed caching policies
- Support flush-when-idle
- Support expanding cache as more memory becomes available to the platform
- Credit throttling to avoid I/O skew and fairness/balanced I/O between the two nodes of the I/O group
- Enable switching of the preferred node without needing to move volumes between I/O Groups

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

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

IBM SAN Volume Controller 4.2.1 Cache Partitioning, REDP-4426, explains the SVC cache partitioning capability:

http://www.redbooks.ibm.com/abstracts/redp4426.html?Open

4.4.4 Port configuration

With the introduction of the 2145-DH8 nodes and the ability to have up to 12 Fibre Channel ports per node, there are a number of different options that are valid for attaching storage to the SVC cluster. Refer to section 4.3.2, “SAN zoning and SAN connections” on page 74 for configurations based on a single, dual, or triple 4 Port Fibre Channel HBA adapter port allocation recommendations:

- If you require a high throughput environment greater than 10 GBps, zoning all ports on the disk backend storage to all ports on the SVC nodes in the cluster is a valid option.
- If you are looking to achieve the lowest latency storage environment, the allocation of four ports per node to intercluster traffic and inter I/O group traffic is the best recommendation. Each of the four ports should be zoned so that it only sees one other port in the same I/O Group. The same ports should be used for the remote traffic zoning. The remaining eight ports per node can be separated into host and storage attachment roles.

> **Note:** A port should be used for the same purpose/attached to the same switch/zones for every node.
>
> The only exception to this is when using mixed hardware types. At which point the lowest ports should be used for the same purposes and the remaining ports can be allocated as required. (The lowest ports are the lowest numbered adapter slots, or the right-most bits in the mask).

Although virtualization with the SVC provides a great deal of flexibility, it does not diminish the necessity to have a SAN and disk subsystems that can deliver the wanted performance.

Essentially, SVC performance improvements are gained by having as many MDisks as possible, therefore creating a greater level of concurrent I/O to the backend without overloading a single disk or array.

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

- Creating a storage pool
- Creating volumes
- Connecting to or configuring hosts that must receive disk space from an SVC clustered system

You can obtain more detailed information about performance and preferred practices for the SVC in *SAN Volume Controller Best Practices and Performance Guidelines*, SG24-7521:

`http://www.redbooks.ibm.com/abstracts/sg247521.html?Open`

4.4.5 Performance monitoring

Performance monitoring must be an integral part of the overall IT environment.

This topic is covered in more detail in the following chapter: Chapter 8, “Performance data and statistics gathering” on page 153.

For the SVC, as for the other IBM storage subsystems, the official IBM product to collect performance statistics and supply a performance report is the IBM Tivoli Storage Productivity Center.

You can obtain more information about using the IBM Tivoli Storage Productivity Center to monitor your storage subsystem in *SAN Storage Performance Management Using Tivoli Storage Productivity Center*, SG24-7364:

`http://www.redbooks.ibm.com/abstracts/sg247364.html?Open`

5

SAN Volume Controller Easy Tier

This chapter describes the history of SAN Volume Controller Easy Tier, and discusses the changes in versions and the enhancements in version 7.3. In addition, we assess the capabilities of the new functionality, compare usage cases, and identify configuration and deployment considerations that should be taken into account during the planning stage of your SVC deployment.

In the following chapter, our intent is to provide only a basic technical overview and focus on the benefits with the new version of Easy Tier. More details for planning and configuration are available in the following IBM Redbooks publications:

- *Implementing IBM Easy Tier with IBM Real-time Compression*, TIPS1072
- *IBM System Storage SAN Volume Controller Best Practices and Performance Guidelines*, SG24-7521
- *IBM DS8000 Easy Tier*, REDP-4667 (this concept is similar to SAN Volume Controller Easy Tier)

5.1 SVC Easy Tier history

IBM Easy Tier is a performance function that automatically and non-disruptively migrates frequently accessed data from magnetic media to solid-state drives (SSD or flash drives). In that way, the most frequently accessed data is stored on the fastest storage tier, and the overall performance is improved.

SVC has benefited from the software development work for the IBM System Storage DS8000 product, in which there have been six versions of Easy Tier. Of those versions, version 1 and 3 have been implemented within SVC.

The *first generation of Easy Tier* introduced automated storage performance management by efficiently boosting enterprise-class performance with flash drives (SSD) and automating storage tiering from enterprise-class drives to flash drives, thus optimizing flash deployments with minimal costs. It also introduced dynamic volume relocation and dynamic extent pool merge.

The *second generation of Easy Tier was only implemented within DS8000.*

The *third generation of Easy Tier* introduces further enhancements that provide automated storage performance and storage economics management across all three drive tiers (Flash, Enterprise, and Nearline storage tiers) as outlined in Figure 5-1. It allows you to consolidate and efficiently manage more workloads on a single SVC system. It also introduces support for *Storage Pool Balancing* in homogeneous pools.

5.1.1 New features in Easy Tier 3

The enhancements of Easy Tier include:

- Support for three tiers of disk or a mixture of any two tiers.
- Storage Pool Balancing.
- Enhancements to the STAT tool including additional graphing from the STAT Utility.

Figure 5-1 shows the supported easy tier pools now available in Easy Tier 3.

Tier 0	Tier 1	Tier 2
Three Tier Pools :		
SSD	Enterprise	Nearline
Two Tier Pools :		
SSD	Enterprise	-
SSD	Nearline	-
-	Enterprise	Nearline
Single tier pools :		
SSD	-	-
-	Enterprise	-
-	-	Nearline

Figure 5-1 Easy Tier three multi-level tiering

By default, Easy Tier is enabled on any pool that contains more than one class of disk drive. Easy Tier manages extent migration using the following four processes:

- Promote/Swap

  Moves the relevant hot extents to higher performing tier

- Warm Demote

  – Prevents performance overload of a tier by demoting a warm extent to the lower tier

  – Triggered when bandwidth or IOPS exceeds predefined threshold

- Cold Demote

  Coldest data moved to lower HDD tier

- Expanded Cold Demote

  Demotes appropriate sequential workloads to the lowest tier to better utilize Nearline bandwidth

- Storage Pool Balancing

  – Redistribute extents within a tier to balance utilization across MDisks for maximum performance

  – Either move or swap

> **Note:** Extent migrations occur only between adjacent tiers.

Figure 5-2 shows the Easy Tier process for extent migration.

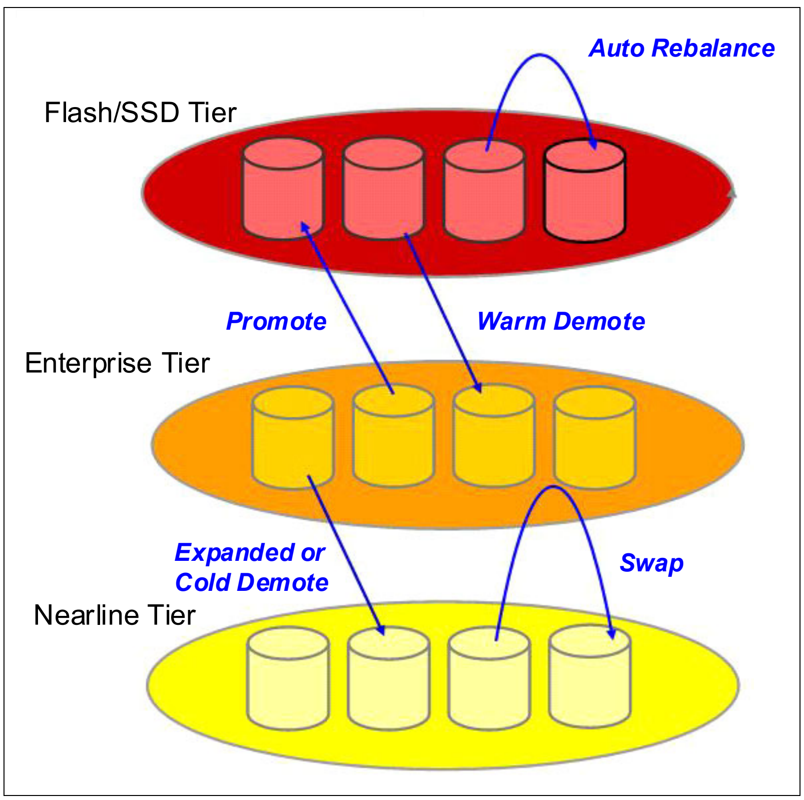


Figure 5-2 Easy Tier extent migration

5.1.2 Storage Pool Balancing

Storage Pool Balancing is a new feature within the 7.3 code, which while associated with Easy Tier, operates independently of Easy Tier and does not require an Easy Tier license. This feature assesses the extents that are written in a pool and balances them automatically across all MDisks within the pool. This process works in conjunction with Easy Tier when multiple classes of disks exist in a single pool.

The process will automatically balance existing data when new MDisks are added into an existing pool even if the pool only contains a single type of drive.

> **Note:** Storage Pool Balancing can be used to balance extents when mixing different size disks of the same performance tier. For example, when adding larger capacity drives to pool with smaller capacity drives of the same class, Storage Pool Balancing redistributes the extents to take advantage of the full capacity of the new MDisks.

5.2 Performance and monitoring considerations

In this section, we briefly explain the impact of Easy Tier on performance and introduce its monitoring tools.

5.2.1 Considerations for optimal performance

With the availability of SVC code 7.3, SVC now has the ability to classify and identify performance profiles based on the category of drive within an MDisk group. SVC recognizes three types of disk: Flash drives, Enterprise drives, and Nearline drives.

However, when a new MDisk is added to SVC, SVC does not automatically classify the MDisk by the type of drive that the MDisk consists of. You need to manually select the MDisk and choose the type of drive and allocate it to the MDisk.

In Figure 5-3 on page 109, `mdisk29` is a flash drive (SSD) allocated from an IBM Storwize V7000 storage controller to the SVC. When you right click the required MDisk, you can choose the option to “Select Tier”.

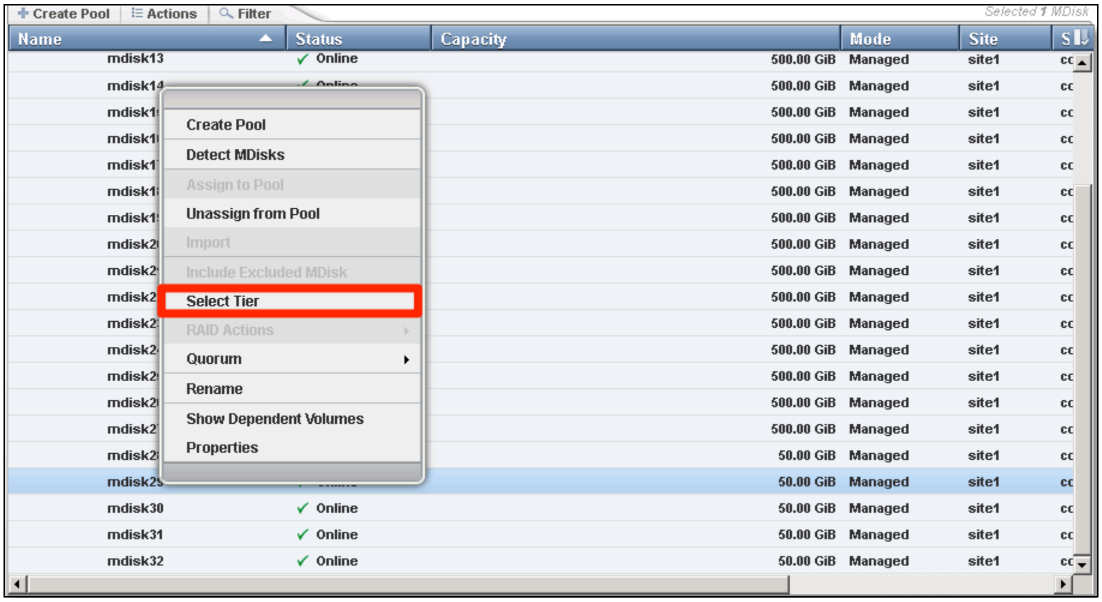


Figure 5-3 Shows the option to select a tier for a specific MDisk

Figure 5-4 shows the options that are available for each MDisk to define the class of drive within the MDisk.

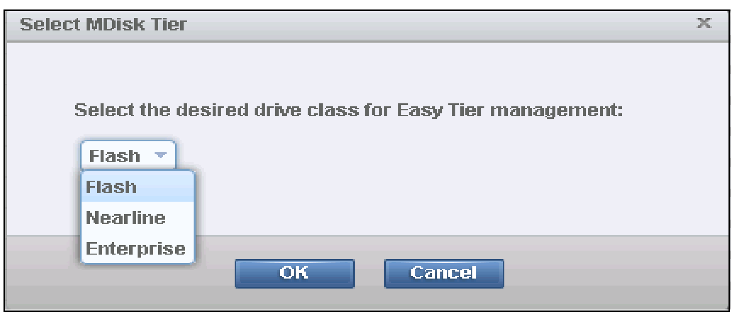


Figure 5-4 Shows the class of drives available within SVC MDisks

Note: When using an IBM Storwize family backend storage controller that supports Storage Pool Balancing, you must disable Storage Pool Balancing by presenting MDisks in the following way:

Single Storwize MDisk → Single Storwize Pool → Single Storwize Volume → Single SVC MDisk

Failure to do so means that Storage Pool Balancing on MDisks within SVC competes with Storage Pool Balancing on the Storwize controller, which causes performance degradation at both levels.

5.2.2 Monitoring tools

The IBM Storage Tier Advisor Tool (STAT) is a Windows console application that analyzes heat data files produced by Easy Tier and produces a graphical display of the amount of “hot” data per volume and predictions of how additional flash drive (SSD) capacity, Enterprise Drive, and Nearline Drive could benefit performance for the system and by storage pool.

Heat data files are produced approximately once a day (that is, every 24 hours) when Easy Tier is active on one or more storage pools and summarizes the activity per volume since the prior heat data file was produced. On SVC and Storwize serial products, the heat data file is in the /dumps directory on the configuration node and is named “dpa_heat.node_name.time_stamp.data”.

Any existing heat data file is erased when it has existed for longer than seven days. The file must be offloaded by the user and Storage Tier Advisor Tool invoked from a Windows command prompt console with the file specified as a parameter. The user can also specify the output directory. The Storage Tier Advisor Tool creates a set of HTML files and the user can then open the resulting “index.html” in a browser to view the results.

Updates to the STAT tool for SVC 7.3 have added additional capability for reporting. As a result, when the STAT tool is run on a heat map file an additional three CSV files are created and placed in the Data_files directory.

The IBM STAT tool can be downloaded from the IBM Support website:

http://www.ibm.com/support/docview.wss?uid=ssg1S4000935

Figure 5-5 shows the CSV files highlighted in the Data_files directory after running the stat tool over an SVC heatmap.

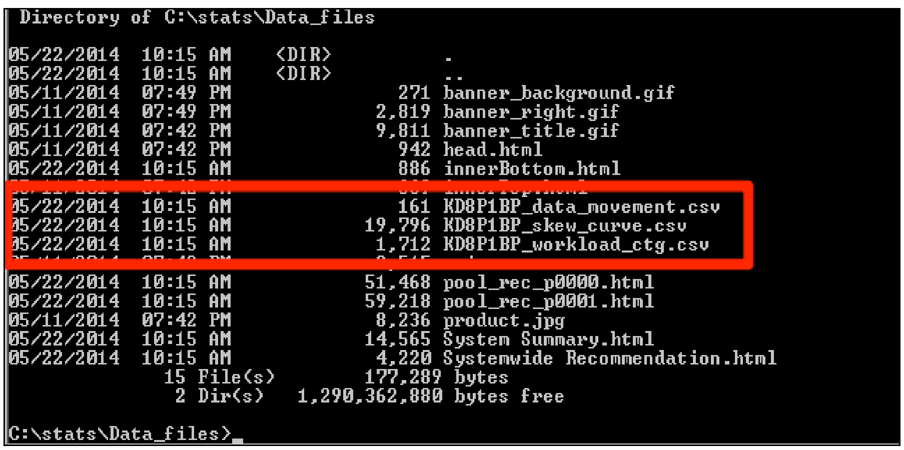


Figure 5-5 CSV files created by the STAT Tool for Easy Tier

In addition to the STAT tool, SVC 7.3 code now has an additional utility, which is a Microsoft SQL file for creating additional graphical reports of the workload that Easy Tier is performing. The IBM STAT Charting Utility takes the output of the three CSV files and turns them into graphs for simple reporting.

The three new graphs display:

- Workload Categorization

  New workload visuals help you compare activity across tiers within and across pools to help determine optimal drive mix for current workloads. The output is illustrated in Figure 5-6 on page 111.

1	index	pool_id	lun_id	tier	active_cap	inactive_ca	active_larg	low_activit	active_read	inactive_re	active_larg	low_activit	active_writ	inactive_wr	active_larg	low_activit	active_kib	inactive_ki
3500	3499	P0	0x1000	0	14	0	0	36	25.35	0	0	51.8	1.38	0	0	3.56	2703.36	0
3501	3500	P0	0x1000	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3502	3501	P0	0x1000	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3503	3502	P0	0x1001	0	11	0	0	39	19.36	0	0	55.9	1.2	0	0	3.4	2205.82	0
3504	3503	P0	0x1001	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3505	3504	P0	0x1001	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3506	3505	P0	0x1002	0	9	0	0	41	14.44	0	0	55.17	1.23	0	0	4.83	1286.81	0
3507	3506	P0	0x1002	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3508	3507	P0	0x1002	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3509	3508	P0	0x1003	0	32	0	0	18	31.64	0	0	18.78	15.89	0	0	7.35	5258.99	0
3510	3509	P0	0x1003	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3511	3510	P0	0x1003	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3860	3859	P0	0x1408	0	0	132	0	0	0	0	0	0	0	0	0	0	0	0
3861	3860	P0	0x1408	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3862	3861	P0	0x1408	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3863	3862	P0	0x1409	0	0	132	0	0	0	0	0	0	0	0	0	0	0	0
3864	3863	P0	0x1409	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3865	3864	P0	0x1409	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17990	17989	P0	0x7000	0	0	0	0	5	0	0	0	6.7	0	0	0	0.32	0	0
17991	17990	P0	0x7000	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17992	17991	P0	0x7000	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17993	17992	P0	0x7001	0	0	0	0	5	0	0	0	6	0	0	0	0.28	0	0
17994	17993	P0	0x7001	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17995	17994	P0	0x7001	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17996	17995	P0	0x7002	0	0	0	0	5	0	0	0	6.62	0	0	0	0.27	0	0

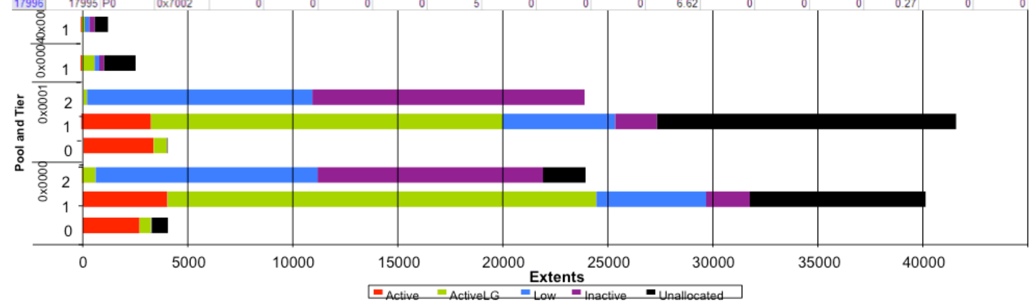


Figure 5-6 STAT Charting Utility Workload Categorization report

- Daily Movement Reporting

  New Easy Tier summary report every 24 hours illustrating data migration activity (5-min. intervals) can help visualize migration types and patterns for current workloads. The output is illustrated in Figure 5-7.

1	index	dss	lun_id	extent	time	type	typedescri	source_rar	target_rank
2	1	IBM.2107-:	0x2100	8	2013-02-0	2	SWAP	R71	R33
3	2	IBM.2107-:	0x2377	15	2013-02-0	2	SWAP	R33	R71
4	3	IBM.2107-:	0x2100	27	2013-02-0	2	SWAP	R71	R44
5	4	IBM.2107-:	0x239a	14	2013-02-0	2	SWAP	R44	R71
6	5	IBM.2107-:	0x2100	2	2013-02-0	2	SWAP	R71	R20
7	6	IBM.2107-:	0x216a	14	2013-02-0	2	SWAP	R20	R71
8	7	IBM.2107-:	0x253f	11	2013-02-0	2	SWAP	R72	R46
9	8	IBM.2107-:	0x2311	0	2013-02-0	2	SWAP	R46	R72
10	9	IBM.2107-:	0x2797	11	2013-02-0	4	MOVE	R69	R67
11	10	IBM.2107-:	0x2797	6	2013-02-0	4	MOVE	R69	R67
12	11	IBM.2107-:	0x2797	5	2013-02-0	4	MOVE	R69	R67
13	12	IBM.2107-:	0x2797	1	2013-02-0	4	MOVE	R69	R67

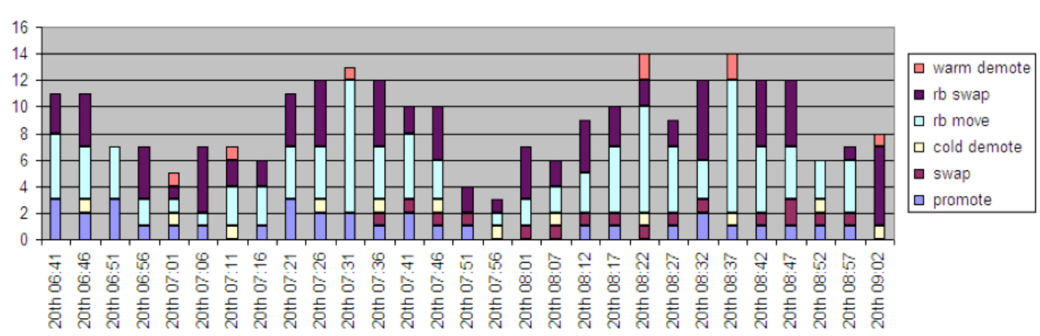
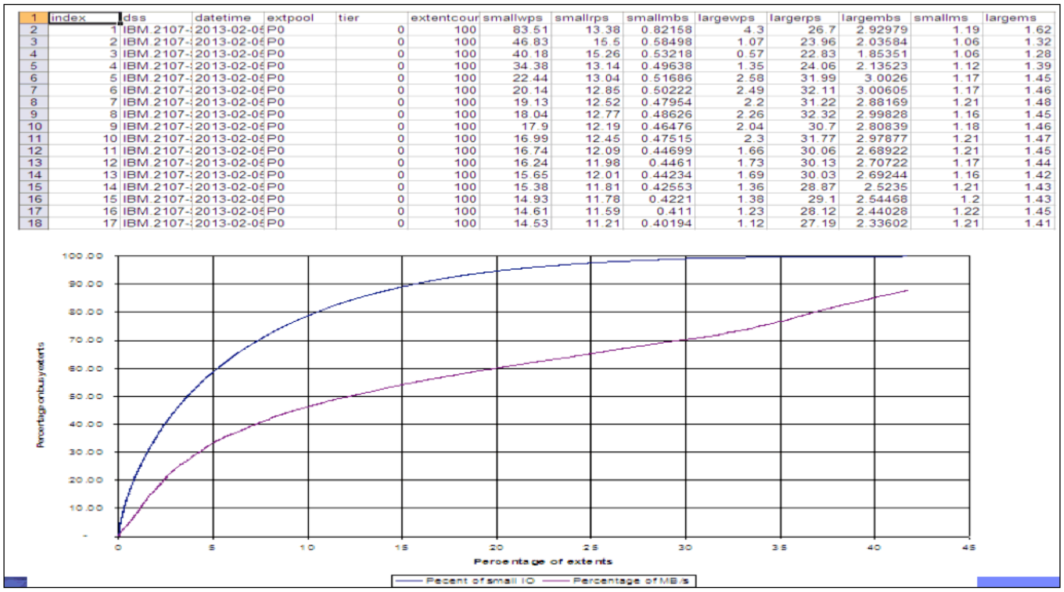


Figure 5-7 STAT Charting Utility Daily Summary report

- Workload Skew

  This shows skew of all workloads across the system in a graph to help clients visualize and accurately tier configurations when adding capacity or a new system. The output is illustrated in Figure 5-8.



1	index	dss	datetime	extpool	tier	extentcour	smallwps	smallrps	smallmbs	largewps	largerps	largembs	smallms	largems
2	1	IBM.2107-:	2013-02-05	P0	0	100	83.51	13.38	0.82158	4.3	26.7	2.92979	1.19	1.62
3	2	IBM.2107-:	2013-02-05	P0	0	100	46.83	15.5	0.58498	1.07	23.96	2.03584	1.06	1.32
4	3	IBM.2107-:	2013-02-05	P0	0	100	40.18	15.26	0.53218	0.57	22.83	1.85351	1.06	1.28
5	4	IBM.2107-:	2013-02-05	P0	0	100	34.38	13.14	0.49638	1.35	24.06	2.13523	1.12	1.39
6	5	IBM.2107-:	2013-02-05	P0	0	100	22.44	13.04	0.51686	2.58	31.99	3.0026	1.17	1.45
7	6	IBM.2107-:	2013-02-05	P0	0	100	20.14	12.85	0.50222	2.49	32.11	3.00605	1.17	1.46
8	7	IBM.2107-:	2013-02-05	P0	0	100	19.13	12.52	0.47954	2.2	31.22	2.88169	1.21	1.48
9	8	IBM.2107-:	2013-02-05	P0	0	100	18.04	12.77	0.48626	2.26	32.32	2.99828	1.16	1.45
10	9	IBM.2107-:	2013-02-05	P0	0	100	17.9	12.19	0.46476	2.04	30.7	2.80839	1.18	1.46
11	10	IBM.2107-:	2013-02-05	P0	0	100	16.99	12.45	0.47515	2.3	31.77	2.97877	1.21	1.47
12	11	IBM.2107-:	2013-02-05	P0	0	100	16.74	12.09	0.44699	1.66	30.06	2.68922	1.21	1.45
13	12	IBM.2107-:	2013-02-05	P0	0	100	16.24	11.98	0.4461	1.73	30.13	2.70722	1.17	1.44
14	13	IBM.2107-:	2013-02-05	P0	0	100	15.65	12.01	0.44234	1.69	30.03	2.69244	1.16	1.42
15	14	IBM.2107-:	2013-02-05	P0	0	100	15.38	11.81	0.42553	1.36	28.87	2.5235	1.21	1.43
16	15	IBM.2107-:	2013-02-05	P0	0	100	14.93	11.78	0.4221	1.38	29.1	2.54468	1.2	1.43
17	16	IBM.2107-:	2013-02-05	P0	0	100	14.61	11.59	0.411	1.23	28.12	2.44028	1.22	1.45
18	17	IBM.2107-:	2013-02-05	P0	0	100	14.53	11.21	0.40194	1.12	27.19	2.33602	1.21	1.41

Figure 5-8 STAT Charting Utility Workload Skew report

The STAT Charting Utility can be downloaded from the IBM support website:

http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS5251

6

SAN Volume Controller initial configuration

In this chapter, we discuss the following topics:

- Managing the cluster
- SAN Volume Controller Hardware Management Console
- SAN Volume Controller initial configuration steps

6.1 Managing the cluster

You can manage the SAN Volume Controller in many ways. The following methods are the most common:

- Using the SAN Volume Controller Management graphical user interface (GUI)
- Using a PuTTY-based SAN Volume Controller command-line interface (CLI)
- Using IBM Tivoli Storage Productivity Center (TPC)

Figure 6-1 shows the various ways to manage a SAN Volume Controller cluster.

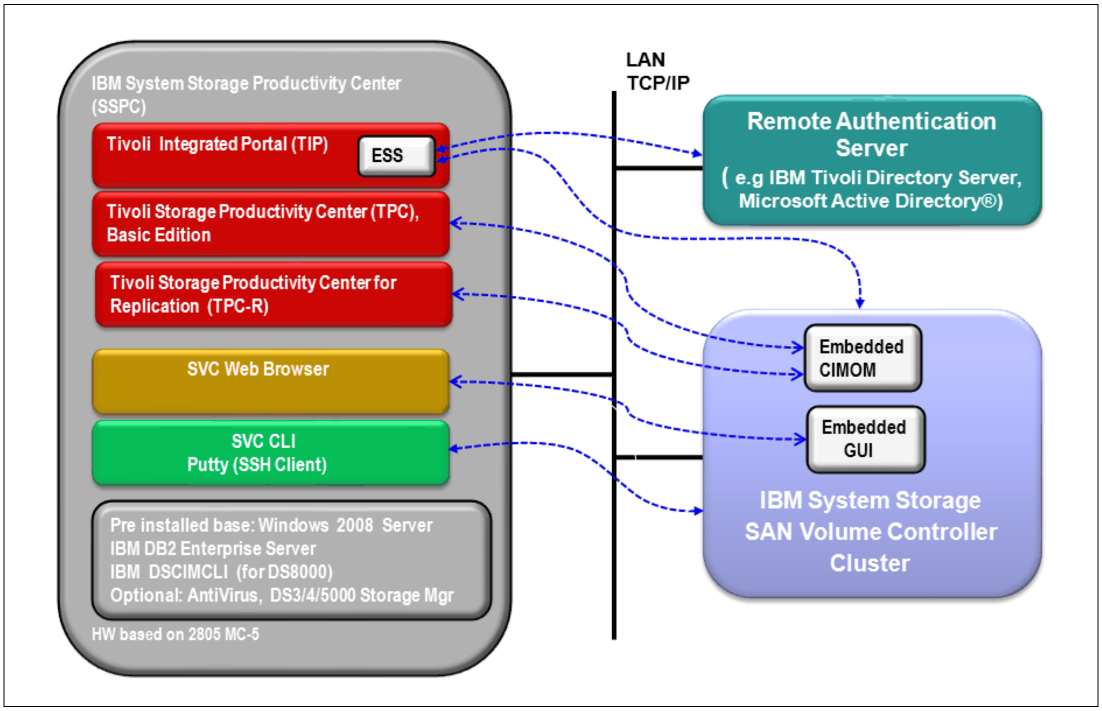


Figure 6-1 SAN Volume Controller cluster management

Note that you have full management control of the SAN Volume Controller regardless of which method you choose. IBM Tivoli Storage Productivity Center is a robust software product with various functions that needs to be purchased separately.

If you already have a previously installed SAN Volume Controller cluster in your environment, it is possible that you are using the SAN Volume Controller Console (Hardware Management Console - HMC). You can still use it together with your IBM Tivoli Storage Productivity Center. When using the specific, retail product called IBM System Storage Productivity Center (SSPC, which is not offered anymore), you can only log in to your SAN Volume Controller from one of them at a time.

If you decide to manage your SAN Volume Controller cluster with the SAN Volume Controller CLI, it does not matter if you are using the SAN Volume Controller Console or IBM Tivoli Storage Productivity Center server because the SAN Volume Controller CLI is located on the cluster and accessed through the Secure Shell (SSH), which can be installed anywhere.

6.1.1 Network requirements for SAN Volume Controller

To plan your installation, consider the TCP/IP address requirements of the IBM SAN Volume Controller cluster and the requirements for the SAN Volume Controller cluster to access other services. You must also plan the address allocation and the Ethernet router, gateway, and firewall configuration to provide the required access and network security.

Figure 6-2 shows the TCP/IP ports and services that are used by the SAN Volume Controller.

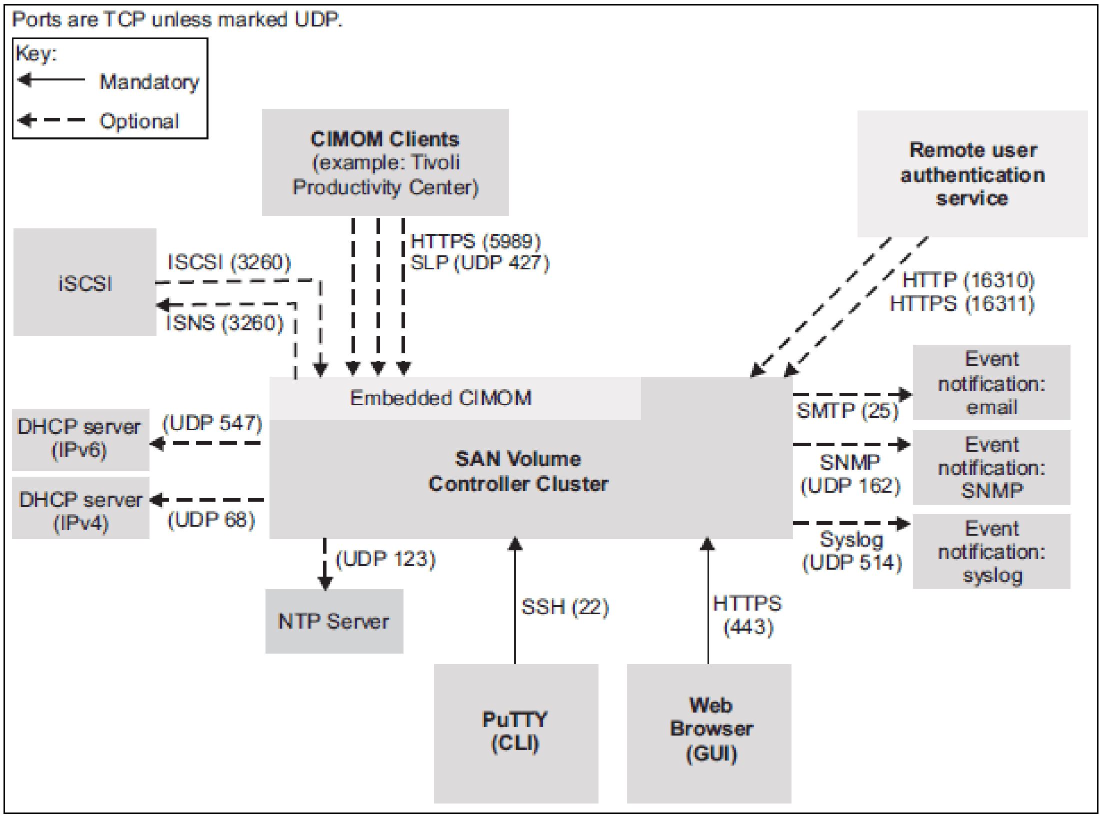


Figure 6-2 TCP/IP ports

For more information about TCP/IP prerequisites, see Chapter 4, "Planning and configuration" on page 57.

6.1.2 Prerequisites

Ensure that the SAN Volume Controller nodes are physically installed and that Ethernet and Fibre Channel (FC) connectivity has been correctly configured. For information about physical connectivity to the SAN Volume Controller, see Chapter 4, "Planning and configuration" on page 57.

Before configuring the cluster, ensure that the following information is available:

- License

  The license indicates whether the client is permitted to use FlashCopy, Metro Mirror, and the Real-time Compression features. It also indicates how much capacity the client is licensed to virtualize.

- For IPv4 addressing:
  - Cluster IPv4 addresses: These addresses include one address for the cluster and another address for the service address.
  - IPv4 subnet mask.
  - Gateway IPv4 address.
- For IPv6 addressing:
  - Cluster IPv6 addresses: These addresses include one address for the cluster and another address for the service address.
  - IPv6 prefix.

Gateway IPv6 address.

6.1.3 Initial configuration of the SAN Volume Controller cluster

For our initial configuration, we are using the following hardware:

- 2 x 2145-DH8 nodes.
- 1 x 32 GB additional memory for each 2145-DH8 node (total 64 GB of memory per node).
- 1 x processor additional for each 2145-DH8 node (total two processors per node).
- 1 x Real-time Compression accelerator card for each 2145-DH8 node.
- 1 x four port HBA adapter in each 2145-DH8 node.
- 2 x SAN Switches (for a redundant SAN fabric).

The back-end storage consists of Storwize V7000 block storage arrays.

The first step is to connect a PC or notebook to the Technician port on the rear of the SAN Volume Controller (SVC) node. See Figure 6-3 on page 117 for the Technician port. The Technician port provides a DHCP IP address V4, so you must ensure that your PC or notebook is configured for DHCP. The “default” IP address for a new node is 192.168.0.1.

The 2145-DH8 does not provide IPv6 IP addresses for the Technician port.

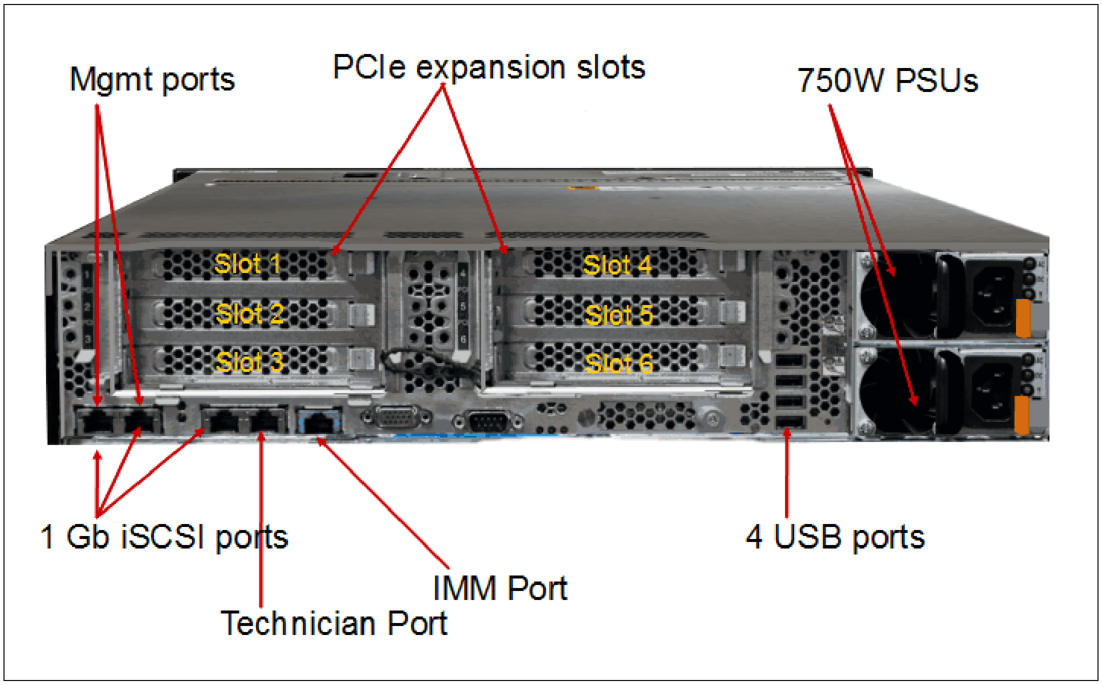


Figure 6-3 Rear of 2145-DH8

Nodes: During the initial configuration, you will see certificate warnings, since the 2145 certificates are self issued. Accept these warnings as they are not harmful.

When your PC or notebook is connected to the Technician port, and you have validated that you have an IP v4 DCHP address, for example, 192.168.0.12 (first IP address the SVC node will assign), open a supported browser, which should automatically redirect you to 192.168.0.1 and the initial configuration of the cluster can start.

Figure 6-4 on page 118 shows the Welcome window and starts the wizard that allows you to configure a new system or expand an existing system.

Figure 6-4 Welcome window

This chapter focuses on setting up a new system, so we mark **Yes** and click **Next**.

> **Note:** If you are adding 2145-DH8 nodes to an existing system, ensure that the existing systems are running code level 7.3 or higher, as the 2145-DH8 only supports code level 7.3 or higher.

The next window will ask you to set an IP address for the cluster. You can choose between IPv4 or IPv6 address. In Figure 6-5 on page 119 we have set an IPv4 address.

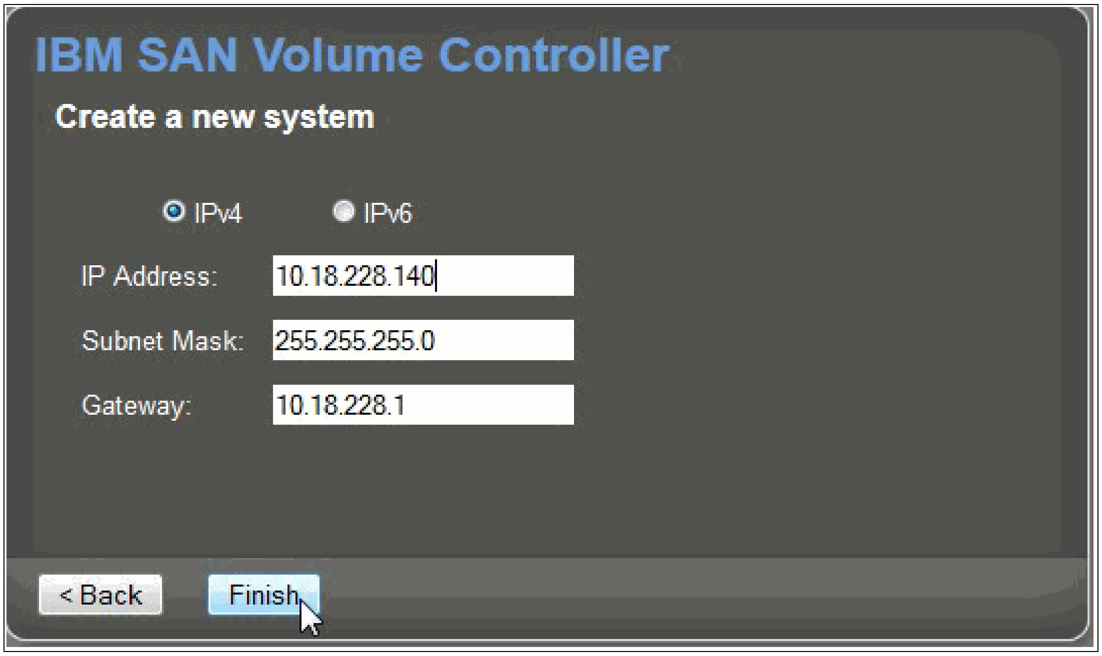


Figure 6-5 Setting the IP address

Click **Finish** when you have entered the IP address/Subnetmask and Gateway. The system will now start to initialize the system (node), and you will see a window like the one in Figure 6-6.

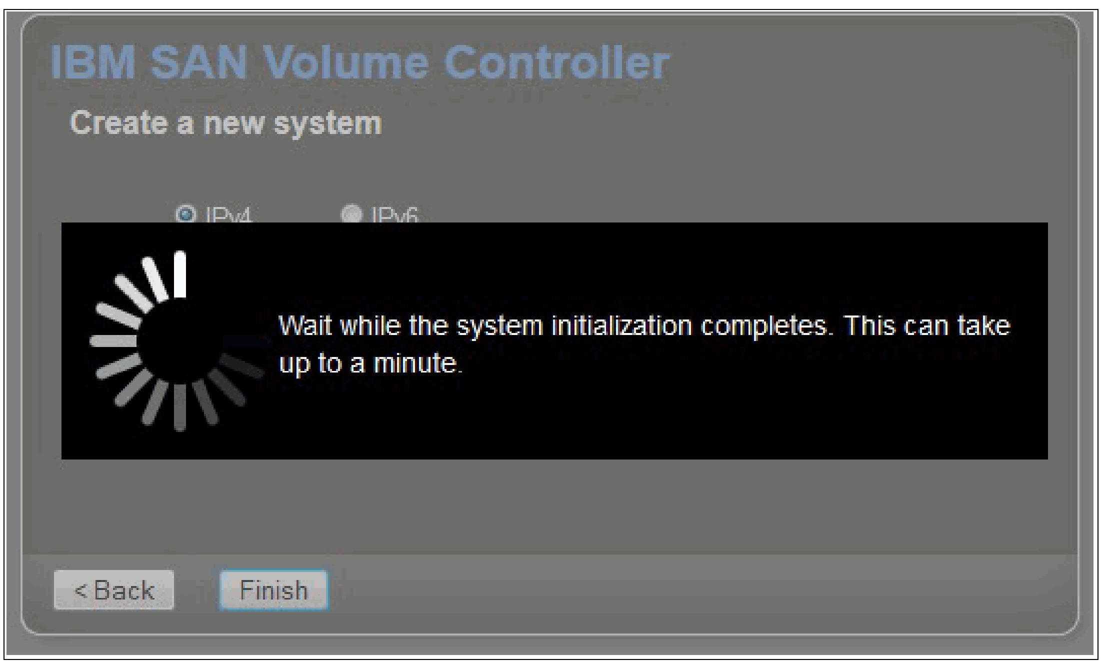


Figure 6-6 Initialization of the cluster

When the initialization is successfully completed, you will see the message shown in Figure 6-7.

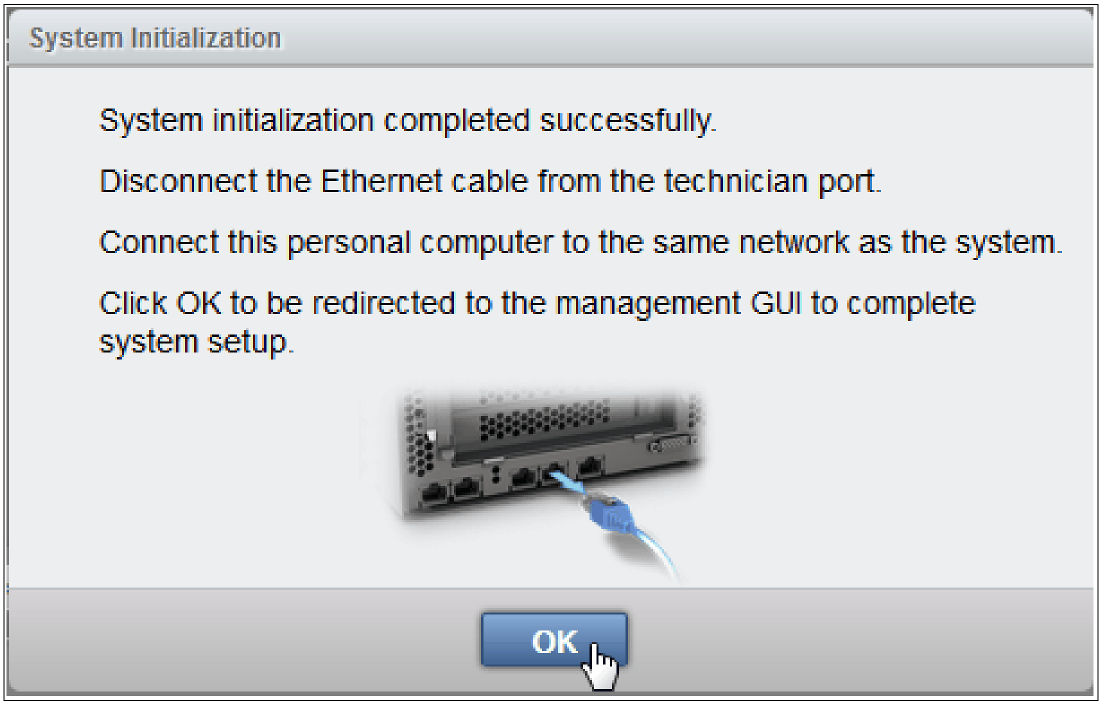


Figure 6-7 Initialization complete

Follow the on-screen instructions: Disconnect the Ethernet cable from the Technician port as well as from your PC or notebook, connect the same PC or notebook to the same network as the system, and upon clicking **OK** you are redirected to the GUI for completion of the system setup. You can connect to the System IP address from any management console that is connected to the same network as the system.

Whether redirected from your PC or notebook or connecting to the Management IP address of the system, you will be taken to the Login window as shown in Figure 6-8 on page 121.

Figure 6-8 Login window

You will then need to type in the default password (passw0rd with a zero). Click “Log in” and you will be asked to change the default password, as shown in Figure 6-9, which can be any combination 6 - 63 characters.

Figure 6-9 Change default password window

When you have changed the password, you will only need to go through a few more steps before the initial configuration is completed.

You will get to the system setup page as shown in Figure 6-10.

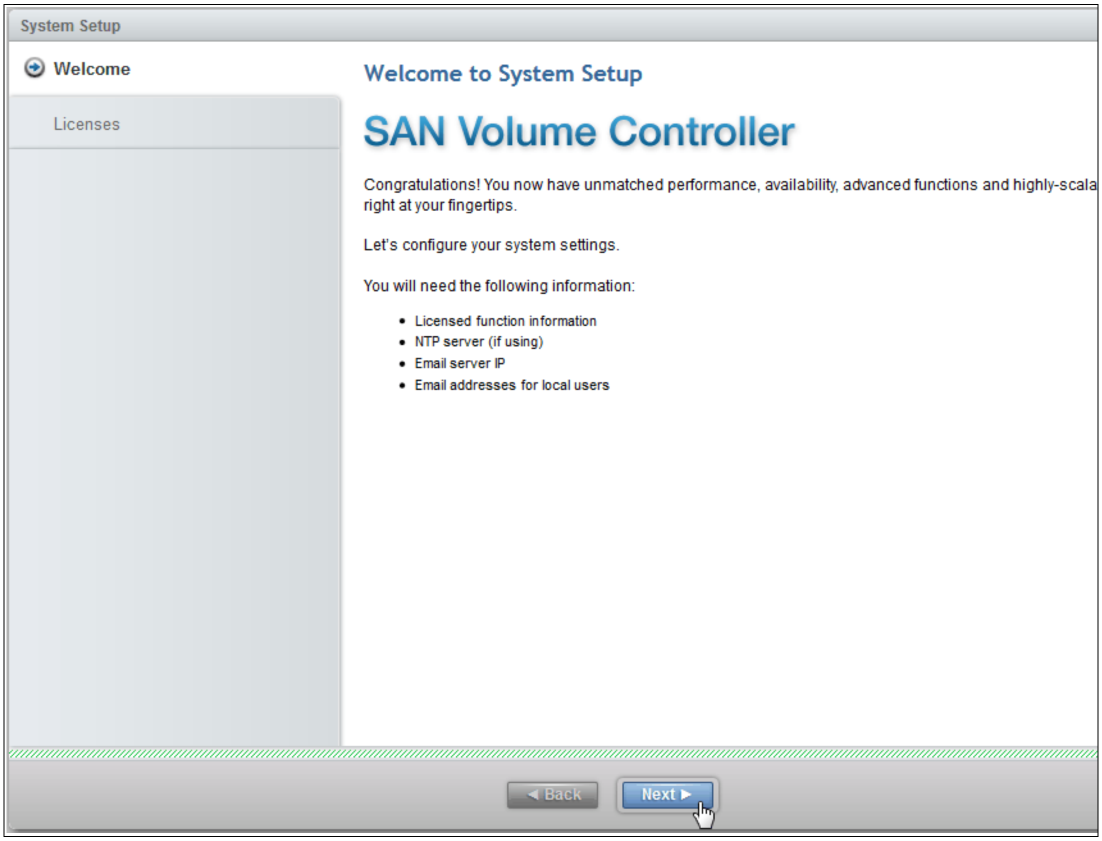


Figure 6-10 Welcome to system setup

Next, you have to read and accept the License agreement in Figure 6-11, and click **Next**.

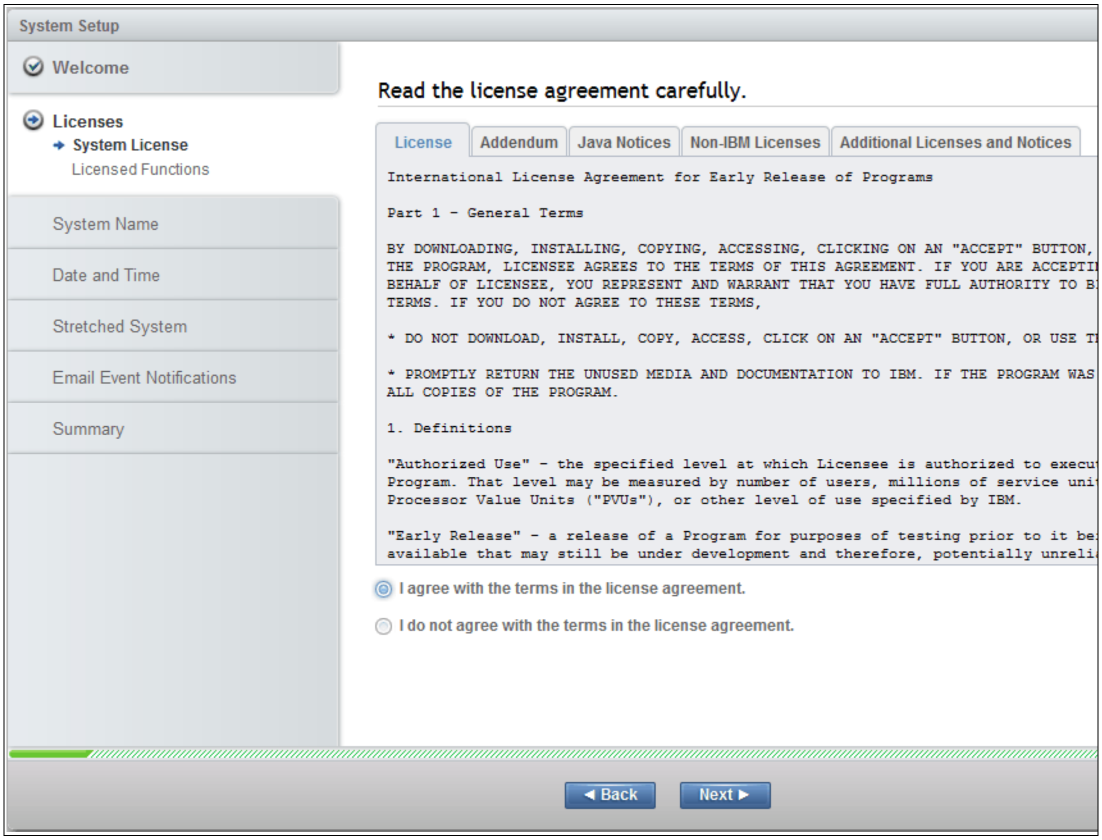


Figure 6-11 License agreement

You can now enter the purchased license for this cluster as shown in Figure 6-12.

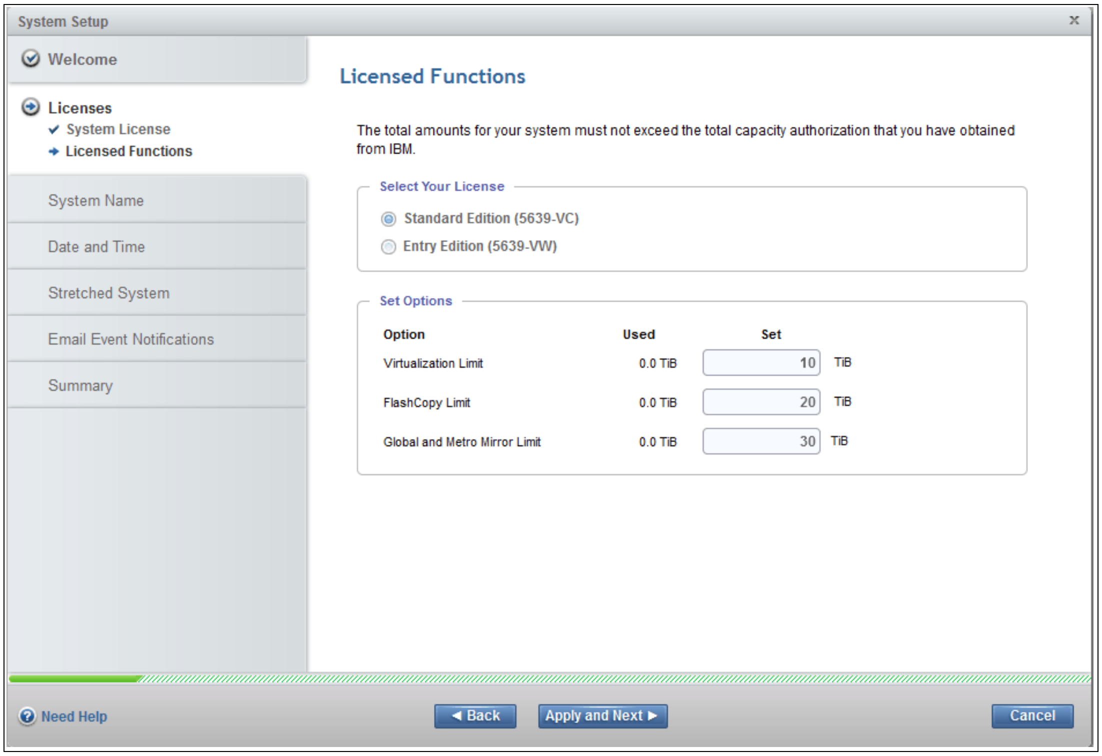


Figure 6-12 Licensed functions

> **Note:** You cannot enter the compression license at this stage. If you have ordered the 2145-DH8 with compression cards, they will need to be hardware activated before you can enter the license for compression. This is described as the last part of the initial configuration as depicted in Figure 6-30 on page 135 and its related text.

When you press **Apply and Next**, the system will configure the license as seen in Figure 6-13 on page 125.

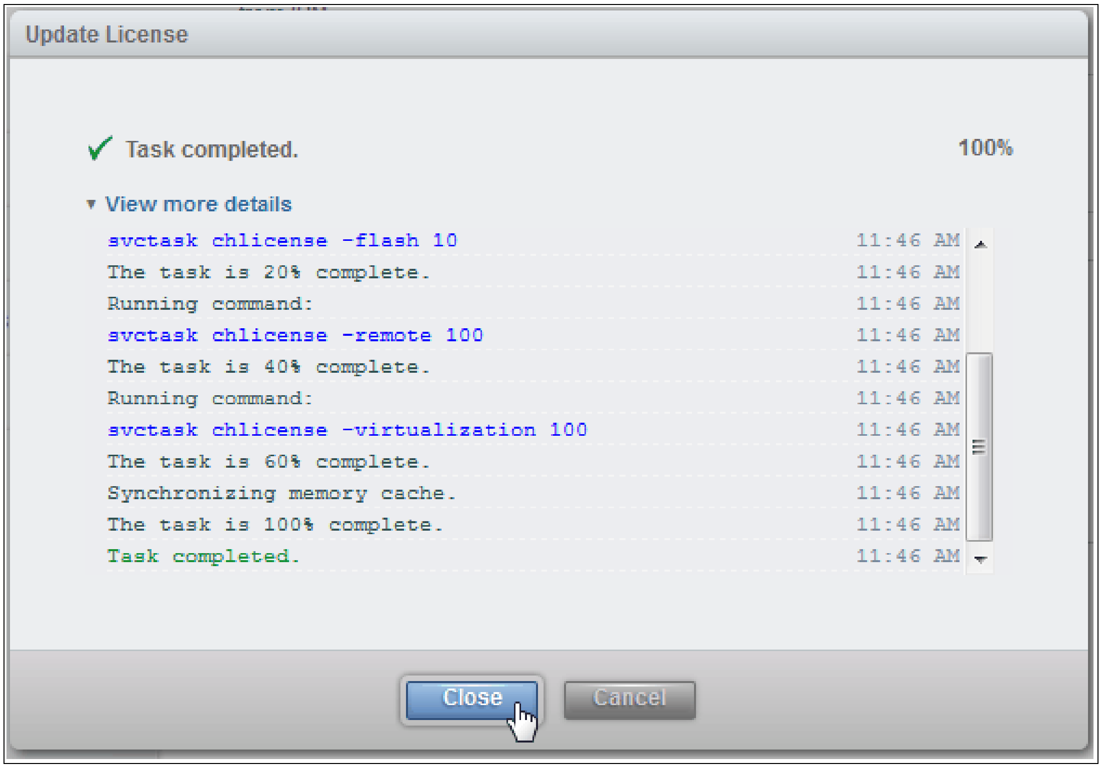


Figure 6-13 Applying license

On the next window, you can choose the wanted name for the system. You can also choose to leave it with the default name and if wanted you can change it later. See Figure 6-14.

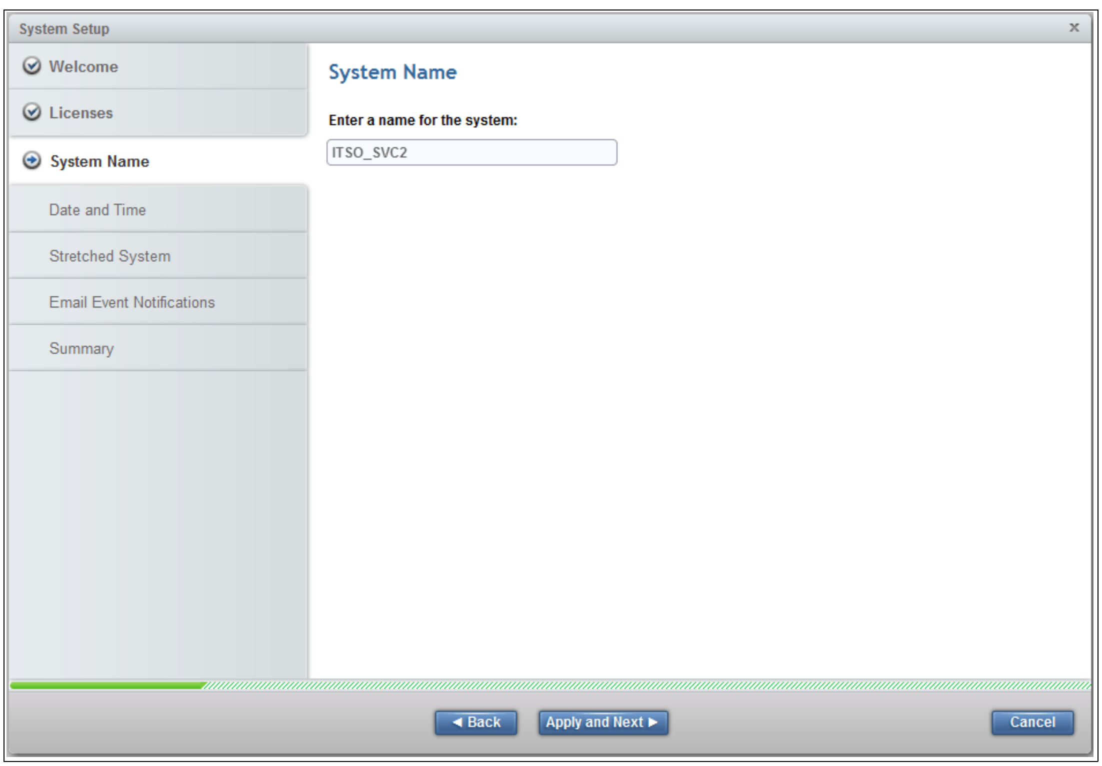


Figure 6-14 System name

When pressing **Apply and Next**, the system will configure the system name, as seen in Figure 6-15.

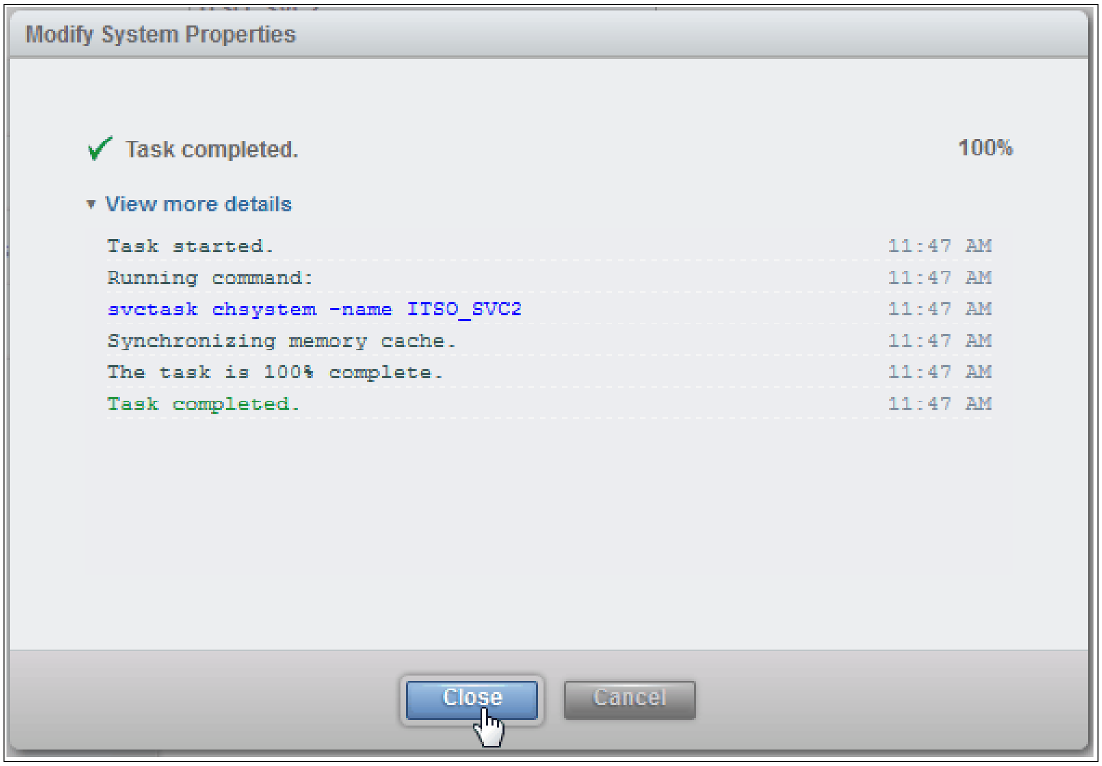


Figure 6-15 Applying system name

Figure 6-16 on page 127 allows you to set the date and time. You can either use manual settings or use an NTP server. Using an NTP server in the environment for all devices is highly recommended, for example, to ensure that you have a common time stamp for troubleshooting. Note that at this stage if you choose manual settings, you can only select 12-hour settings AM/PM. This can be changed later to be a 24-hour setting. See Figure 6-16 on page 127.

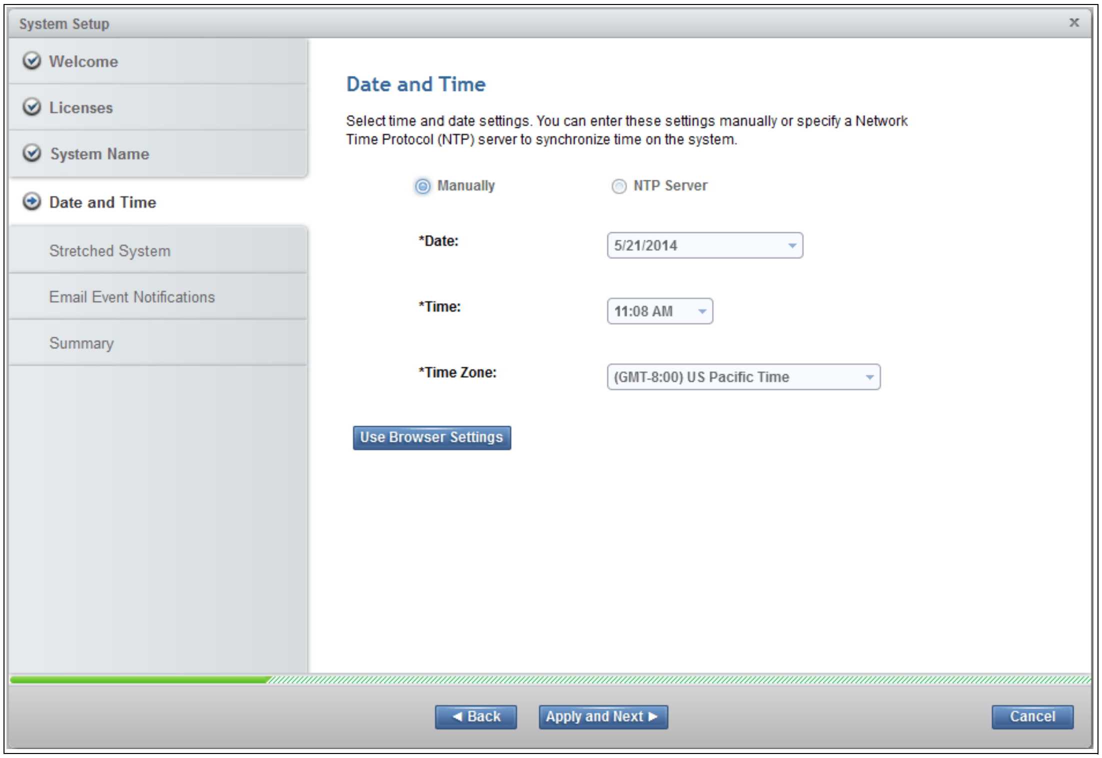


Figure 6-16 Setting Date and Time

After pressing **Apply and Next**, the system will configure the Date and Time as seen in Figure 6-17.

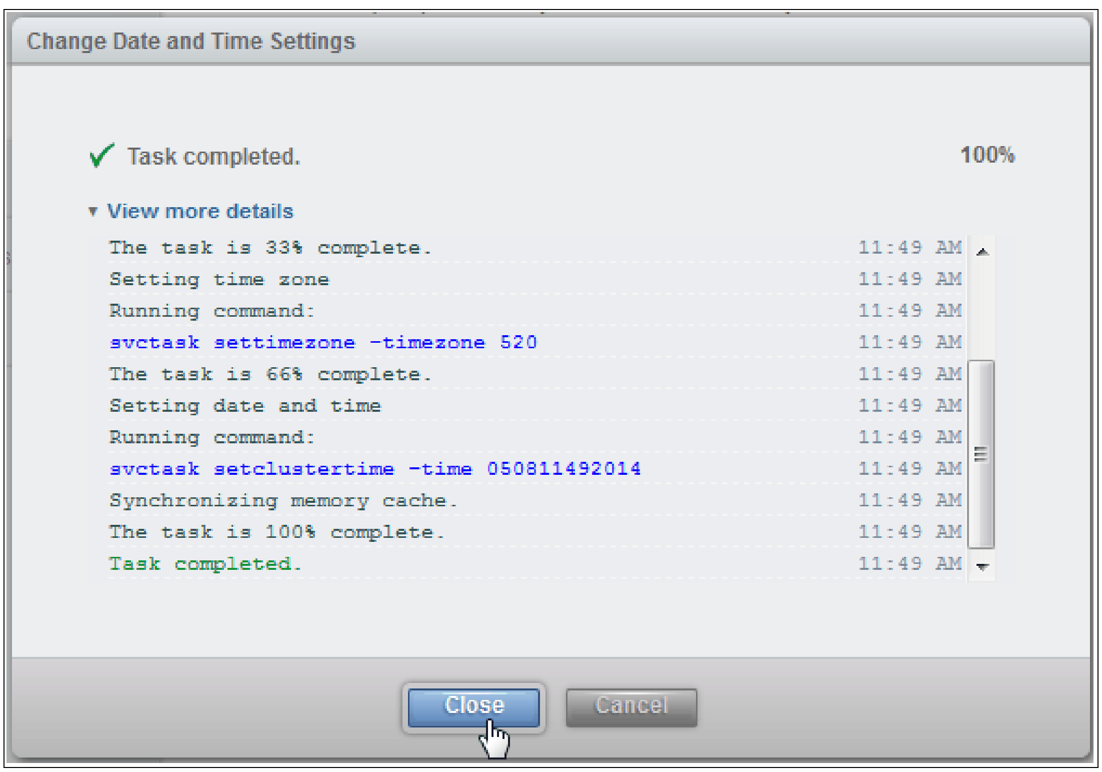


Figure 6-17 Applying Date and Time

You will then be asked if this is a *Stretched System* across multiple sites. See Figure 6-18. This is also known as *setting the topology*.

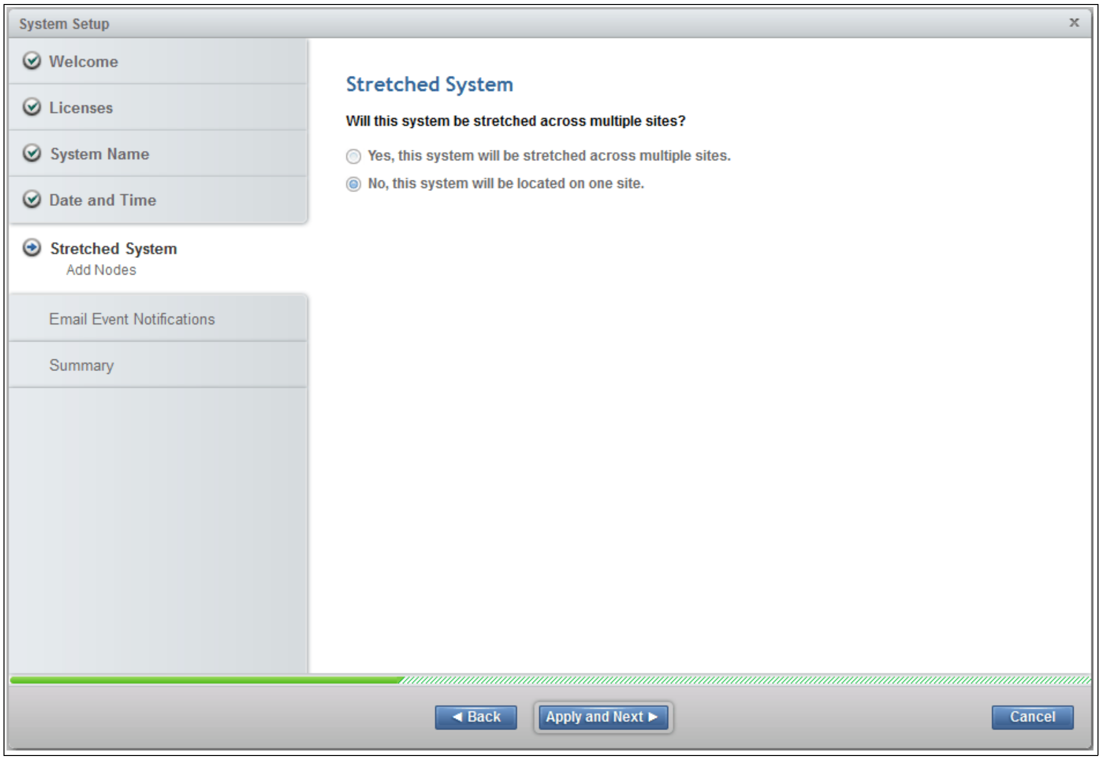


Figure 6-18 Stretched System or one site system

Note: If you are configuring a Stretched System, you are not able to use the disk expansion unit with solid-state drives for that purpose. See Chapter 11, "SAN Volume Controller Stretched System" on page 345.

After pressing **Apply and Next**, the system will configure the selected choice as shown in Figure 6-19 on page 129.

Note: The term *Topology* was introduced in version 7.2 as a part of an enhanced stretch system and site awareness. See also Chapter 11, "SAN Volume Controller Stretched System" on page 345.

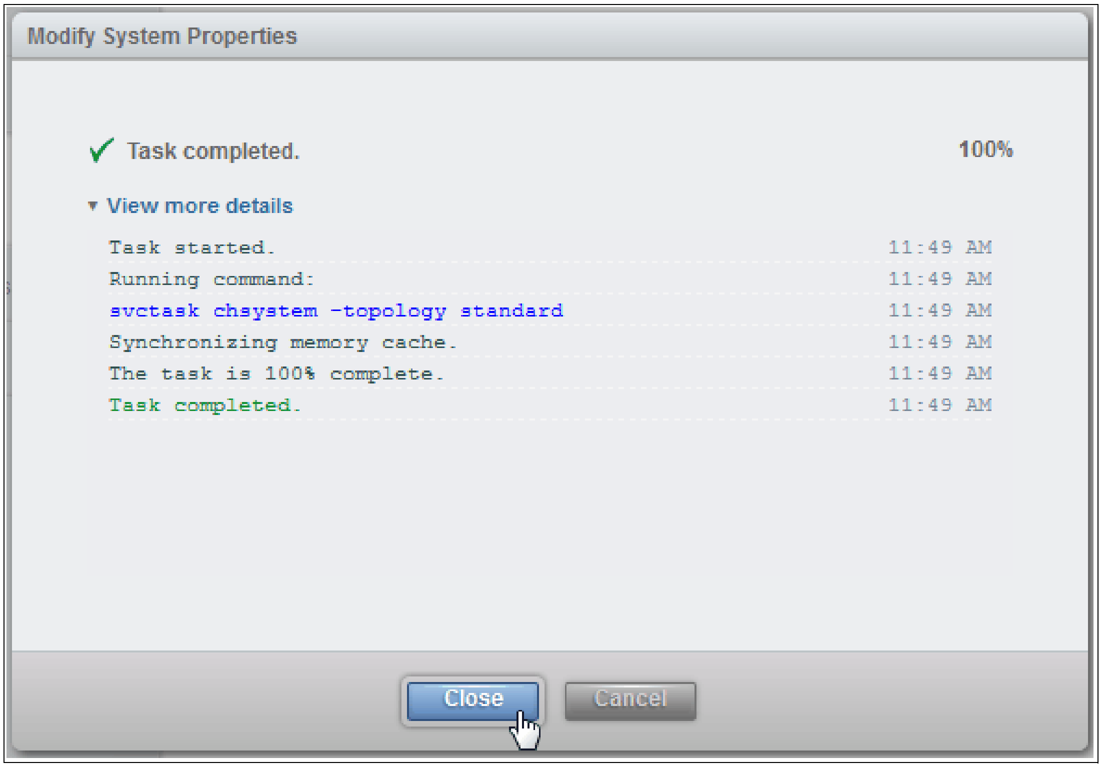


Figure 6-19 Applying the topology

Clicking **Close** will take you to the next menu, which is where you can Add Nodes to the system, as shown in Figure 6-20.

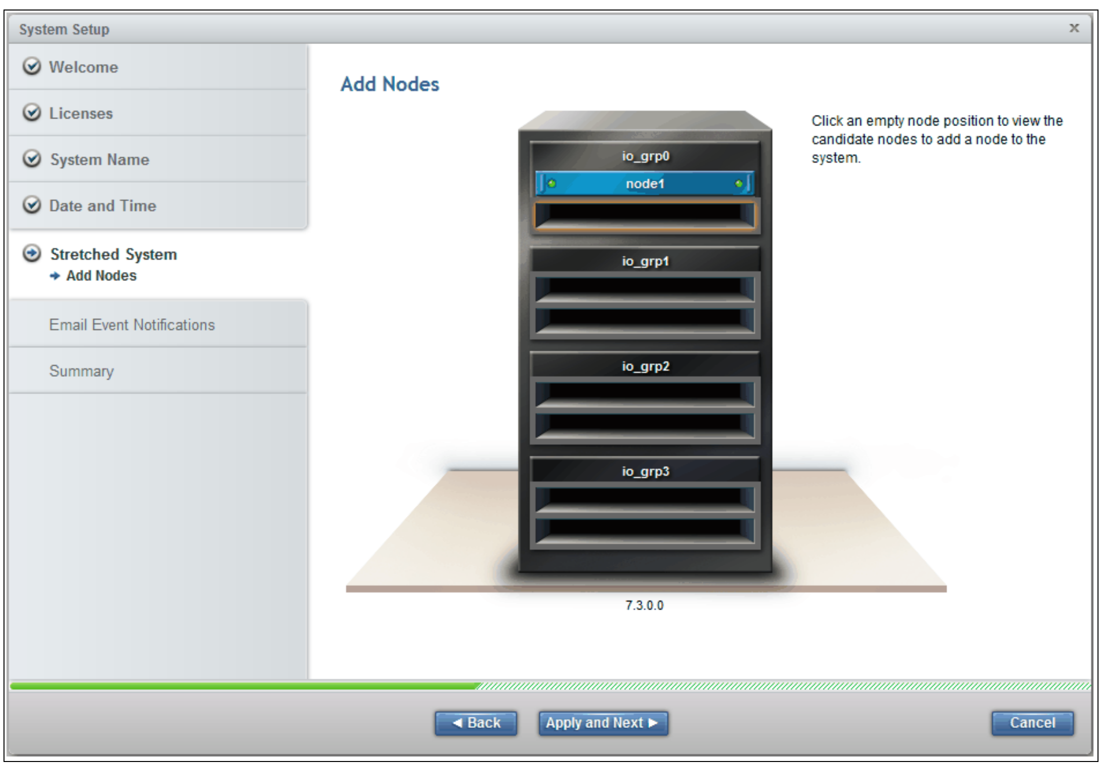


Figure 6-20 System with the first node

Now it is time to add the rest of the nodes that are to be used in the system. Click an empty node position to view the candidate nodes to add a node to the system, as seen in Figure 6-21. In this scenario, we have selected slot 2, which is the same IO group as the first node.

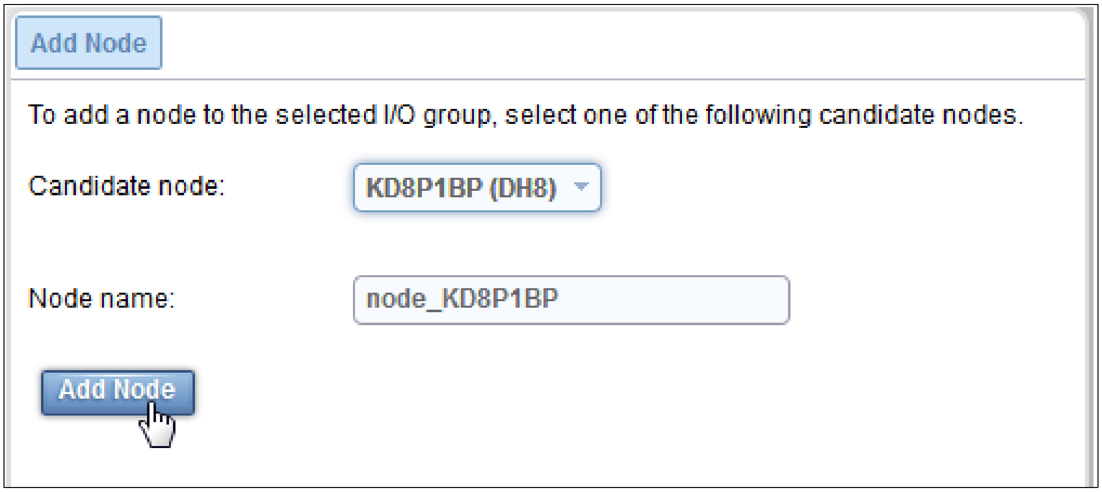


Figure 6-21   Adding a node to the system

You will now be presented with a warning. See Figure 6-22.

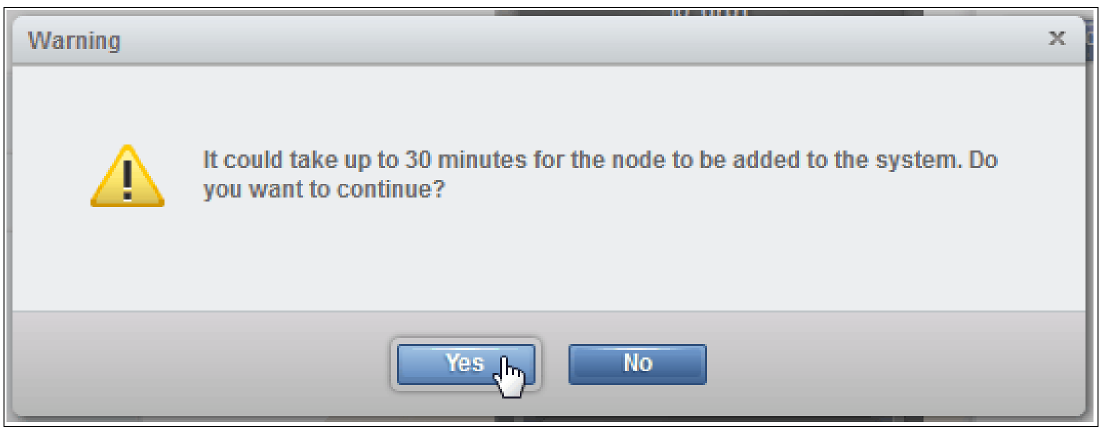


Figure 6-22   Adding a node warning

Click **Yes**, and the window as shown in Figure 6-23 on page 131 appears.

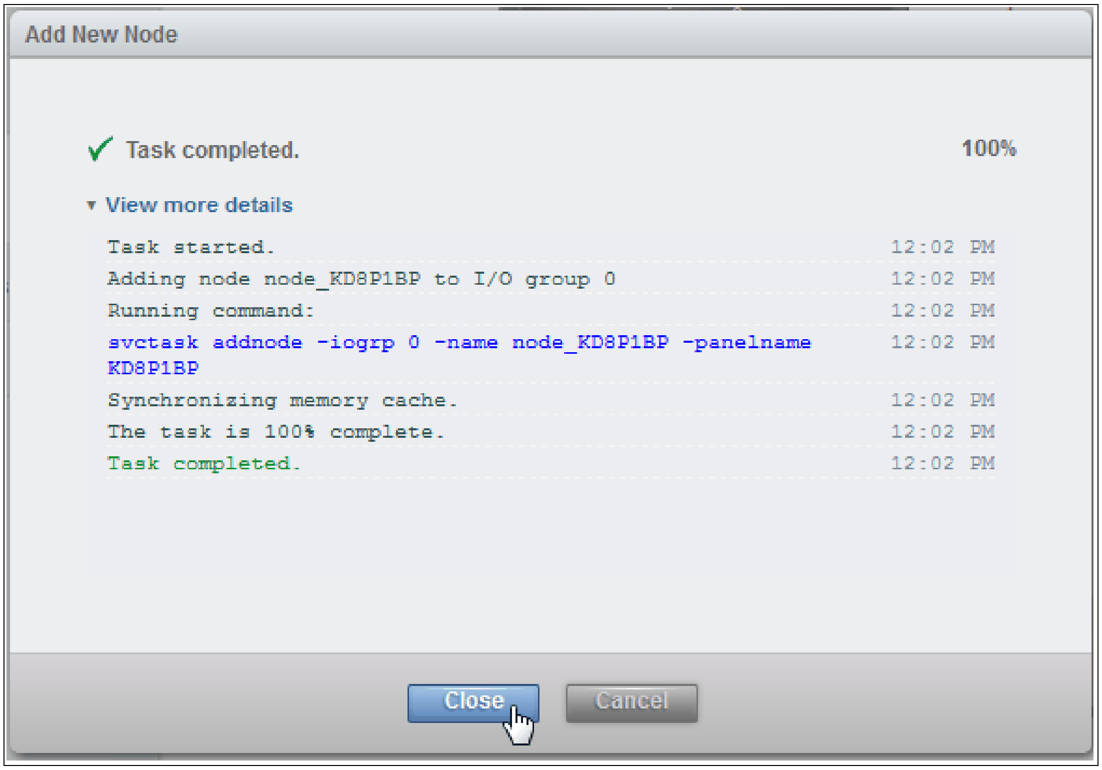


Figure 6-23 Adding a new node

Continue this process for all nodes that are to be added to the system.

Note: Ensure that you have collected serial numbers for all SVC nodes before you reach this step. This should be done as a part of the physical installation, according to the wanted rack layout. See 4.1, “General planning rules” on page 58.

Clicking **Close** takes you to the System Setup window where the nodes have been added, as shown in Figure 6-24 on page 132.

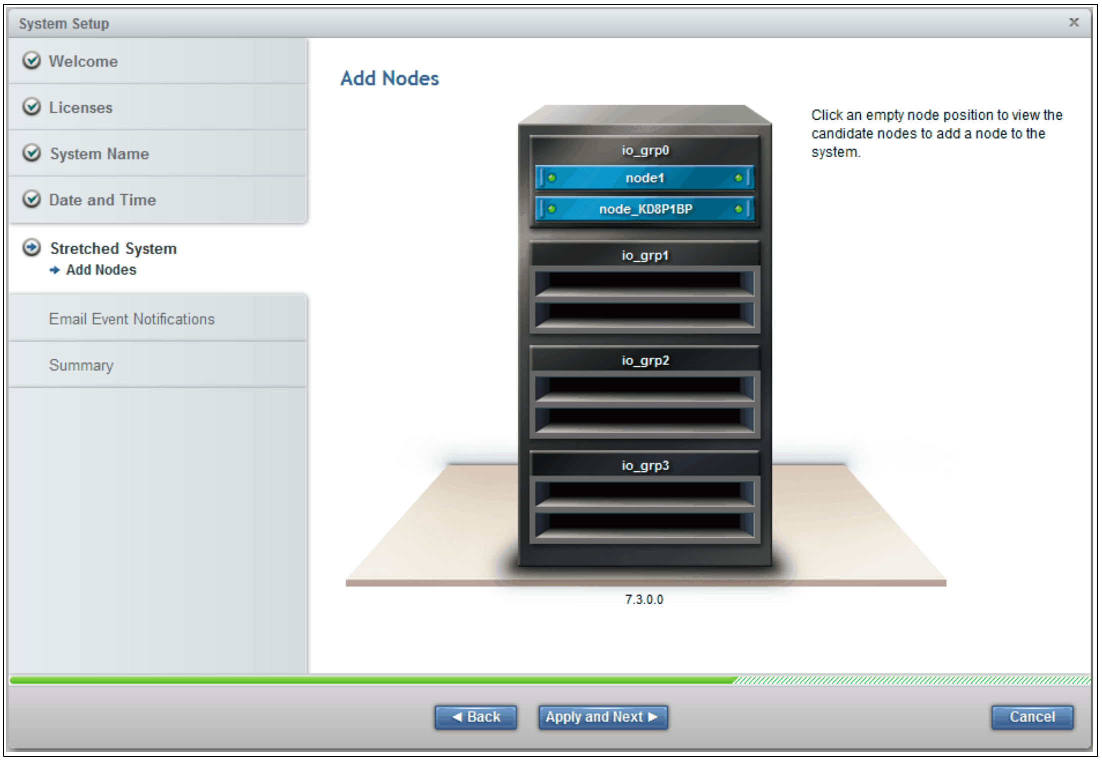


Figure 6-24 System Setup window

Press **Apply and Next**, and you will be asked if you want to set up Email Event Notifications, as shown in Figure 6-25.

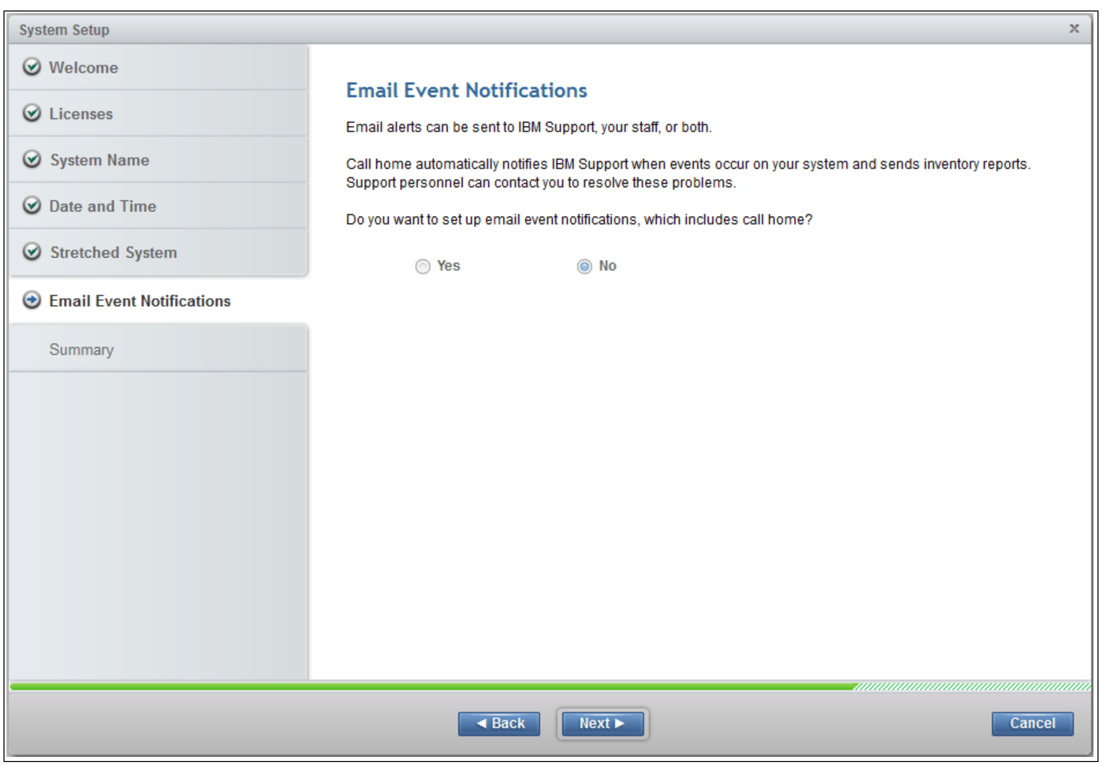


Figure 6-25 Email Event Notification

It is highly recommended to set up this function, however it can be configured later. If you choose to say No to this option now, a warning will presented to you, as shown in Figure 6-26.

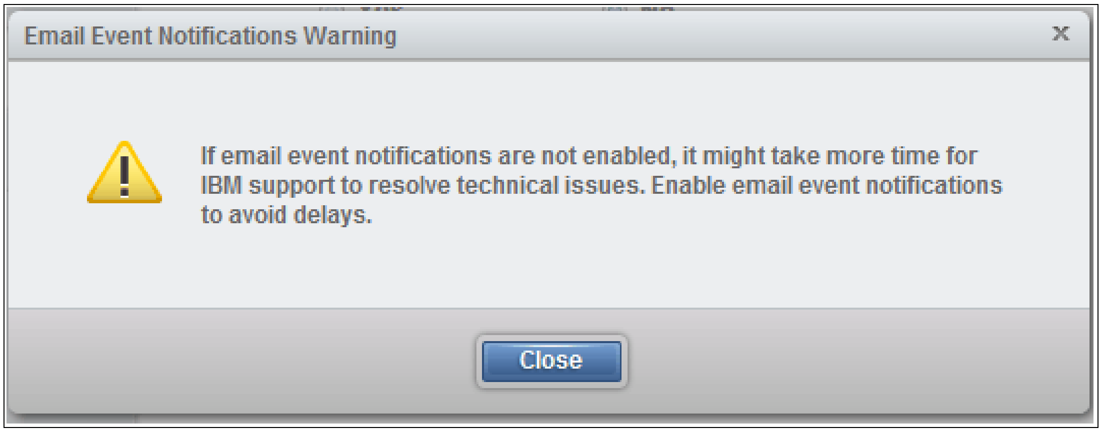


Figure 6-26 No email warning

Note: You must have access to an SMTP server (by IP address) to be able to configure Email Event Notifications.

Now you are only a few clicks away from completing the initial setup. When you click **Close** on the email warning, you will get to the Summary window, as shown in Figure 6-27.

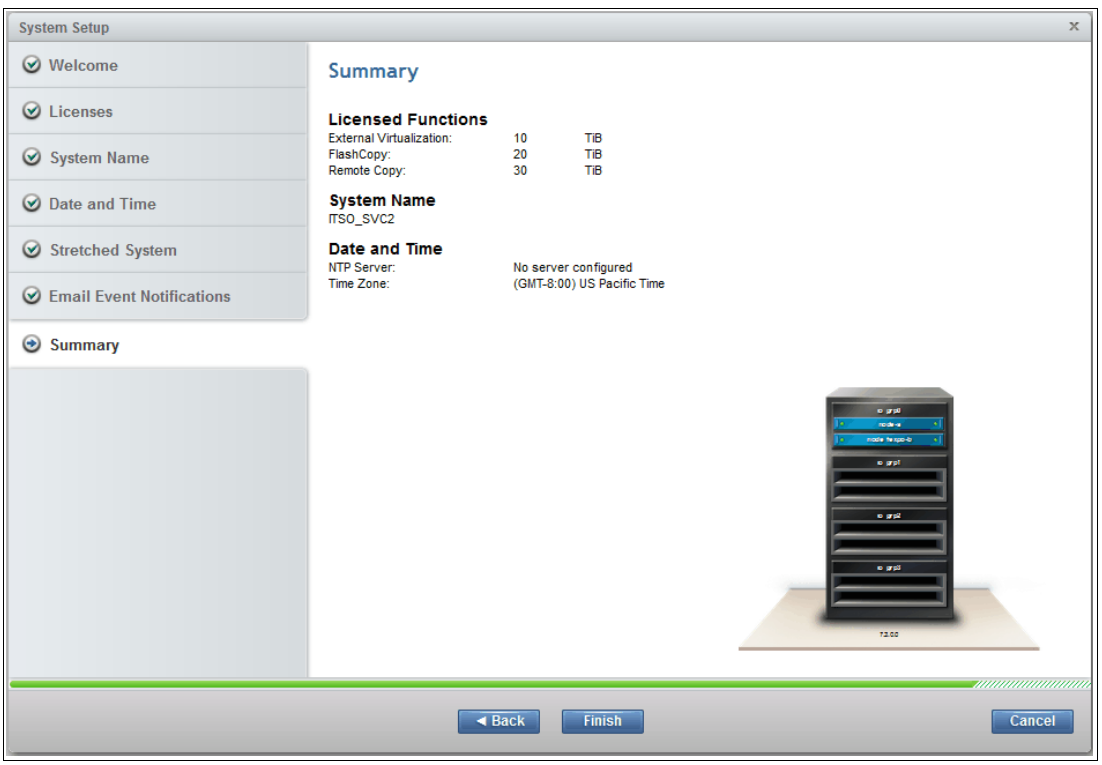


Figure 6-27 Summary

Click **Finish**, and the system will complete the initial setup as shown in Figure 6-28.

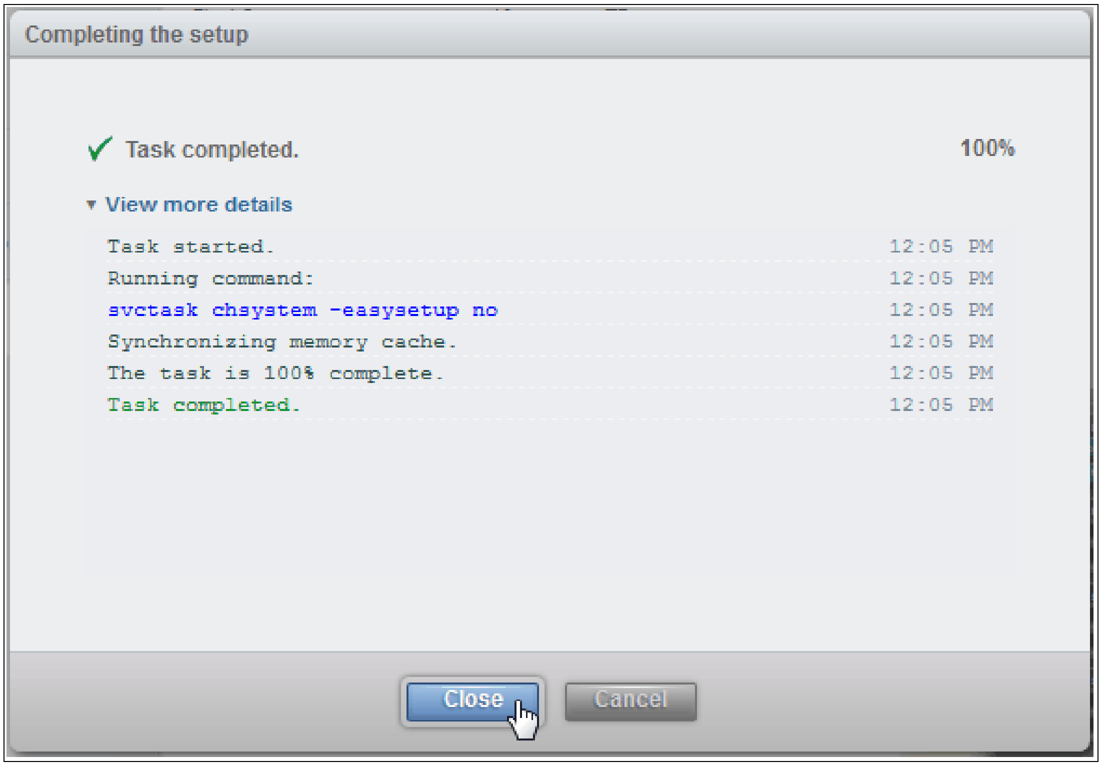


Figure 6-28 Initial configuration complete

When you click **Close**, you will then be directed to the System overview, as shown in Figure 6-29.

Figure 6-29 System overview

Now you have completed the initial configuration, and storage systems, hosts, and so on, can be configured. If you have purchased the 2145-DH8 with compression cards or any other adapter cards, they will need to be hardware activated before you can use the compression feature. We will show this process in the following sections.

If compression cards, or any other adapter cards, are installed in the nodes, after the completion of the initial setup you will briefly see a Status Alert in the lower right corner, as seen in Figure 6-30, which states that the detected hardware needs activation.

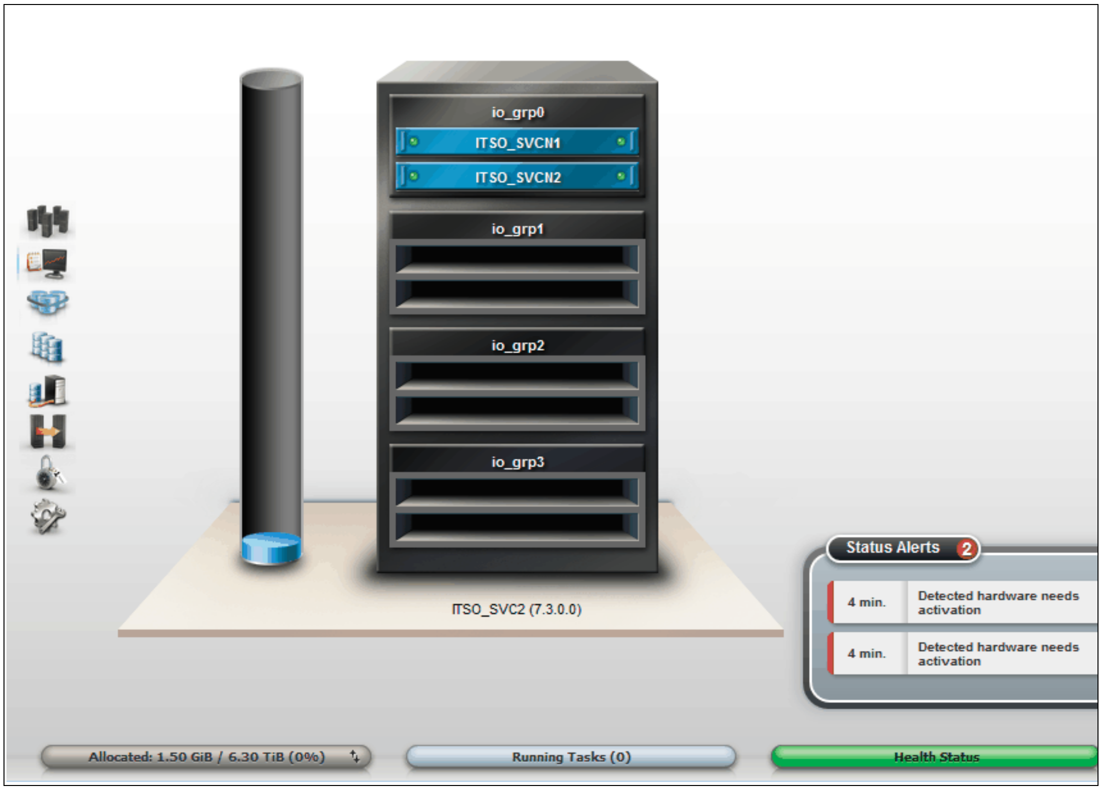


Figure 6-30 Detected hardware needs activation

When you click either one of the Status Alerts, you will be redirected to the event menu. See Figure 6-31 on page 136. Errorcode 1199 means that there is hardware that requires hardware activation.

For a complete list of errorcodes and warnings, go to the IBM SVC Knowledge Center:

http://pic.dhe.ibm.com/infocenter/svc/ic/index.jsp

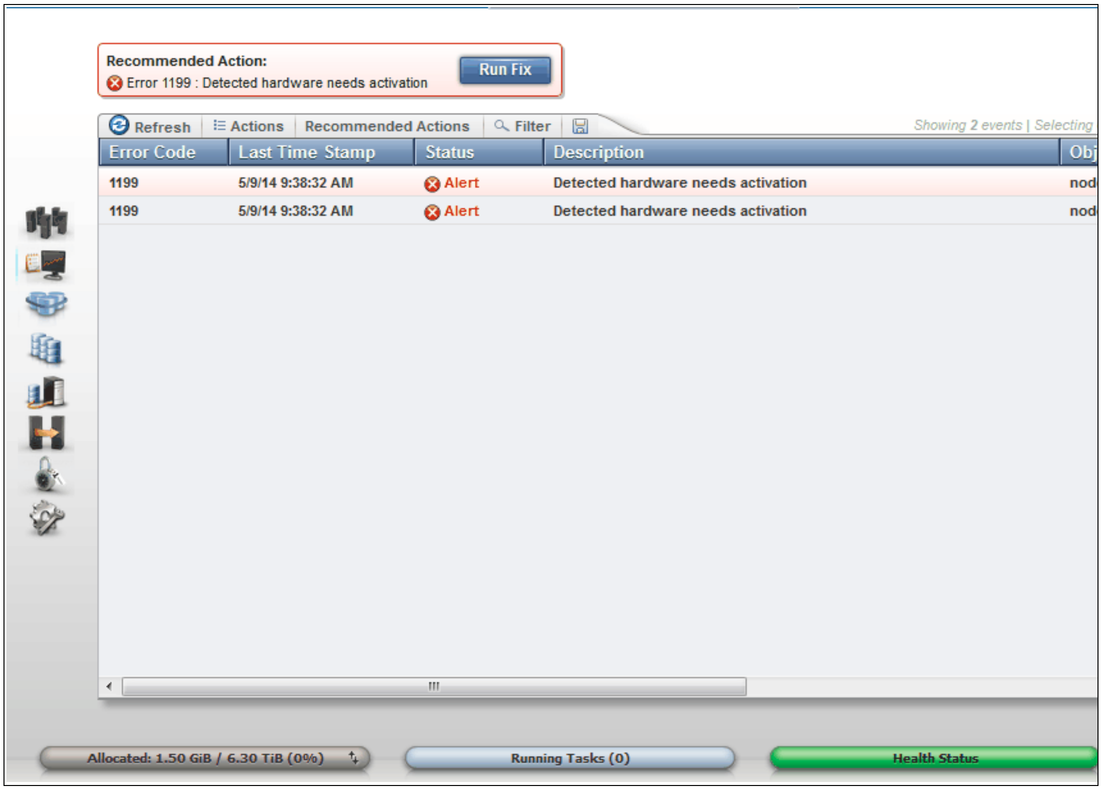
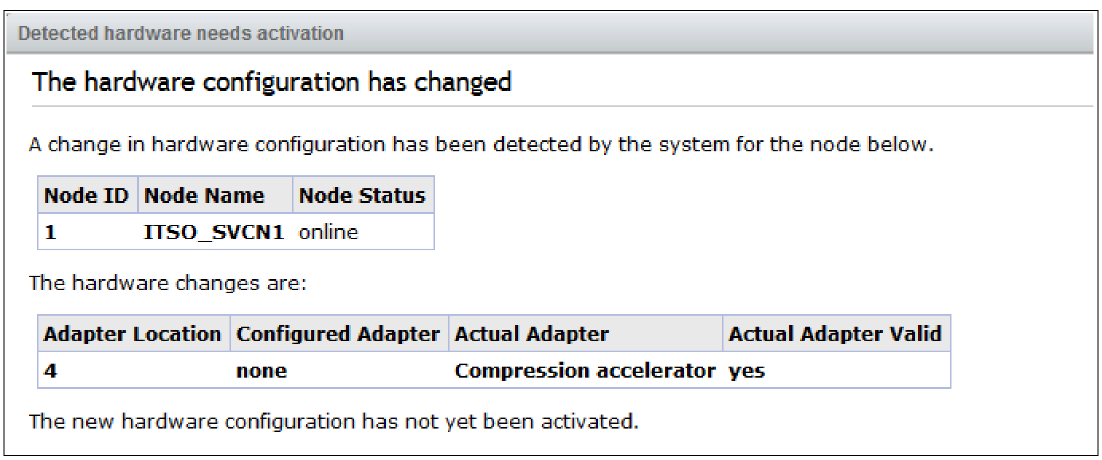


Figure 6-31 Event notifications

Select **Run Fix**, and you will see the window that is shown in Figure 6-32.


Detected hardware needs activation

The hardware configuration has changed

A change in hardware configuration has been detected by the system for the node below.

Node ID	Node Name	Node Status
1	ITSO_SVCN1	online

The hardware changes are:

Adapter Location	Configured Adapter	Actual Adapter	Actual Adapter Valid
4	none	Compression accelerator	yes

The new hardware configuration has not yet been activated.

Figure 6-32 Hardware configuration change

After reading the warning, when you click the radio button as shown in Figure 6-33, and have selected **Accept new hardware**, the node will reboot before the changes take effect.

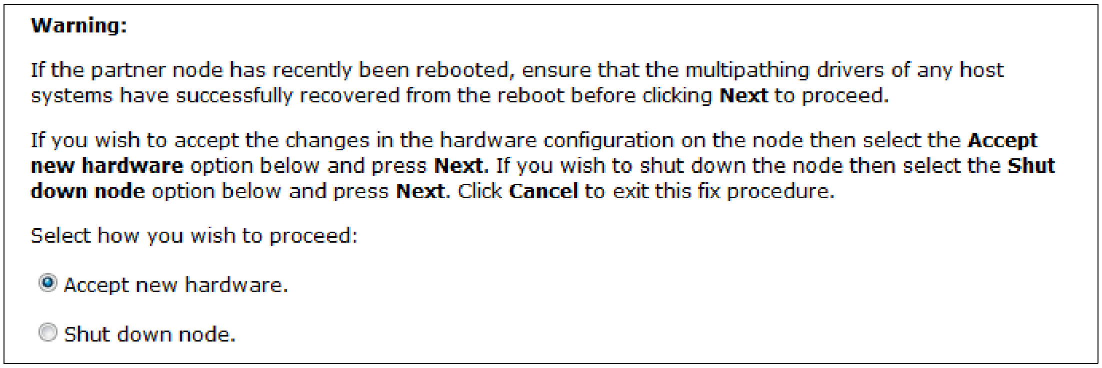


Figure 6-33 Accept new hardware

Figure 6-34 shows that the new hardware has been accepted, and is activated after the node has rebooted.

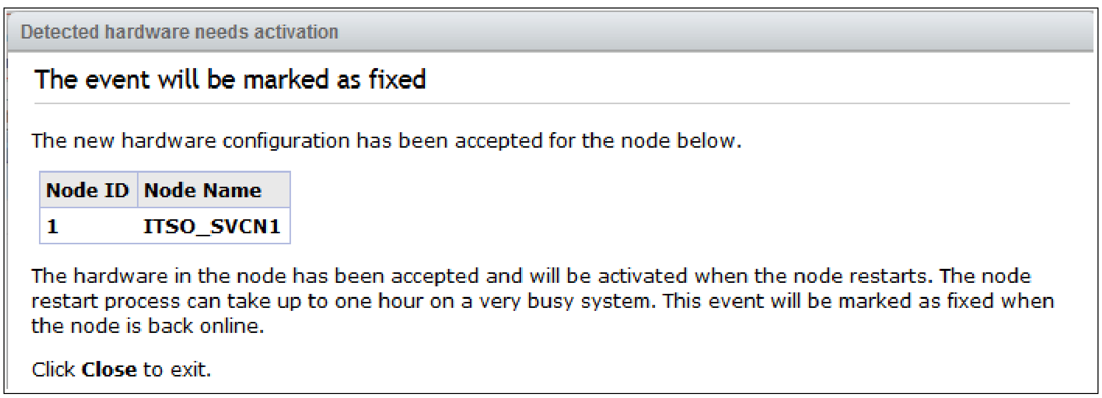


Figure 6-34 New hardware accepted

You will need to perform the same procedure for all the nodes that have compression cards, or any other adapter cards installed, and when the node has rebooted, the Status Alert warnings will automatically disappear.

You can verify that the new hardware is active by entering the **Monitoring Menu** → **System Details** → **Hardware** under each node. You can see the details for installed adapters as shown in Figure 6-35 on page 138.

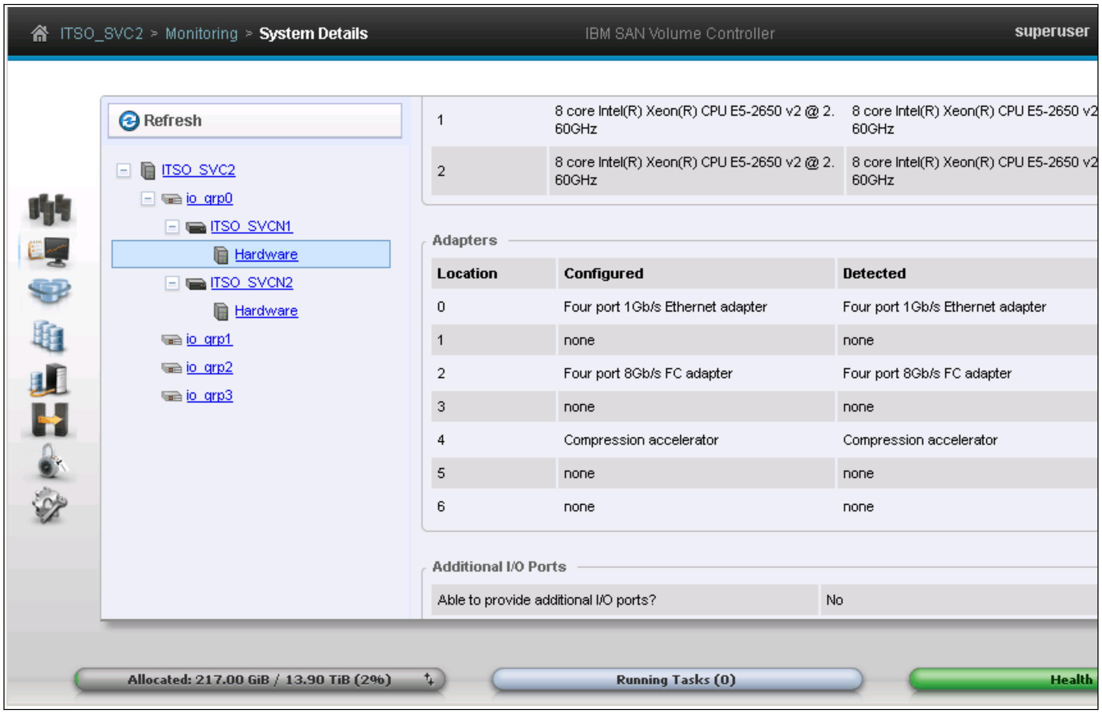


Figure 6-35 Configured adapters

> **Note:** If you are having trouble trying to activate the new hardware/adapters using the GUI, you can try to access the system by using the CLI.
>
> When you have access to the system by CLI, issue the following command to do a rescan of the installed hardware:
>
> **`chnodehw -force “nodename/nodeid`**
>
> If hardware activation problems persist, contact the IBM Support team in your region to solve the problem.

You can now start to add storage systems to the SAN Volume Controller system, and configure the added storage for the SAN Volume Controller. Details for how to manage storage under the SAN Volume Controller can be found in the following publication: *IBM System Storage SAN Volume Controller and Storwize V7000 Best Practices and Performance Guidelines*, SG24-7521.

7

Real-time Compression and the SVC 2145-DH8 Node

This chapter provides details about the Real-time Compression feature as it pertains to the SAN Volume Controller 2145-DH8. We provide an overview of Real-time Compression and its applicable use cases, offer a deeper look into Real-time Compression and the Random Access Compression Engine (RACE), and examine the specific hardware options and improvements in the 2145-DH8 node that specifically benefit the use of Real-time Compression.

7.1 Real-time Compression background, overview, and value proposition

Businesses and organizations around the world are challenged with tough economic conditions. The IT environment, which was historically viewed as an expense, is now viewed as a source of innovation that must drive future revenue. However, ever-increasing data storage requirements consume the available resources and disrupt attempts to innovate the IT environment.

The modern IT department has numerous challenges, including:

- Support for increasing data storage requirements: Shrinking IT budgets are pressuring IT managers to increase the lifetime of existing storage systems. Traditional methods of cleanup of unneeded data and archival of files to auxiliary storage are time consuming. They shift one resource constraint, physical storage, to another: The human work of storage administrators.
- Power, cooling, and floor space: A data center provides the means to host the storage systems. However, the physical characteristics of the hard disk drive-based systems limit the amount of data that can be stored per rack unit. Power consumption and heat dissipation are major concerns for IT managers who must fit the storage systems into a limited data center. This conflicts with the increasing demand for computing power needed to support new types of applications.
- High availability of data: Digital information has become the basis for any service in use today. As a result, the underlying systems that provide access to digital information are expected to be online all the time. This requirement has made it impossible to introduce data reduction solutions that impose any type of downtime. This restriction is true whether it is an actual inability to access the data, or merely a major slowdown when accessing an optimized data set.

Compression of primary storage provides an innovative approach designed to overcome these challenges.

7.1.1 The solution: IBM Real-time Compression

The IBM Real-time Compression solution addresses the challenges listed in the previous section because it was designed from the ground up for primary storage. Implementing Real-time Compression in SAN Volume Controller provides the following benefits:

- Compression for active primary data: IBM Real-time Compression can be used with active primary data. Therefore, it supports workloads that are not candidates for compression in other solutions. The solution supports online compression of existing data. It allows storage administrators to regain free disk space in an existing storage system without requiring administrators and users to clean up or archive data. This configuration significantly enhances the value of existing storage assets, and the benefits to the business are immediate. The capital expense of upgrading or expanding the storage system is delayed.
- Compression for replicated/mirrored data: Remote volume copies can be compressed in addition to the volumes at the primary storage tier. This process reduces storage requirements in Metro Mirror and Global Mirror destination volumes as well.
- No changes to the existing environment are required: IBM Real-time Compression is part of the storage system. It was designed with transparency in mind so that it can be implemented without changes to applications, hosts, networks, fabrics, or external storage

systems. The solution is not apparent to hosts, so users and applications continue to work as-is. Compression occurs within the SAN Volume Controller system itself.

- Overall savings in operational expenses: More data is stored in a rack space, so fewer storage expansion enclosures are required to store a data set. This reduced rack space has the following benefits:
  - Reduced power and cooling requirements: More data is stored in a system, therefore requiring less power and cooling per gigabyte of used capacity.
  - Reduced software licensing for additional functions in the system: More data stored per enclosure reduces the overall spending on licensing.
- Disk space savings are immediate: The space reduction occurs when the host writes the data. This process is unlike other compression solutions in which some or all of the reduction is realized only after a post-process compression batch job is run.

7.1.2 Common use cases

This section addresses the most common use cases for implementing compression:

- General-purpose volumes
- Databases
- Virtualized infrastructures

General-purpose volumes

Most general-purpose volumes are used for highly compressible data types, such as home directories, CAD/CAM, oil and gas geoseismic data, and log data. Storing such types of data in compressed volumes provides immediate capacity reduction to the overall consumed space. More space can be provided to users without any change to the environment.

There can be many file types stored in general-purpose servers. However, for practical information, the estimated compression ratios are based on actual field experience. Expected compression ratios are 50% to 60%.

File systems that contain audio, video files, and compressed files are not good candidates for compression. The overall capacity savings on these file types are minimal.

Databases

Database information is stored in table space files. It is common to observe high compression ratios in database volumes. Examples of databases that can greatly benefit from real-time compression are IBM DB2®, Oracle, and Microsoft SQL Server. Expected compression ratios are 50% to 80%.

Attention: Some databases offer optional built-in compression. Generally, do not compress already compressed database files.

Virtualized infrastructures

The proliferation of open systems virtualization in the market has increased the use of storage space, with more virtual server images and backups kept online. The use of compression reduces the storage requirements at the source.

Examples of virtualization solutions that can greatly benefit from real-time compression are VMWare, Microsoft Hyper-V, and KVM. Expected compression ratios are 45% to 75%.

Tip: Virtual machines with file systems that contain compressed files are not good candidates for compression as described in “General-purpose volumes”.

7.2 IBM Real-time Compression technology

IBM Real-time Compression technology is based on the RACE. RACE is an integral part of the software stack of SAN Volume Controller version 6.4.0 and later. This integration does not alter the behavior of the system, so that previously existing features are supported for compressed volumes. RACE leverages a lossless data compression algorithm along with a real-time compression technology, allowing it to meet the industry requirements of performance, reliability, scalability.

RACE technology leverages over 50 patents, many of which are not about compression. Rather, they define how to make industry standard Lempel-Ziv (L)-based compression of primary storage operate in real time while allowing random access. The primary intellectual property behind this is the RACE engine. At a high level, the IBM RACE component compresses data written into the storage system dynamically. This compression occurs transparently, so Fibre Channel and iSCSI connected hosts are not aware of the compression. RACE is an in-line compression technology, meaning that each host write is compressed as it passes through the SAN Volume Controller software to the disks. This has a clear benefit over other compression technologies that are post-processing in nature. These alternative technologies do not provide immediate capacity savings, and therefore are not a good fit for primary storage workloads such as databases and active data set applications.

RACE is based on the Lempel-Ziv lossless data compression algorithm and operates in a real-time method. When a host sends a write request, it is acknowledged by the upper level write cache of the system, and then destaged to the storage pool. As part of its destaging, it passes through the compression engine, and is then stored in compressed format onto the storage pool. Writes are therefore acknowledged immediately after being received by the upper write cache, with compression occurring as part of the destaging to internal or external physical storage.

Capacity is saved when the data is written by the host because the host writes are smaller when written to the storage pool.

IBM Real-time Compression is a self-tuning solution, similar to the SAN Volume Controller system itself. It adapts to the workload that runs on the system at any particular moment.

7.2.1 Random Access Compression Engine

To understand why RACE is unique, you need to review the traditional compression techniques. This description is not about the compression algorithm itself, that is, how the data structure is reduced in size mathematically. Rather, the description is about how the data is laid out within the resulting compressed output.

Compression utilities

Compression is probably most known to users because of the widespread use of compression utilities such as the zip and gzip utilities. At a high level, these utilities take a file as their input, and parse the data by using a sliding window technique. Repetitions of data are detected within the sliding window history, most often 32 KB. Repetitions outside of the window cannot be referenced. Therefore, the file cannot be reduced in size unless data is repeated when the window “slides” to the next 32 KB slot.

Figure 7-1 shows compression that uses a sliding window, where the first two repetitions of the string “ABCDEF” fall within the same compression window, and can therefore be compressed using the same dictionary. Note, the third repetition of the string falls outside of this window, and can therefore not be compressed using the same compression dictionary as the first two repetitions, reducing the overall achieved compression ratio.

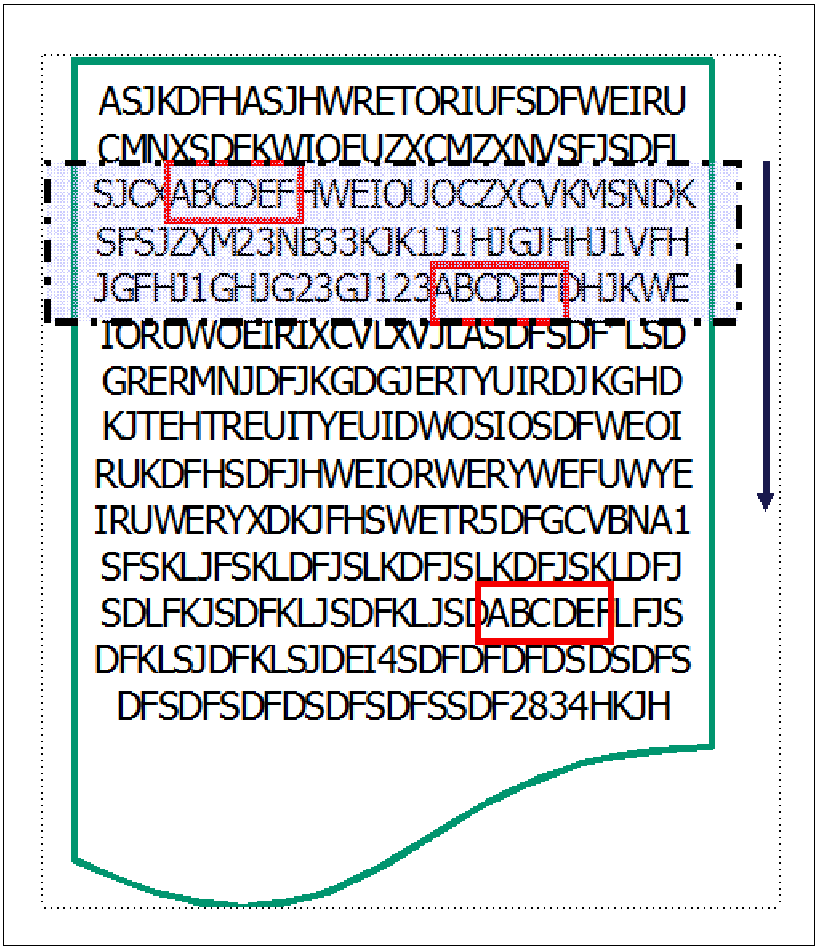


Figure 7-1 Compression that uses a sliding window

Traditional data compression in storage systems

The traditional approach taken to implement data compression in storage systems is an extension of how compression works in the compression utilities previously mentioned. Similar to compression utilities, the incoming data is broken into fixed chunks, and then each chunk is compressed and extracted independently.

However there are drawbacks to this approach. An update to a chunk requires a read of the chunk followed by a recompression of the chunk to include the update. The larger the chunk size chosen, the heavier the I/O penalty to recompress the chunk. If a small chunk size is chosen, the compression ratio is reduced because the repetition detection potential is reduced.

Figure 7-2 shows an example of how the data is broken into fixed size chunks (in the upper-left side of the figure). It also shows how each chunk gets compressed independently into variable length compressed chunks (in the upper-right side of the figure). The resulting compressed chunks are stored sequentially in the compressed output.

Although this approach is an evolution from compression utilities, it is limited to low performance use cases. This limitation is mainly because it does not provide real random access to the data.

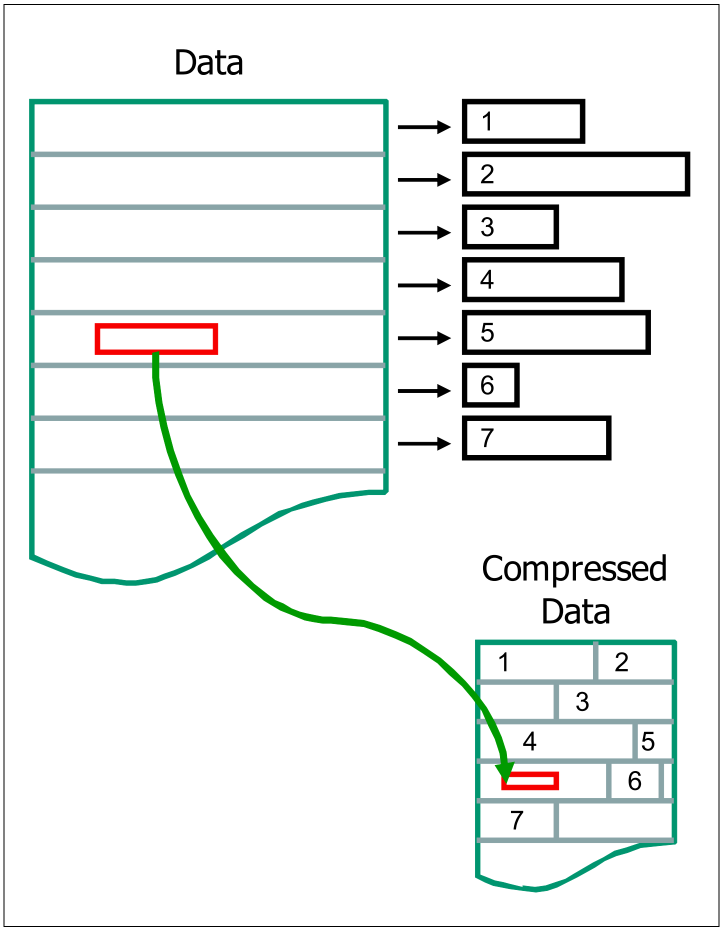


Figure 7-2 Traditional data compression in storage systems

Random Access Compression Engine

The IBM patented Random Access Compression Engine implements an inverted approach when compared to traditional approaches to compression. RACE uses variable-size chunks for the input, and produces fixed-size chunks for the output.

This method enables an efficient and consistent method to index the compressed data because it is stored in fixed-size containers.

Figure 7-3 on page 145 shows Random Access Compression.

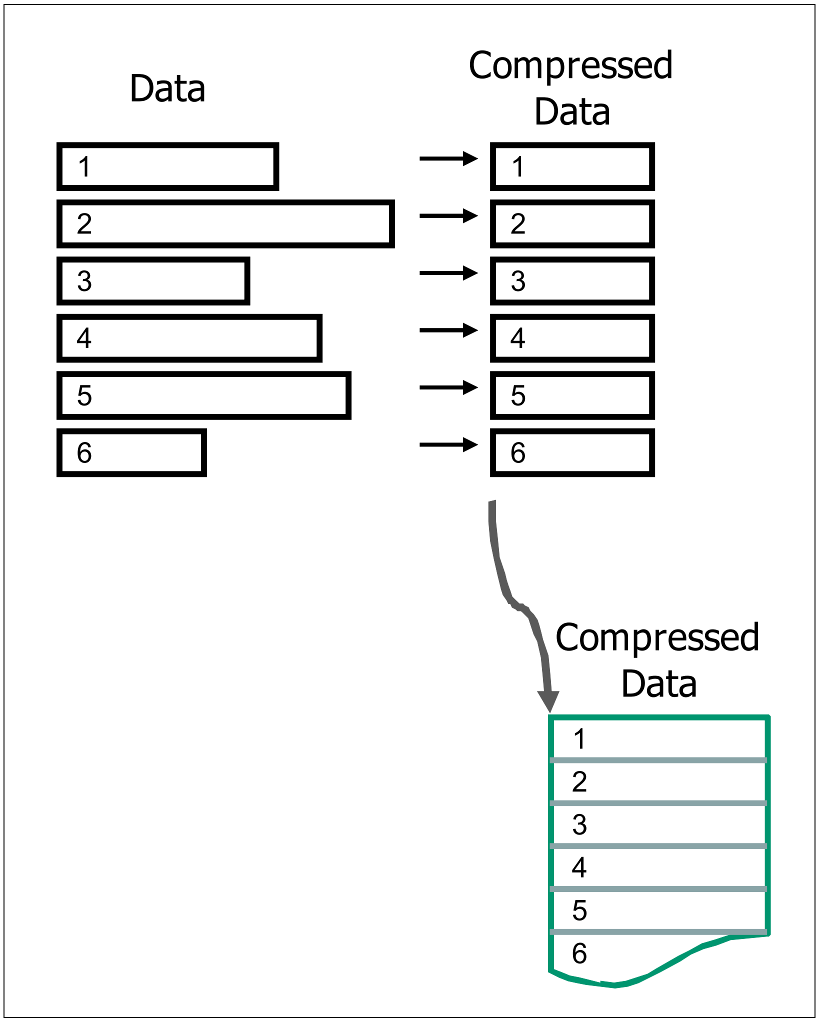


Figure 7-3 Random Access Compression

Location-based compression

Both compression utilities and traditional storage systems compression compress data by finding repetitions of bytes within the chunk that is being compressed. The compression ratio of this chunk depends on how many repetitions can be detected within the chunk. The number of repetitions is affected by how much the bytes stored in the chunk are related to each other. The relation between bytes is driven by the format of the object. For example, an office document might contain textual information, and an embedded drawing (like this page). Because the chunking of the file is arbitrary, it has no notion of how the data is laid out within the document. Therefore, a compressed chunk can be a mixture of the textual information and part of the drawing. This process yields a lower compression ratio because the different data types mixed together cause a suboptimal dictionary of repetitions. That is, fewer repetitions can be detected because a repetition of bytes in a text object is unlikely to be found in a drawing.

This traditional approach to data compression is also called *location-based compression*. The data repetition detection is based on the location of data within the same chunk.

Temporal compression

RACE offers a technology leap beyond location-based compression, temporal compression.

When host writes arrive to RACE, they are compressed and fill up fixed size chunks also called *compressed blocks*. Multiple compressed writes can be aggregated into a single compressed block. A dictionary of the detected repetitions is stored within the compressed block. When applications write new data or update existing data, it is typically sent from the host to the storage system as a series of writes. Because these writes are likely to originate from the same application and be from the same data type, more repetitions are usually detected by the compression algorithm.

This type of data compression is called *temporal compression* because the data repetition detection is based on the time the data was written into the same compressed block. Temporal compression adds the time dimension that is not available to other compression algorithms. It offers a higher compression ratio because the compressed data in a block represents a more homogeneous set of input data.

Figure 7-4 on page 147 shows (in the upper part) how three writes sent one after the other by a host end up in different chunks. They get compressed in different chunks because their location in the volume is not adjacent. This yields a lower compression ratio because the same data must be compressed non-natively by using three separate dictionaries. When the same three writes are sent through RACE (in the lower part of the figure), the writes are compressed together by using a single dictionary. This yields a higher compression ratio than location-based compression.

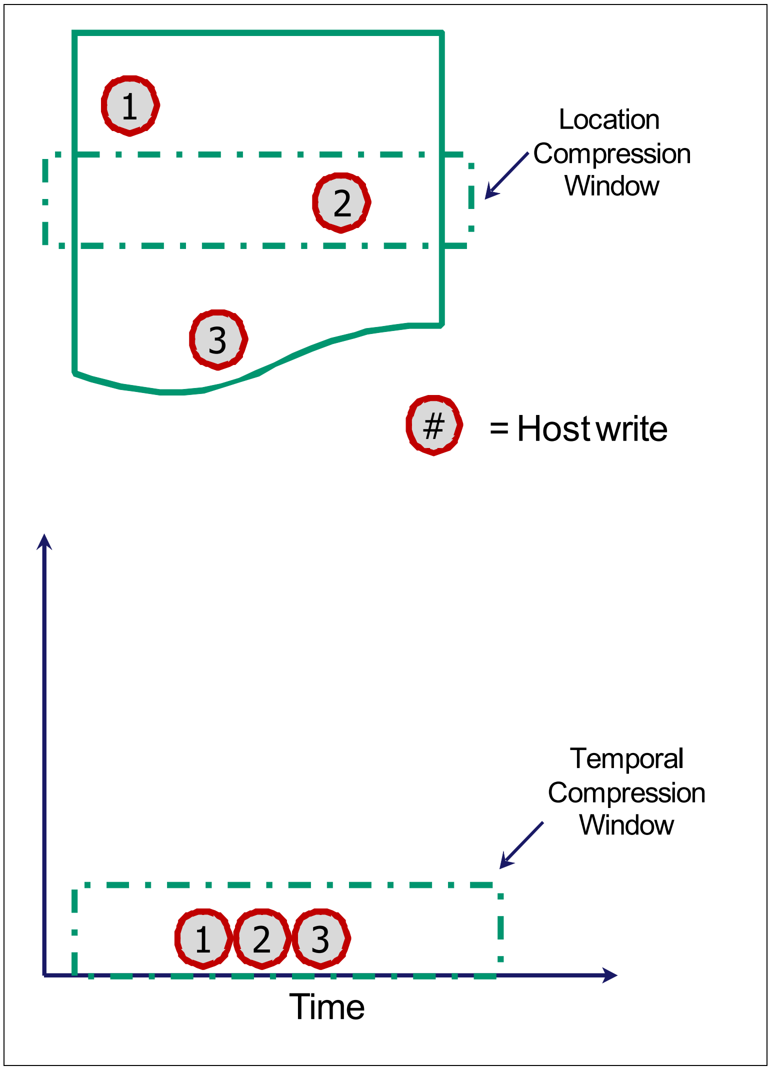


Figure 7-4 Location-based versus temporal compression

7.2.2 RACE in SAN Volume Controller software stack

It is important to understand where the RACE technology is implemented in the SAN Volume Controller software stack. This location determines how it applies to SAN Volume Controller components.

RACE technology is implemented into the SAN Volume Controller thin provisioning layer, and is an organic part of the stack. The SAN Volume Controller software stack is shown in Figure 7-5 on page 148. Compression is transparently integrated with existing system management design. All of the SAN Volume Controller advanced features are supported on compressed volumes. You can create, delete, migrate, map (assign), and unmap (unassign) a compressed volume as though it were a fully allocated volume. In addition, you can utilize Real-time Compression along with Easy Tier on the same volumes. This compression method provides nondisruptive conversion between compressed and decompressed volumes. This conversion provides a uniform user-experience and eliminates the need for special procedures when dealing with compressed volumes.

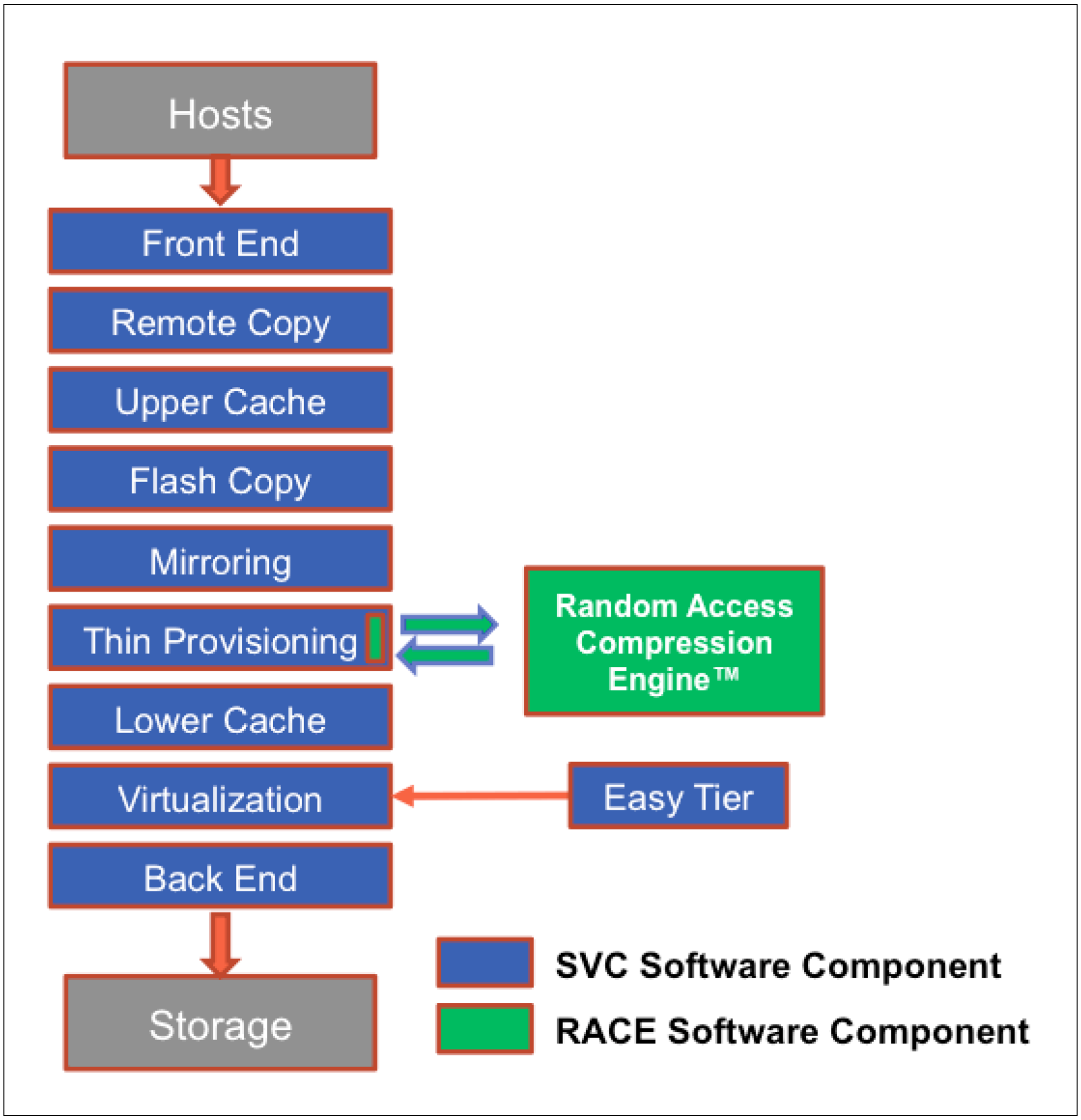


Figure 7-5 SAN Volume Controller 7.3 software stack

7.2.3 Data write flow

When a host sends a write request to SAN Volume Controller, it reaches the upper cache layer. The host is immediately sent an acknowledgment of its I/Os.

When the upper cache layer destages to the RACE, the I/Os are sent to the thin-provisioning layer. They are then sent to RACE, and if necessary, the original host write or writes. The metadata that holds the index of the compressed volume is updated if needed, and is compressed as well.

7.2.4 Data read flow

When a host sends a read request to the SAN Volume Controller for compressed data, it is forwarded directly to the Real-time Compression (RtC) component:

- If the RtC component contains the requested data, the SAN Volume Controller cache replies to the host with the requested data without having to read the data from the lower level cache or disk.
- If the RtC component does not contain the requested data, the request is forwarded to the SAN Volume Controller lower-level cache.
- If the lower-level cache contains the requested data, it is sent up the stack and returned to the host without accessing the storage.
- If the lower-level cache does not contain the requested data, it sends a read request to the storage for the requested data.

7.2.5 Compression of existing data

In addition to compressing data in real time, it is also possible to compress existing data sets. This compression adds a compressed mirrored copy to an existing volume. You then delete the original copy after the synchronization of the compressed copy is complete. This process is nondisruptive, so the data remains online and accessible by applications and users.

This capability enables customers to regain space from the storage pool, which can then be reused for other applications.

With virtualization of external storage systems, the ability to compress already stored data significantly enhances and accelerates the benefit to users. It allows them to see a tremendous return on their SAN Volume Controller investment. On initial purchase of a SAN Volume Controller with Real-time Compression, customers can defer their purchase of new storage. As new storage needs to be acquired, IT purchases a lower amount of the required storage before compression.

7.3 SVC 2145-DH8 node software and hardware updates that enhance Real-time Compression

The SVC 2145-DH8 node and SAN Volume Controller software version 7.3 introduce significant software and hardware improvements that enhance and extend the applicability of the Real-time Compression feature. In this section, we provide an overview of these enhancements:

- Software enhancements
  - SAN Volume Controller cache rearchitecture
- Hardware enhancements
  - Additional/enhanced CPU options
  - Increased memory options
  - Optional Intel Assist Acceleration Technology (Coletto Creek) compression acceleration cards

7.3.1 Software enhancements

As mentioned in Chapter 1, “Introduction to IBM storage virtualization” on page 1, SAN Volume Controller software version 7.3 introduces an enhanced, dual-level caching model. This model differs from the single-level cache model of previous software versions.

In the previous model, the Real-time Compression software component sits below the single-level read/write cache. The benefit of this model is that the upper-level read/write cache masks from the host any latency introduced by the Real-time Compression software component. However, in this single-level caching model, the destaging of writes for compressed I/Os to disk might not be optimal for certain workloads because the RtC component is interacting direct with uncached storage.

In the new, dual-level caching model, the Real-time Compression software component sits below the upper-level fast write cache and above the lower-level advanced read/write cache. There are several advantages to this dual-level model regarding Real-time Compression:

- Host writes, whether to compressed or decompressed volumes are still serviced directly via the upper-level write cache, preserving low host write I/O latency. Response time can improve with this model as the upper cache flushes less data to RACE more frequently.
- The performance of the destaging of compressed write I/Os to storage is improved because these I/Os are now destaged via the advanced lower-level cached, as opposed to directly to storage.
- The existence of a lower-level write cache below the Real-time Compression component in the software stack allows for the coalescing of compressed writes, and as a result, a reduction in back-end I/Os due to the ability to perform full-stride writes for compressed data.
- The existence of a lower-level read cache below the Real-time Compression component in the software stack allows the temporal locality nature of RtC to benefit from pre-fetching from the backend storage.
- The main (lower level) cache now stores compressed data for compressed volumes, increasing the effective size of the lower-level cache.
- Support for larger numbers of compressed volumes.

7.3.2 Hardware updates

As mentioned in Chapter 1, “Introduction to IBM storage virtualization” on page 1, the SAN Volume Controller 2145-DH8 node introduces numerous hardware enhancements. Several of these enhancements relate directly to the Real-time Compression feature and offer significant performance and scalability improvements over previous hardware versions.

Additional/enhanced CPU options

The 2145-DH8 node offers an updated primary CPU that contains 8 cores as compared to the 4 and 6 core CPUs available in previous hardware versions. Additionally, the 2145-DH8 node offers the option of a secondary 8 core CPU for use with Real-time Compression. This additional, compression-dedicated CPU allows for improved overall system performance when utilizing compression over previous hardware models.

> **Note:** In order to use the Real-time Compression feature on 2145-DH8 nodes, the secondary CPU is required.

Increased memory options

The 2145-DH8 node offers the option to increase the node memory from the base 32 GB to 64 GB, for use with Real-time Compression. This additional, compression-dedicated memory allows for improved overall system performance when utilizing compression over previous hardware models.

Note: In order to use the Real-time Compression feature on 2145-DH8 nodes, the additional 32 GB memory option is required.

Optional Intel Quick Assist Acceleration Technology (Coletto Creek) compression acceleration cards

The 2145-DH8 node offers the option to include one or two Intel Quick Assist compression acceleration cards based on the Coletto Creek chipset. The introduction of these Intel based compression acceleration cards in the SAN Volume Controller 2145-DH8 node is an industry first, providing dedicated processing power and greater throughout over previous models.

Note: In order to use the Real-time Compression feature on 2145-DH8 nodes, at least one Quick Assist compression acceleration card is required. With a single card, the maximum number of compressed volumes per I/O group is 200. With the addition of a second Quick Assist card, the maximum number of compressed volumes per I/O group is 512.

For additional details about the compression accelerator cards, refer to 3.1.4, “Compression Accelerator card” on page 44.

8

Performance data and statistics gathering

In this chapter, we provide a brief overview of the performance analysis capabilities of the IBM System Storage SAN Volume Controller (SVC) 7.3. We also describe a method that you can use to collect and process SVC performance statistics.

It is beyond the intended scope of this book to provide an in-depth understanding of performance statistics or explain how to interpret them. For a more comprehensive look at the performance of the SVC, see *SAN Volume Controller Best Practices and Performance Guidelines*, SG24-7521, which is available at the following IBM Redbooks publication site:

http://www.redbooks.ibm.com/abstracts/sg247521.html?Open

For SVC, as with all other IBM storage subsystems, the official IBM tool for the collection of performance statistics and to supply performance reporting, is IBM Tivoli Storage Productivity Center.

You can obtain more information about IBM Tivoli Storage Productivity Center usage and configuration in *SAN Storage Performance Management Using Tivoli Storage Productivity Center*, SG24-7364*:*

http://www.redbooks.ibm.com/abstracts/sg247364.html?Open

Chapter 10, “SAN Volume Controller operations using the GUI” on page 327 contains information about collecting performance statistics.

8.1 SAN Volume Controller performance overview

Although storage virtualization with SVC provides many administrative benefits, it can also provide a substantial increase in performance for various workloads. The caching capability of the SVC and its ability to stripe volumes across multiple disk arrays can provide a significant performance improvement over what can otherwise be achieved when using midrange disk subsystems.

To ensure that the wanted performance levels of your system are maintained, monitor performance periodically to provide visibility to potential problems that exist or are developing so that they can be addressed in a timely manner.

8.1.1 Performance considerations

When designing an SVC storage infrastructure or maintaining an existing infrastructure, you need to consider many factors in terms of their potential effect on performance. These factors include but are not limited to: Dissimilar workloads competing for the same resources, overloaded resources, insufficient resources available, poor performing resources, and similar performance constraints.

Remember the following high-level rules when designing your SAN and SVC layout:

- Host-to-SVC inter-switch link (ISL) oversubscription: This area is the most significant I/O load across ISLs. The recommendation is to maintain a maximum of 7-to-1 oversubscription. Going higher is possible, but it tends to lead to I/O bottlenecks. This suggestion also assumes a core-edge design, where the hosts are on the edge and the SVC is on the core.
- Storage-to-SVC ISL oversubscription: This area is the second most significant I/O load across ISLs. The maximum oversubscription is 7-to-1. Going higher is not supported. Again, this suggestion assumes a multiple-switch SAN fabric design.
- Node-to-node ISL oversubscription: This area is the least significant load of the three possible oversubscription bottlenecks. In standard setups, this load can be ignored; while it is not entirely negligible, it does not contribute significantly to ISL load. However, it is mentioned here regarding the split-cluster capability that was made available with 6.3.0. When running in this manner, the number of ISL links becomes much more important. As with the Storage-to-SVC ISL oversubscription, this load also has a requirement for a maximum of 7-to-1 oversubscription. Exercise caution and careful planning when you determine the number of ISLs to implement. If you need additional assistance, we recommend that you contact your IBM representative and request technical assistance.
- ISL trunking/port channeling: For the best performance and availability, we highly recommend that you use ISL trunking/port channeling. Independent ISL links can easily become overloaded and turn into performance bottlenecks. Bonded or trunked ISLs automatically share load and provide better redundancy in the case of a failure.
- Number of paths per host multipath device: The maximum supported number of paths per multipath device that is visible on the host is eight. Although the Subsystem Device Driver Path Control Module (SDDPCM), related products, and most vendor multipathing software can support more paths, the SVC expects a maximum of eight paths. In general, you see only an effect on performance from more paths than eight. Although the SVC can work with more than eight paths, this design is technically unsupported.
- Do not intermix dissimilar array types or sizes: Although the SVC supports an intermix of differing storage within storage pools, it is best to always use the same array model, same

Redundant Array of Independent Disks (RAID) mode, same RAID size (RAID 5 6+P+S does not mix well with RAID 6 14+2), and same drive speeds.

- When using Real-time Compression, ensure that you use both the Comprestimator tool to assess the expected compression ratio for the workload that you will be delivering, as well as Disk Magic or similar performance sizing tool that allows you to size your performance requirements including the requirements for compression. With the addition of the compression acceleration cards in the 2145-DH8 SVC node, there are additional performance benefits for compressed workloads.

Note: Comprestimator is a command-line host-based utility that can be used to estimate an expected compression rate for block devices. It can be downloaded from the following site:

http://www-01.ibm.com/support/docview.wss?uid=ssg1S4001012

Rules and guidelines are no substitution for monitoring performance. Monitoring performance can both provide a validation that design expectations are met and identify opportunities for improvement.

8.1.2 SAN Volume Controller performance perspectives

The SVC is a combination product that consists of software and hardware. The software was developed by the IBM Research Group and was designed to run on commodity hardware (mass-produced Intel based CPUs with mass-produced expansion cards), while providing distributed cache and a scalable cluster architecture. One of the main goals of this design was to be able to use refreshes in hardware. Currently, the SVC cluster is scalable up to eight nodes and these nodes can be swapped for newer hardware while online. This capability provides a great investment value because the nodes are relatively inexpensive and a node swap can be done online. This capability provides an instant performance boost with no license changes.

Newer nodes, such as the 2145-DH8 model, with additional CPU, Memory, and IO Cards, provide an extra benefit on top of the typical refresh cycle. These benefits include enhancements such as the number of compressed volumes that can be managed by a single IO Group, the number of Fibre Channel Ports per IO Group, and the amount of Read Cache available to an IO Group. The following link provides the node replacement/swap and node addition instructions:

http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TD104437

The performance is near linear when adding additional nodes into the cluster until performance eventually becomes limited by the attached components. Also, while virtualization with the SVC provides significant flexibility in terms of the components that are used, it does not diminish the necessity of designing the system around the components so that it can deliver the wanted level of performance. The key item for planning is your SAN layout. Switch vendors have slightly different planning requirements, but the end goal is that you always want to maximize the bandwidth that is available to the SVC ports. The SVC is one of the few devices capable of driving ports to their limits on average, so it is imperative that you put significant thought into planning the SAN layout.

Essentially, SVC performance improvements are gained by spreading the workload across a greater number of back-end resources and additional caching that are provided by the SVC cluster. Eventually, however, the performance of individual resources becomes the limiting factor.

8.2 Performance monitoring

In this section, we highlight several performance monitoring techniques.

8.2.1 Collecting performance statistics

The SVC is constantly collecting performance statistics. The default frequency by which files are created is at 5-minute intervals. The collection interval can be changed using the **startstats** command. The statistics files (named VDisk, MDisk, and Node) are saved at the end of the sampling interval and a maximum of 16 files (each) are stored before they are overlaid in a rotating log fashion. This design provides statistics for the most recent 80-minute period if using the default five-minute sampling interval. The SVC supports user-defined sampling intervals of from 1 to 60 minutes. The maximum space that is required for a performance statistics file is 1,153,482 bytes. There can be up to 128 (16 per each of the three types across eight nodes) different files across eight SVC nodes. This design makes the total space requirement a maximum of 147,645,694 bytes for all performance statistics from all nodes in an SVC cluster. Make note of this maximum when you are in time-critical situations. The required size is not otherwise important because SVC node hardware is more than capable.

You can define the sampling interval by using the **startstats -interval 2** command to collect statistics at 2-minute intervals; see 9.9.7, “Starting statistics collection” on page 227.

> **Collection intervals:** Although more frequent collection intervals provide a more detailed view of what happens within the SVC, they shorten the amount of time that the historical data is available on the SVC. For example, instead of an 80-minute period of data with the default 5-minute interval, if you adjust to 2-minute intervals, you have a 32-minute period instead.

Since SVC 5.1.0, cluster-level statistics are no longer supported. Instead, use the per node statistics that are collected. The sampling of the internal performance counters is coordinated across the cluster so that when a sample is taken, all nodes sample their internal counters at the same time. It is important to collect all files from all nodes for a complete analysis. Tools, such as Tivoli Storage Productivity Center, perform this intensive data collection for you.

Statistics file naming

The files that are generated are written to the `/dumps/iostats/` directory. The file name is in the following formats:

- `Nm_stats_`<*node_serial_number*>`_`<*date*>`_`<*time*> for MDisk statistics
- `Nv_stats_`<*node_serial_number*>`_`<*date*>`_`<*time*> for VDisk statistics
- `Nn_stats_`<*node_serial_number*>`_`<*date*>`_`<*time*> for node statistics
- `Nd_stats_`<*node_serial_number*>`_`<*date*>`_`<*time*> for disk drive statistics, not used for SVC

The *node_serial_number* is of the node on which the statistics were collected. The date is in the form <*yymmdd*> and the time is in the form <*hhmmss*>. The following example shows an MDisk statistics file name:

```
Nm_stats_KD8P1CG_140515_171848
```

Figure 8-1 shows typical MDisk Volume, Node, and disk drive statistics file names.

```
IBM_2145:ITSO_SVC2:admin>svcinfo lsiostatsdumps
id iostat_filename
0  Nd_stats_KD8P1CG_140515_171848
1  Nv_stats_KD8P1BP_140519_103220
2  Nd_stats_KD8P1BP_140519_103220
3  Nn_stats_KD8P1BP_140519_103220
4  Nm_stats_KD8P1BP_140519_103220
```

Figure 8-1 Shows the file names for the different types of statistics collected

> **Tip:** The performance statistics files can be copied from the SVC nodes to a local drive on your workstation using the `pscp.exe` (included with PuTTY) from an MS-DOS command line, as shown in this example:
>
> ```
> C:\Program Files\PuTTY>pscp -unsafe -load ITSO-SVC3
> admin@10.18.229.81:/dumps/iostats/* c:\statsfiles
> ```
>
> Use the **`-load`** parameter to specify the session that is defined in PuTTY.
>
> Specify the **`-unsafe`** parameter when you use wildcards.

The performance statistics files are in `.xml` format. They can be manipulated using various tools and techniques. An example of a tool that you can use to analyze these files is the SVC Performance Monitor (**`svcmon`**).

> **svcmon tool:** The **`svcmon`** tool is not an officially supported tool. It is provided on an “as is” basis.

You can obtain this tool from the following website:

https://www.ibm.com/developerworks/community/blogs/svcmon/entry/svcmon_available?lang=en

8.2.2 Real-time performance monitoring

Starting with Version 6.2.0, SVC supports real-time performance monitoring. Real-time performance statistics provide short-term status information for the SVC. The statistics are shown as graphs in the management graphical user interface (GUI) or can be viewed from the command-line interface (CLI). With system-level statistics, you can quickly view the CPU utilization and the bandwidth of volumes, interfaces, and MDisks. Each graph displays the current bandwidth in either megabytes per second (MBps) or I/Os per second (IOPS), as well as a view of bandwidth over time. Each node collects various performance statistics, mostly at 5-second intervals, and the statistics that are available from the config node in a clustered environment. This information can help you determine the performance effect of a specific node. As with system statistics, node statistics help you to evaluate whether the node is operating within normal performance metrics.

Real-time performance monitoring gathers the following system-level performance statistics:

- CPU utilization

- Port utilization and I/O rates
- Volume and MDisk I/O rates
- Bandwidth
- Latency

Real-time statistics are not a configurable option and cannot be disabled.

Real-time performance monitoring with the CLI

The following commands are available for monitoring the statistics through the CLI: **lsnodestats** and **lssystemstats**. Next, we show you examples of how to use them.

The **lsnodestats** command provides performance statistics for the nodes that are part of a clustered system, as shown in Example 8-1 (note that the output is truncated and shows only part of the available statistics). You can also specify a node name in the command to limit the output for a specific node.

Example 8-1 lsnodestats command output

```
IBM_2145:ITSO_SVC3:superuser>lsnodestats
node_id node_name  stat_name         stat_current stat_peak stat_peak_time
1       ITSOSVC3N1 cpu_pc            1            2         111003154220
1       ITSOSVC3N1 fc_mb             0            9         111003154220
1       ITSOSVC3N1 fc_io             1724         1799      111003153930
...
2       ITSOSVC3N2 cpu_pc            1            1         111003154246
2       ITSOSVC3N2 fc_mb             0            0         111003154246
2       ITSOSVC3N2 fc_io             1689         1770      111003153857
...
```

The previous example shows statistics for the two node members of cluster ITSO_SVC3: nodes ITSOSVC3N1 and ITSOSVC3N2. For each node, the following columns are displayed:

- stat_name: Provides the name of the statistic field
- stat_current: The current value of the statistic field
- stat_peak: The peak value of the statistic field in the last 5 minutes
- stat_peak_time: The time that the peak occurred

On the other side, the **lssystemstats** command lists the same set of statistics listed with the **lsnodestats** command, but representing all nodes in the cluster. The values for these statistics are calculated from the node statistics values in the following way:

- Bandwidth: Sum of bandwidth of all nodes
- Latency: Average latency for the cluster, which is calculated using data from the whole cluster, not an average of the single node values
- IOPS: Total IOPS of all nodes
- CPU percentage: Average CPU percentage of all nodes

Example 8-2 shows the resulting output of the **lssystemstats** command.

Example 8-2 lssystemstats command output

```
IBM_2145:ITSO_SVC3:superuser>lssystemstats
stat_name          stat_current stat_peak stat_peak_time
cpu_pc             1            1         111003160859
fc_mb              0            0         111003160859
fc_io              1291         1420      111003160504
...
```

Table 8-1 has a brief description of each of the statistics presented by the **lssystemstats** and **lsnodestats** commands.

Table 8-1 lssystemstats and lsnodestats statistics field name descriptions

Field name	Unit	Description
cpu_pc	Percentage	Utilization of node CPUs
fc_mb	MBps	Fibre Channel bandwidth
fc_io	IOPS	Fibre Channel throughput
sas_mb	MBps	SAS bandwidth
sas_io	IOPS	SAS throughput
iscsi_mb	MBps	IP-based Small Computer System Interface (iSCSI) bandwidth
iscsi_io	IOPS	iSCSI throughput
write_cache_pc	Percentage	Write cache fullness. Updated every 10 seconds.
total_cache_pc	Percentage	Total cache fullness. Updated every 10 seconds.
vdisk_mb	MBps	Total VDisk bandwidth
vdisk_io	IOPS	Total VDisk throughput
vdisk_ms	Milliseconds	Average VDisk latency
mdisk_mb	MBps	MDisk (SAN and RAID) bandwidth
mdisk_io	IOPS	MDisk (SAN and RAID) throughput
mdisk_ms	Milliseconds	Average MDisk latency
drive_mb	MBps	Drive bandwidth
drive_io	IOPS	Drive throughput
drive_ms	Milliseconds	Average drive latency
vdisk_w_mb	MBps	VDisk write bandwidth
vdisk_w_io	IOPS	VDisk write throughput
vdisk_w_ms	Milliseconds	Average VDisk write latency
mdisk_w_mb	MBps	MDisk (SAN and RAID) write bandwidth
mdisk_w_io	IOPS	MDisk (SAN and RAID) write throughput
mdisk_w_ms	Milliseconds	Average MDisk write latency

Field name	Unit	Description
drive_w_mb	MBps	Drive write bandwidth
drive_w_io	IOPS	Drive write throughput
drive_w_ms	Milliseconds	Average drive write latency
vdisk_r_mb	MBps	VDisk read bandwidth
vdisk_r_io	IOPS	VDisk read throughput
vdisk_r_ms	Milliseconds	Average VDisk read latency
mdisk_r_mb	MBps	MDisk (SAN and RAID) read bandwidth
mdisk_r_io	IOPS	MDisk (SAN and RAID) read throughput
mdisk_r_ms	Milliseconds	Average MDisk read latency
drive_r_mb	MBps	Drive read bandwidth
drive_r_io	IOPS	Drive read throughput
drive_r_ms	Milliseconds	Average drive read latency

Real-time performance monitoring with the GUI

The real-time statistics are also available from the SVC GUI. Go to **Monitoring** → **Performance**, as shown in Figure 8-2, to open the Performance Monitoring window.

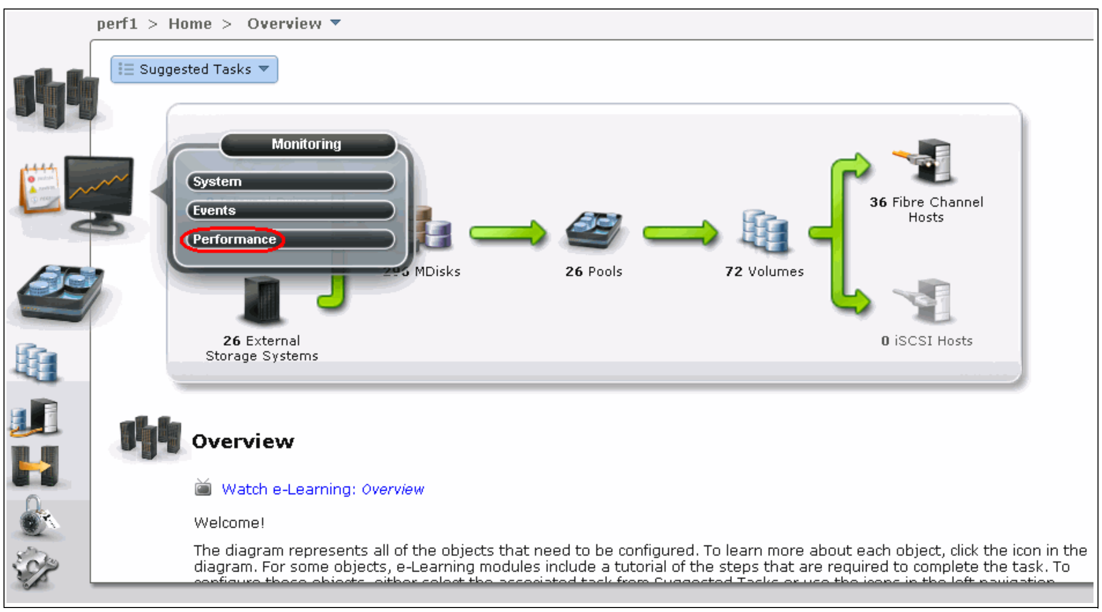


Figure 8-2 SAN Volume Controller Monitoring menu

The Performance Monitoring window, as shown in Figure 8-3 on page 161, is divided into four sections that provide utilization views for the following resources:

- CPU Utilization
  - Shows the CPU usage for general SVC tasks %
  - Shows the CPU usage for compression (when enabled)%

- Volumes

  Shows the overall volume utilization with the following fields:

  – Read
  – Write
  – Read latency
  – Write latency

- Interfaces

  Shows the overall statistics for each of the available interfaces:

  – Fibre Channel
  – iSCSI
  – SAS
  – IP Replication

- MDisks

  Shows the following overall statistics for the MDisks:

  – Read
  – Write
  – Read latency
  – Write latency

Figure 8-3 shows Real Time Performance Graphs.

Figure 8-3 SVC Real Time Performance Graphs

You can also select to view performance statistics for each of the available nodes of the system, as shown in Figure 8-4.

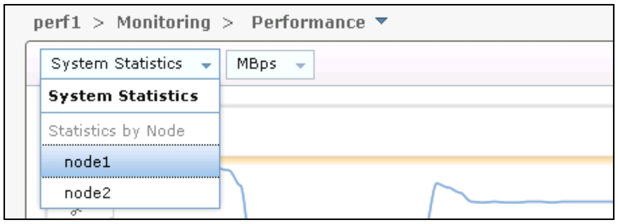


Figure 8-4 Select system node

It is also possible to change the metric between MBps or IOPS (Figure 8-5).

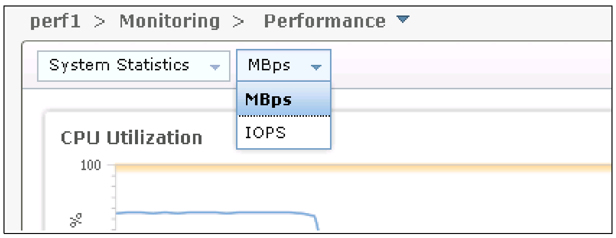


Figure 8-5 Changing to MBps

On any of these views, you can select any point in time with your cursor to know the exact value and when it occurred. As soon as you place your cursor over the timeline, it becomes a dotted line with the various values gathered (Figure 8-6).

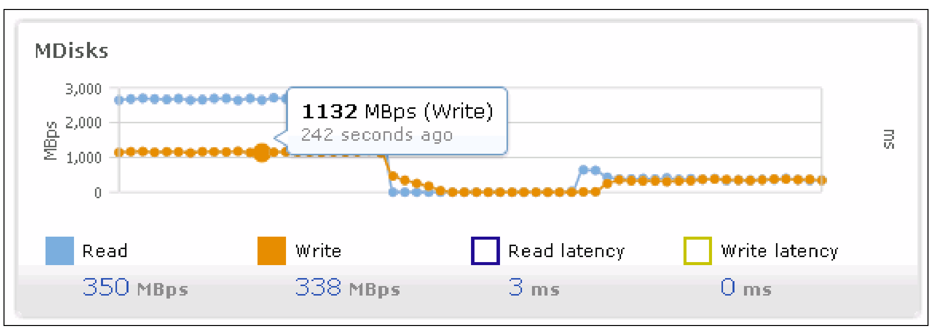


Figure 8-6 Detailed resource utilization

For each of the resources, there are various values that you can view by selecting the check box next to a value. For example, for the MDisks view, as shown in Figure 8-7, the four available fields are selected: Read, Write, Read latency, and Write latency.

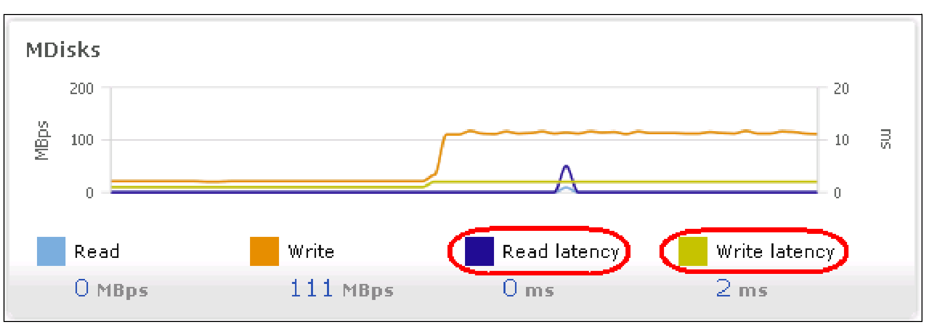


Figure 8-7 Detailed resource utilization

8.2.3 Performance data collection and Tivoli Storage Productivity Center for Disk

Although you can obtain performance statistics in standard `.xml` files, using `.xml` files is a less practical and less user-friendly method to analyze the SVC performance statistics. Tivoli Storage Productivity Center for Disk is the supported IBM tool to collect and analyze SVC performance statistics.

For more information about using Tivoli Storage Productivity Center to monitor your storage subsystem, see *SAN Storage Performance Management Using Tivoli Storage Productivity Center*, SG24-7364, which is available at the following website:

http://www.redbooks.ibm.com/abstracts/sg247364.html?Open

9

SAN Volume Controller operations using the command-line interface

In this chapter, we describe operational management. We use the command-line interface (CLI) to demonstrate both normal operation and then advanced operation.

You can use either the CLI or graphical user interface (GUI) to manage IBM System Storage SAN Volume Controller (SVC) operations. We use the CLI in this chapter. You can script these operations, and we think it is easier to create the documentation for the scripts using the CLI.

This chapter assumes a fully functional SAN Volume Controller environment.

9.1 Normal operations using CLI

In the following topics, we describe the commands that best represent normal operational commands.

9.1.1 Command syntax and online help

> **Command prefix changes:** The **svctask** and **svcinfo** command prefixes are no longer needed when issuing a command. If you have existing scripts that use those prefixes, they will continue to function. You do not need to change your scripts.

Two major command sets are available:

- The **svcinfo** command set allows you to query the various components within the SAN Volume Controller environment.
- The **svctask** command set allows you to make changes to the various components within the SAN Volume Controller.

When the command syntax is shown, you will see certain parameters in square brackets, for example `[parameter]`. These brackets indicate that the parameter is optional in most if not all instances. Any information that is not in square brackets is required information. You can view the syntax of a command by entering one of the following commands:

svcinfo -?	Shows a complete list of informational commands
svctask -?	Shows a complete list of task commands
svcinfo commandname -?	Shows the syntax of informational commands
svctask commandname -?	Shows the syntax of task commands
svcinfo commandname -filtervalue?	Shows the filters that you can use to reduce the output of the informational commands

> **Help:** You can also use **-h** instead of **-?**, for example, the **svcinfo -h** or **svctask commandname -h** command.

If you look at the syntax of the command by typing **svcinfo command name -?**, you often see **-filter** listed as a parameter. Be aware that the *correct* parameter is **-filtervalue**.

> **Tip:** You can use the up and down arrow keys on your keyboard to recall commands that were recently issued. Then, you can use the left and right, Backspace, and Delete keys to edit commands before you resubmit them.

Using shortcuts

You can use this command to display a list of display or execution commands. This command produces an alphabetical list of actions that are supported. The ***command*** parameter must be **svcinfo** for display commands or **svctask** for execution commands. The ***model*** parameter allows for different shortcuts on different platforms: 2145 or 2076.

```
<command> Shortcuts <model>
```

See Example 9-1 (lines have been removed from the command output for brevity).

Example 9-1 shortcuts command

```
IBM_2145:ITSO_SVC1:admin>svctask shortcuts 2145
addcontrolenclosure
addhostiogrp
addhostport
addmdisk
addnode
addvdiskcopy
applydrivesoftware
applysoftware
cancellivedump
cfgportip
chhost
chiogrp
chldap
chldapserver
chlicense
chmdisk
chmdiskgrp
chnode
chnodehw
chpartnership
chquorum
chrcconsistgrp
mkemailserver
mkemailuser
mkfcconsistgrp
mkfcmap
mkhost
mkldapserver
mkmdiskgrp
mkpartnership
mkrcconsistgrp
mkrcrelationship
mksnmpserver
mksyslogserver
mkuser
mkusergrp
mkvdisk
mkvdiskhostmap
prmmdisk
rmmdiskgrp
rmnode
rmpartnership
rmportip
rmrcconsistgrp
triggerlivedump
writesernum
```

Using reverse-i-search

If you work on your SAN Volume Controller with the same PuTTY session for many hours and enter many commands, scrolling back to find your previous or similar commands can be a time-intensive task. In this case, using the **reverse-i-search** command can help you quickly and easily find any command that you have already issued in the history of your commands by using the Ctrl+R keys. Ctrl+R will allow you to interactively search through the command history as you type commands. Pressing Ctrl+R at an empty command prompt will give you a prompt, as shown in Example 9-2.

Example 9-2 Using reverse-i-search

```
IBM_2145:ITSO_SVC1:admin>lsiogrp
id name           node_count vdisk_count host_count
0  io_grp0        2          10          8
1  io_grp1        2          10          8
2  io_grp2        0          0           0
3  io_grp3        0          0           0
4  recovery_io_grp 0         0           0
(reverse-i-search)`i': lsiogrp
```

As shown in Example 9-2, we had executed an **lsiogrp** command. By pressing Ctrl+R and typing **sv**, the command that we needed was recalled from history.

9.2 New commands

The following commands are introduced in version 7.3.0.0. They are mainly available for Stretched System configuration and IP replication. Example 9-3 provides the list of these commands.

Example 9-3 New commands in 7.3

```
Battery:
   chnodebattery
   lsnodebattery

Boot drive(s):
   chbootdrive
   chnodebootdrive
   lsbootnodedrive
   lsbootdrive
   satask chvpd

fan modules:
   lsenclosurefanmodule
```

SAN Volume Controller 7.3.0.0 also includes some changed commands, and has added some attributes and variables for existing commands. See the command reference or help for details.

The following examples shows in detail the new commands.

Example 9-4 on page 169 shows **chnodebattery**. Use the **chnodebattery** command to set or clear the light-emitting diode (LED) on a hot-swappable battery (in a node). This command applies to SAN Volume Controller 2145-DH8 systems.

Example 9-4 chnodebattery

```
chnodebattery -identify yes -battery 1 3

The resulting output:

No feedback

An invocation example to remove battery 1 in node 3

chnodebattery -remove -battery 1 3

The resulting output:

No feedback
```

Example 9-5 shows the **lsnodebattery** command. Use this command to display information about the batteries in a node. This command applies to SAN Volume Controller 2145-DH8 systems.

Example 9-5 lsnodebattery

```
lsnodebattery

The resulting output:

node_id node_name battery_id status  charging_status recondition_needed
node_percentage_charge   end_of_life_warning
1      node1    1        online  charged       no              50
no
1       node1     2          offline
2      node2    1        online  charged       no              50
no
2       node2     2          online  reconditioning  yes                 50

A concise invocation example

If battery 2 in node 1 is failing to charge:

lsnodebattery

The resulting output:

node_id node_name battery_id status   charging_status recondition_needed
node_percentage_charge   end_of_life_warning
1      node1    1        online   charged       no              50
no
1      node1    2        offline  idle          no              50
no
2      node2    1        online   charged       no              100
no
2       node2     2          online   charged         no                  100

A concise invocation example
```

```
If battery 2 in node 1 is removed (last known status is presented):

lsnodebattery

The resulting output:

node_id node_name battery_id status  charging_status recondition_needed
node_percentage_charge   end_of_life_warning
1     node1     1        online  charged       no             50
no
1     node1     2        offline charged       no             50
no
2     node2     1        online  charged       no             50
no
2       node2      2          online  reconditioning  yes                50

A detailed invocation example

lsnodebattery -battery 2 2

The resulting output:

node_id 2
node_name node2
battery_id 2
status online
charging_status reconditioning
present yes
redundant yes
recondition_needed yes
remaining_charge_capacity_mAh 1600
full_charge_capacity_mAh 1950
end_of_life_warning no
FRU_part_number FRU0001
FRU_identity 11SYM30BG123456MAN0001
compatability_level 5
last_recondition_timestamp 0
powered_on_hours 12345
cycle_count 2
node_percentage_charge 50
error_sequence_number
```

Example 9-6 shows **chbootdrive**. Use the **chbootdrive** command to synchronize a broken drive or field-replaceable unit (FRU) replacement drive. This command applies to SAN Volume Controller 2145-DH8 systems.

Example 9-6 chbootdrive

```
An invocation example

satask chbootdrive -sync

The resulting output:

No feedback
```

Example 9-7 shows **chnodebootdrive**. Use the **chnodebootdrive** command to identify a drive or synchronize the drives on a system if there is a broken drive or FRU replacement drive. This command applies to SAN Volume Controller 2145-DH8 systems.

Example 9-7 chnodebootdrive

```
An invocation example

chnodebootdrive -identify yes -slot 1 1

The resulting output:

No feedback
```

Example 9-8 shows **lsnodebootdrive**. Use the **lsnodebootdrive** command to return information about the internal boot drives for nodes. This command applies to SAN Volume Controller 2145-DH8 systems.

Example 9-8 lsnodebootdrive

```
In this example, bfn1, bfn2, bfn3, are online nodes and bfn4 is an offline node.

lsnodebootdrive -delim ' '

The resulting output:

node_id node_name can_sync slot_id booted status      actual_drive_sn
configured_drive_sn actual_node_sn
1      bfn1     no      1      yes    online     1234567        1234567
bfbfbf1
1       bfn1      no       2       no     missing                    1234568
1       bfn1      no       3              empty
1       bfn1      no       4              empty
1       bfn1      no       5              empty
1       bfn1      no       6              empty
1       bfn1      no       7              empty
1       bfn1      no       8       no     unsupported 12BD345
2      bfn2     no      1      yes    online     9234567        9234567
bfbfbf2
2      bfn2     no      2      no     wrong_node  1234568       9234577
bfbfbf2
3      bfn3     no      1      yes    online     3333333        3333333
bfbfbf3
3      bfn3     yes     2      no     out_of_sync 3333334       3333334
bfbfbf3
4      bfn4     unknown  1      yes    unknown    4444444       4444444
bfbfbf4
4      bfn4     unknown  2      no     unknown    4444445       4444445
bfbfbf4

A concise invocation example

lsnodebootdrive bfn4

The resulting output:
```

```
node_id node_name can_sync slot_id booted status      actual_drive_sn
configured_drive_sn actual_node_sn
4      bfn4     unknown  1      yes    unknown    4444444        4444444
bfbfbf4
4      bfn4     unknown  2      no     unknown    4444445        4444445
bfbfbf4

A detailed invocation example

lsnodebootdrive -slot 1 bfn1

The resulting output:

node_id 1
node_name bfn1
can_sync no
slot_id 1
booted yes
status online
actual_drive_sn 1234567
configured_drive_sn 1234567
actual_node_sn bfbfbf1
identify off
FRU_part_number 90Y9998
FRU_identity 11S49Y7427YXXX6X

A detailed invocation example

lsnodebootdrive -slot 1 homer4

The resulting output:

node_id 4
node_name homer4
can_sync unknown
slot_id 1
booted yes
status unknown
actual_drive_sn 4444444
configured_drive_sn 4444444
actual_node_sn bfbfbf4
identify off
FRU_part_number 90Y9999
FRU_identity 11S49Y7427YXXX6Y
```

Example 9-9 shows **lsbootdrive**. Use the **lsbootdrive** command to return the drive information of the internal boot drives for the given node. This command applies to SAN Volume Controller 2145-DH8 systems.

Example 9-9 lsbootdrive

```
An invocation example

sainfo lsbootdrive

The resulting output:
```

```
panel_id node_id node_name can_sync slot_id booted status     actual_drive_sn
configured_drive_sn actual_node_sn identify FRU_part_number FRU_identity
bfbfbf1  1       bfn1      yes      1       yes    online      1234567
1234567            bfbfbf1        off      90Y8879         11S49Y7427YXXX6XK
bfbfbf1  1       bfn1      no       2       no     missing
1234568                           off
bfbfbf1  1       bfn1      no       3              empty
bfbfbf1  1       bfn1      no       4              empty
bfbfbf1  1       bfn1      no       5              empty
bfbfbf1  1       bfn1      no       6              empty
bfbfbf1  1       bfn1      no       7              empty
bfbfbf1  1       bfn1      no       8       no     unsupported 12BD345
```

Example 9-10 shows **chvpd**. Use the **chvpd** command to set vital product data (VPD) such as serial number and machine type.

Example 9-10   chvpd

```
An invocation example

This example applies to SAN Volume Controller 2145-DH8 systems.

satask chvpd -wwnn 1111111111111111

The resulting output:

 No Feedback

An invocation example

satask chvpd -serial 8675309

The resulting output:

 No Feedback
```

Example 9-11shows **lsenclosurefanmodule**. Use the **lsenclosurefanmodule** command to report the status of each fan module and the contained fans in an enclosure.

Example 9-11   lsenclosurefanmodule

```
An invocation example

lsenclosurefanmodule 2

The resulting output:

enclosure_id fan_module_id status
2            1             online
2            2             online

An invocation example
```

```
lsenclosurefanmodule -fanmodule 1 1

The resulting output:

enclosure_id 1
fan_module_id 1
status online
error_sequence_number
FRU_part_number 31P1847
FRU_identity 11S31P1846YM10BG3B101N
```

9.3 Working with managed disks and disk controller systems

This section details the various configuration and administrative tasks that you can perform on the managed disks (MDisks) within the SAN Volume Controller environment and the tasks that you can perform at a disk controller level.

9.3.1 Viewing disk controller details

Use the **lscontroller** command to display summary information about all available back-end storage systems.

To display more detailed information about a specific controller, run the command again and append the controller name parameter, for example, controller ID 2, as shown in Example 9-12.

Example 9-12 lscontroller command

```
IBM_2145:ITSO_SVC1:admin>lscontroller 2
id 2
controller_name DS3500
WWNN 20080080E51B09E8
mdisk_link_count 10
max_mdisk_link_count 10
degraded no
vendor_id LSI
product_id_low INF-01-0
product_id_high 0
product_revision 0770
ctrl_s/n    b    Ns M
allow_quorum yes
WWPN 20680080E51B09E8
path_count 12
max_path_count 24
WWPN 20690080E51B09E8
path_count 8
max_path_count 20
WWPN 20580080E51B09E8
path_count 12
max_path_count 12
WWPN 20590080E51B09E8
path_count 8
max_path_count 20
IBM_2145:ITSO_SVC1:admin>
```

9.3.2 Renaming a controller

Use the **chcontroller** command to change the name of a storage controller. To verify the change, run the **lscontroller** command. Example 9-13 shows both of these commands.

Example 9-13 chcontroller command

```
IBM_2145:ITSO_SVC1:admin>chcontroller -name ITSO-DS3500 DS3500
IBM_2145:ITSO_SVC1:admin>lscontroller
id controller_name ctrl_s/n           vendor_id          product_id_low
product_id_high
0  ITSO-DS5000                        LSI                INF-01-0
0
2  ITSO-DS3500       b    Ns M        LSI                INF-01-0
0
IBM_2145:ITSO_SVC1:admin>
```

This command renames the controller named `controller0` to `DS4500`.

Choosing a new name: The **chcontroller** command specifies the new name first. You can use letters A to Z, a to z, numbers 0 - 9, the dash (-), and the underscore (_). The new name can be between 1 and 63 characters in length. However, the new name cannot start with a number, dash, or the word "controller" (because this prefix is reserved for SAN Volume Controller assignment only).

9.3.3 Discovery status

Use the **lsdiscoverystatus** command, as shown in Example 9-14, to determine if a discovery operation is in progress. The output of this command is a status of active or inactive.

Example 9-14 lsdiscoverystatus command

```
IBM_2145:ITSO_SVC1:admin>lsdiscoverystatus
id scope     IO_group_id IO_group_name status
0  fc_fabric                         inactive
```

This command displays the state of all discoveries in the clustered system. During discovery, the system updates the drive and MDisk records. You must wait until the discovery has finished and is inactive before you attempt to use the system. This command displays one of the following results:

- active: There is a discovery operation in progress at the time that the command is issued.
- inactive: There are no discovery operations in progress at the time that the command is issued.

9.3.4 Discovering MDisks

In general, the clustered system detects the MDisks automatically when they appear in the network. However, certain Fibre Channel (FC) controllers do not send the required Small Computer System Interface (SCSI) primitives that are necessary to automatically discover the new MDisks.

If new storage has been attached and the clustered system has not detected it, it might be necessary to run this command before the system can detect the new MDisks.

Use the **detectmdisk** command to scan for newly added MDisks (Example 9-15).

Example 9-15 detectmdisk

```
IBM_2145:ITSO_SVC1:admin>detectmdisk
```

To check whether any newly added MDisks were successfully detected, run the **lsmdisk** command and look for new unmanaged MDisks.

If the disks do not appear, check that the disk is appropriately assigned to the SAN Volume Controller in the disk subsystem, and that the zones are set up properly.

> **Discovery process:** If you have assigned a large number of logical unit numbers (LUNs) to your SAN Volume Controller, the discovery process can take time. Check several times by using the **lsmdisk** command to see if all the expected MDisks are present.

When all the disks allocated to the SAN Volume Controller are seen from the SAN Volume Controller system, the following procedure is a useful way to verify which MDisks are unmanaged and ready to be added to the storage pool.

Perform the following steps to display MDisks:

1. Enter the **lsmdiskcandidate** command, as shown in Example 9-16. This command displays all detected MDisks that are not currently part of a storage pool.

   Example 9-16 lsmdiskcandidate command

   ```
   IBM_2145:ITSO_SVC1:admin>lsmdiskcandidate
   id
   0
   1
   2
   .
   .
   ```

   Alternatively, you can list all MDisks (managed or unmanaged) by issuing the **lsmdisk** command, as shown in Example 9-17.

   Example 9-17 lsmdisk command

   ```
   IBM_2145:ITSO_SVC1:admin>lsmdisk -filtervalue controller_name=ITSO-DS3500
   id name   status mode      mdisk_grp_id mdisk_grp_name capacity ctrl_LUN_#
   controller_name UID tier
   0  mdisk0 online managed   0 STGPool_DS3500-1 128.0GB  0000000000000000 ITSO-DS3500
   60080e50001b0b62000007b04e731e4d00000000000000000000000000000000 generic_hdd
   1  mdisk1 online managed   0 STGPool_DS3500-1 128.0GB  0000000000000001 ITSO-DS3500
   60080e50001b0b62000007b24e731e6000000000000000000000000000000000 generic_hdd
   2  mdisk2 online managed   0 STGPool_DS3500-1 128.0GB  0000000000000002 ITSO-DS3500
   60080e50001b09e8000006f44e731bdc00000000000000000000000000000000 generic_hdd
   3  mdisk3 online managed   1 STGPool_DS3500-2 128.0GB  0000000000000003 ITSO-DS3500
   60080e50001b0b62000007b44e731e8400000000000000000000000000000000 generic_hdd
   4  mdisk4 online managed   1 STGPool_DS3500-2 128.0GB  0000000000000004 ITSO-DS3500
   60080e50001b09e8000006f64e731bff00000000000000000000000000000000 generic_hdd
   5  mdisk5 online managed   1 STGPool_DS3500-2 128.0GB  0000000000000005 ITSO-DS3500
   60080e50001b0b62000007b64e731ea900000000000000000000000000000000 generic_hdd
   ```

```
6  mdisk6 online unmanaged 10.0GB   0000000000000006 ITSO-DS3500
60080e50001b09e80000085f4e7d60dd00000000000000000000000000000000 generic_hdd
```

From this output, you can see additional information, such as the current status, about each MDisk. For the purpose of our current task, we are only interested in the unmanaged disks because they are candidates for a storage pool.

Tip: The **`-delim`** parameter collapses output instead of wrapping text over multiple lines.

2. If not all the MDisks that you expected are visible, rescan the available FC network by entering the **`detectmdisk`** command, as shown in Example 9-18.

Example 9-18 detectmdisk

```
IBM_2145:ITSO_SVC1:admin>detectmdisk
```

3. If you run the **`lsmdiskcandidate`** command again and your MDisk or MDisks are still not visible, check that the LUNs from your subsystem have been properly assigned to the SAN Volume Controller and that appropriate zoning is in place (for example, the SAN Volume Controller can see the disk subsystem). See Chapter 4, “Planning and configuration” on page 57 for details about setting up your storage area network (SAN) fabric.

9.3.5 Viewing MDisk information

When viewing information about the MDisks (managed or unmanaged), we can use the **`lsmdisk`** command to display overall summary information about all available managed disks. To display more detailed information about a specific MDisk, run the command again and append the **`-mdisk`** name parameter (for example, `mdisk0`).

The overview command is **`lsmdisk -delim`**, as shown in Example 9-19.

The summary for an individual MDisk is **`lsmdisk`** (*name/ID* of the MDisk from which you want the information), as shown in Example 9-20 on page 178.

Example 9-19 lsmdisk command

```
IBM_2145:ITSO_SVC1:admin>lsmdisk -delim :
id:name:status:mode:mdisk_grp_id:mdisk_grp_name:capacity:ctrl_LUN_#:controller_name:UID:tie
r
0:mdisk0:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000000:ITSO-DS3500:60080e50001
b0b62000007b04e731e4d00000000000000000000000000000000:generic_hdd
1:mdisk1:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000001:ITSO-DS3500:60080e50001
b0b62000007b24e731e6000000000000000000000000000000000:generic_hdd
2:mdisk2:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000002:ITSO-DS3500:60080e50001
b09e8000006f44e731bdc00000000000000000000000000000000:generic_hdd
3:mdisk3:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000003:ITSO-DS3500:60080e50001
b0b62000007b44e731e8400000000000000000000000000000000:generic_hdd
4:mdisk4:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000004:ITSO-DS3500:60080e50001
b09e8000006f64e731bff00000000000000000000000000000000:generic_hdd
5:mdisk5:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000005:ITSO-DS3500:60080e50001
b0b62000007b64e731ea900000000000000000000000000000000:generic_hdd
6:mdisk6:online:unmanaged:::10.0GB:0000000000000006:ITSO-DS3500:60080e50001b09e80000085f4e7
d60dd00000000000000000000000000000000:generic_hdd
```

Example 9-20 shows a summary for a single MDisk.

Example 9-20 Usage of the command lsmdisk (ID)

```
IBM_2145:ITSO_SVC1:admin>lsmdisk 0
id 0
name mdisk0
status online
mode managed
mdisk_grp_id 0
mdisk_grp_name STGPool_DS3500-1
capacity 128.0GB
quorum_index 1
block_size 512
controller_name ITSO-DS3500
ctrl_type 4
ctrl_WWNN 20080080E51B09E8
controller_id 2
path_count 4
max_path_count 4
ctrl_LUN_# 0000000000000000
UID 60080e50001b0b62000007b04e731e4d00000000000000000000000000000000
preferred_WWPN 20580080E51B09E8
active_WWPN 20580080E51B09E8
fast_write_state empty
raid_status
raid_level
redundancy
strip_size
spare_goal
spare_protection_min
balanced
tier generic_hdd
```

9.3.6 Renaming an MDisk

Use the **chmdisk** command to change the name of an MDisk. When using the command, be aware that the new name comes first and then the ID/name of the MDisk being renamed. Use this format: **chmdisk -name (*new name*) (*current ID*/*name*)**. Use the **lsmdisk** command to verify the change. Example 9-21 show both of these commands.

Example 9-21 chmdisk command

```
IBM_2145:ITSO_SVC1:admin>chmdisk -name mdisk_0 mdisk0
```

This command renamed the MDisk named `mdisk0` to `mdisk_0`.

> **The chmdisk command:** The **chmdisk** command specifies the new name first. You can use letters A to Z, a to z, numbers 0 - 9, the dash (-), and the underscore (_). The new name can be between 1 and 63 characters in length. However, the new name cannot start with a number, dash, or the word "`mdisk`" (because this prefix is reserved for SAN Volume Controller assignment only).

9.3.7 Including an MDisk

If a significant number of errors occur on an MDisk, the SAN Volume Controller automatically excludes it. These errors can result from a hardware problem, a SAN problem, or the result of poorly planned maintenance. If it is a hardware fault, you can receive a Simple Network Management Protocol (SNMP) alert about the state of the disk subsystem (before the disk was excluded), and you can undertake preventive maintenance. If not, the hosts that were using virtual disks (VDisks), which used the excluded MDisk, now have I/O errors.

By running the **lsmdisk** command, you can see that mdisk0 is excluded in Example 9-22.

Example 9-22 lsmdisk command: Excluded MDisk

```
IBM_2145:ITSO_SVC1:admin>lsmdisk -delim :
id:name:status:mode:mdisk_grp_id:mdisk_grp_name:capacity:ctrl_LUN_#:controller_name:UID:tie
r
0:mdisk0:excluded:managed:0:STGPool_DS3500-1:128.0GB:0000000000000000:ITSO-DS3500:60080e500
01b0b62000007b04e731e4d00000000000000000000000000000000:generic_hdd
1:mdisk1:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000001:ITSO-DS3500:60080e50001
b0b62000007b24e731e6000000000000000000000000000000000:generic_hdd
2:mdisk2:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000002:ITSO-DS3500:60080e50001
b09e8000006f44e731bdc00000000000000000000000000000000:generic_hdd
3:mdisk3:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000003:ITSO-DS3500:60080e50001
b0b62000007b44e731e8400000000000000000000000000000000:generic_hdd
4:mdisk4:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000004:ITSO-DS3500:60080e50001
b09e8000006f64e731bff00000000000000000000000000000000:generic_hdd
5:mdisk5:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000005:ITSO-DS3500:60080e50001
b0b62000007b64e731ea900000000000000000000000000000000:generic_hdd
6:mdisk6:online:unmanaged:::10.0GB:0000000000000006:ITSO-DS3500:60080e50001b09e80000085f4e7
d60dd00000000000000000000000000000000:generic_hdd
```

After taking the necessary corrective action to repair the MDisk (for example, replace the failed disk and repair the SAN zones), we need to include the MDisk again. We issue the **includemdisk** command (Example 9-23) because the SAN Volume Controller system does not include the MDisk automatically.

Example 9-23 includemdisk

```
IBM_2145:ITSO_SVC1:admin>includemdisk mdisk0
```

Running the **lsmdisk** command again shows that mdisk0 is online again; see Example 9-24.

Example 9-24 lsmdisk command: Verifying that MDisk is included

```
IBM_2145:ITSO_SVC1:admin>lsmdisk -delim :
id:name:status:mode:mdisk_grp_id:mdisk_grp_name:capacity:ctrl_LUN_#:controller_name:UID:tie
r
0:mdisk0:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000000:ITSO-DS3500:60080e50001
b0b62000007b04e731e4d00000000000000000000000000000000:generic_hdd
1:mdisk1:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000001:ITSO-DS3500:60080e50001
b0b62000007b24e731e6000000000000000000000000000000000:generic_hdd
2:mdisk2:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000002:ITSO-DS3500:60080e50001
b09e8000006f44e731bdc00000000000000000000000000000000:generic_hdd
3:mdisk3:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000003:ITSO-DS3500:60080e50001
b0b62000007b44e731e8400000000000000000000000000000000:generic_hdd
4:mdisk4:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000004:ITSO-DS3500:60080e50001
b09e8000006f64e731bff00000000000000000000000000000000:generic_hdd
5:mdisk5:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000005:ITSO-DS3500:60080e50001
b0b62000007b64e731ea900000000000000000000000000000000:generic_hdd
```

```
6:mdisk6:online:unmanaged:::10.0GB:0000000000000006:ITSO-DS3500:60080e50001b09e80000085f4e7
d60dd00000000000000000000000000000000:generic_hdd
```

9.3.8 Adding MDisks to a storage pool

If you created an empty storage pool or you simply assign additional MDisks to your already configured storage pool, you can use the **addmdisk** command to populate the storage pool (Example 9-25).

Example 9-25 addmdisk command

```
IBM_2145:ITSO_SVC1:admin>addmdisk -mdisk mdisk6 STGPool_Multi_Tier
```

You can *only* add unmanaged MDisks to a storage pool. This command adds the MDisk named mdisk6 to the storage pool named STGPool_Multi_Tier.

Important: Do not add this MDisk to a storage pool if you want to create an image mode volume from the MDisk that you are adding. As soon as you add an MDisk to a storage pool, it becomes managed, and extent mapping is not necessarily one-to-one anymore.

9.3.9 Showing MDisks in a storage pool

Use the **lsmdisk -filtervalue** command, as shown in Example 9-26, to see which MDisks are part of a specific storage pool. This command shows all the MDisks that are part of a storage pool if they belong to the Storage Subsystem named STGPool_DS3500-1.

Example 9-26 lsmdisk -filtervalue: MDisks in the managed disk group (MDG)

```
IBM_2145:ITSO_SVC1:admin>lsmdisk -filtervalue mdisk_grp_name=STGPool_DS3500-1
id name   status mode    mdisk_grp_id mdisk_grp_name capacity ctrl_LUN_#
controller_name UID tier
0  mdisk0 online managed 0 STGPool_DS3500-1 128.0GB  0000000000000000 DS3500
60080e50001b0b62000007b04e731e4d00000000000000000000000000000000 generic_hdd
1  mdisk1 online managed 0 STGPool_DS3500-1 128.0GB  0000000000000001 DS3500
60080e50001b0b62000007b24e731e6000000000000000000000000000000000 generic_hdd
2  mdisk2 online managed 0 STGPool_DS3500-1 128.0GB  0000000000000002 DS3500
60080e50001b09e8000006f44e731bdc00000000000000000000000000000000 generic_hdd
```

As you can see in Example 9-26, with this command, by using a wildcard, you will be able to see all the MDisks present in the storage pools named STGPool_* where the asterisk (*) is a wildcard.

9.3.10 Working with a storage pool

Before we can create any volumes on the SAN Volume Controller clustered system, we need to virtualize the allocated storage that is assigned to the SAN Volume Controller. After we have assigned volumes to the SAN Volume Controller’s “managed disks”, we cannot start using them until they are members of a storage pool. Therefore, one of our first operations is to create a storage pool where we can place our MDisks.

This section describes the operations using MDisks and the storage pool. It explains the tasks that we can perform at the storage pool level.

9.3.11 Creating a storage pool

After a successful login to the CLI interface of the SAN Volume Controller, we create the storage pool.

Using the **mkmdiskgrp** command, create a storage pool, as shown in Example 9-27.

Example 9-27 mkmdiskgrp

```
IBM_2145:ITSO_SVC1:admin>mkmdiskgrp -name STGPool_Multi_Tier -ext 256
MDisk Group, id [3], successfully created
```

This command creates a storage pool called STGPool_Multi_Tier. The extent size that is used within this group is 256 MB. We have not added any MDisks to the storage pool yet, so it is an empty storage pool.

You can add unmanaged MDisks and create the storage pool in the same command. Use the command **mkmdiskgrp** with the **-mdisk** parameter and enter the IDs or names of the MDisks, which adds the MDisks immediately after the storage pool is created.

Before the creation of the storage pool, enter the **lsmdisk** command, as shown in Example 9-28. This command lists all of the available MDisks that are seen by the SAN Volume Controller system.

Example 9-28 Listing available MDisks

```
IBM_2145:ITSO_SVC1:admin>lsmdisk -filtervalue controller_name=ITSO-DS3500 -delim :
id:name:status:mode:mdisk_grp_id:mdisk_grp_name:capacity:ctrl_LUN_#:controller_name:UID:tie
r
0:mdisk0:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000000:ITSO-DS3500:60080e50001
b0b62000007b04e731e4d00000000000000000000000000000000:generic_hdd
1:mdisk1:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000001:ITSO-DS3500:60080e50001
b0b62000007b24e731e6000000000000000000000000000000000:generic_hdd
2:mdisk2:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000002:ITSO-DS3500:60080e50001
b09e8000006f44e731bdc0000000000000000000000000000000:generic_hdd
3:mdisk3:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000003:ITSO-DS3500:60080e50001
b0b62000007b44e731e8400000000000000000000000000000000:generic_hdd
4:mdisk4:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000004:ITSO-DS3500:60080e50001
b09e8000006f64e731bff0000000000000000000000000000000:generic_hdd
5:mdisk5:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000005:ITSO-DS3500:60080e50001
b0b62000007b64e731ea9000000000000000000000000000000000:generic_hdd
6:mdisk6:online:unmanaged:::10.0GB:0000000000000006:DS3500:60080e50001b09e80000085f4e7d60dd
00000000000000000000000000000000:generic_hdd
8:mdisk7:online:unmanaged:::10.0GB:00
00000000000008:DS3500:60080e50001b09e8000008614e7d8a2c00000000000000000000000000000000:gene
ric_hdd
```

By using the same command as before (**mkmdiskgrp**) and knowing the MDisk IDs that we are using, we can add multiple MDisks to the storage pool at the same time. We now add the unmanaged MDisks to the storage pool that we created, as shown in Example 9-29.

Example 9-29 Creating a storage pool and adding available MDisks

```
IBM_2145:ITSO_SVC1:admin>mkmdiskgrp -name STGPool_DS5000 -ext 256 -mdisk 6:8
MDisk Group, id [2], successfully created
```

This command creates a storage pool called STGPool_DS5000. The extent size that is used within this group is 256 MB, and two MDisks (6 and 8) are added to the storage pool.

Storage pool name: The **-name** and **-mdisk** parameters are optional. If you do not enter a **-name**, the default is MDiskgrp*x*, where *x* is the ID sequence number that is assigned by the SAN Volume Controller internally. If you do not enter the **-mdisk** parameter, an empty storage pool is created.

If you want to provide a name, you can use letters A to Z, a to z, numbers 0 - 9, and the underscore. The name can be between 1 and 63 characters in length, but it cannot start with a number or the word "MDiskgrp" (because this prefix is reserved for SAN Volume Controller assignment only).

By running the **lsmdisk** command, you now see the MDisks as "managed" and as part of the STGPool_DS3500-1, as shown in Example 9-30.

Example 9-30 lsmdisk command

```
IBM_2145:ITSO_SVC1:admin>lsmdisk -filtervalue controller_name=ITSO-DS3500 -delim :
id:name:status:mode:mdisk_grp_id:mdisk_grp_name:capacity:ctrl_LUN_#:controller_name:UID:tie
r
0:mdisk0:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000000:ITSO-DS3500:60080e50001
b0b62000007b04e731e4d00000000000000000000000000000000:generic_hdd
1:mdisk1:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000001:ITSO-DS3500:60080e50001
b0b62000007b24e731e6000000000000000000000000000000000:generic_hdd
2:mdisk2:online:managed:0:STGPool_DS3500-1:128.0GB:0000000000000002:ITSO-DS3500:60080e50001
b09e8000006f44e731bdc00000000000000000000000000000000:generic_hdd
3:mdisk3:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000003:ITSO-DS3500:60080e50001
b0b62000007b44e731e8400000000000000000000000000000000:generic_hdd
4:mdisk4:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000004:ITSO-DS3500:60080e50001
b09e8000006f64e731bff00000000000000000000000000000000:generic_hdd
5:mdisk5:online:managed:1:STGPool_DS3500-2:128.0GB:0000000000000005:ITSO-DS3500:60080e50001
b0b62000007b64e731ea900000000000000000000000000000000:generic_hdd
6:mdisk6:online:managed:2:STGPool_DS3500-2:10.0GB:0000000000000006:ITSO-DS3500:60080e50001b
09e80000085f4e7d60dd00000000000000000000000000000000:generic_hdd
7:mdisk7:online:managed:3:STGPool_Multi_Tier:10.0GB:0000000000000007:ITSO-DS3500:60080e5000
1b0b620000091f4e7d8c9400000000000000000000000000000000:generic_hdd
8:mdisk8:online:managed:2:STGPool_DS3500-2:10.0GB:0000000000000008:ITSO-DS3500:60080e50001b
09e8000008614e7d8a2c00000000000000000000000000000000:generic_hdd
9:mdisk9:online:managed:3:STGPool_Multi_Tier:10.0GB:0000000000000009:ITSO-DS3500:60080e5000
1b0b62000009214e7d928000000000000000000000000000000000:generic_hdd
```

At this point, you have completed the tasks that are required to create a new storage pool.

9.3.12 Viewing storage pool information

Use the **lsmdiskgrp** command, as shown in Example 9-31, to display information about the storage pools that are defined in the SAN Volume Controller.

Example 9-31 lsmdiskgrp command

```
IBM_2145:ITSO_SVC1:admin>lsmdiskgrp -delim :
id:name:status:mdisk_count:vdisk_count:capacity:extent_size:free_capacity:virtual_capacity:
used_capacity:real_capacity:overallocation:warning:easy_tier:easy_tier_status
0:STGPool_DS3500-1:online:3:11:382.50GB:256:62.50GB:320.00GB:320.00GB:320.00GB:83:0:auto:in
active
1:STGPool_DS3500-2:online:3:11:384.00GB:256:262.00GB:122.00GB:122.00GB:122.00GB:31:0:auto:i
nactive
2:STGPool_DS5000-1:online:2:0:20.00GB:256:20.00GB:0.00MB:0.00MB:0.00MB:0:0:auto:inactive
```

```
3:STGPool_Multi_Tier:online:2:0:20.00GB:256:20.00GB:0.00MB:0.00MB:0.00MB:0:0:auto:inactive
```

9.3.13 Renaming a storage pool

Use the **chmdiskgrp** command to change the name of a storage pool. To verify the change, run the **lsmdiskgrp** command. Example 9-32 shows both of these commands.

Example 9-32 chmdiskgrp command

```
IBM_2145:ITSO_SVC1:admin>chmdiskgrp -name STGPool_DS3500-2_new 1
IBM_2145:ITSO_SVC1:admin>lsmdiskgrp -delim :
id:name:status:mdisk_count:vdisk_count:capacity:extent_size:free_capacity:virtual_
capacity:used_capacity:real_capacity:overallocation:warning:easy_tier:easy_tier_st
atus
0:STGPool_DS3500-1:online:3:11:382.50GB:256:62.50GB:320.00GB:320.00GB:320.00GB:83:
0:auto:inactive
1:STGPool_DS3500-2_new:online:3:11:384.00GB:256:262.00GB:122.00GB:122.00GB:122.00G
B:31:0:auto:inactive
2:STGPool_DS5000-1:online:2:0:20.00GB:256:20.00GB:0.00MB:0.00MB:0.00MB:0:0:auto:in
active
3:STGPool_Multi_Tier:online:2:0:20.00GB:256:20.00GB:0.00MB:0.00MB:0.00MB:0:0:auto:
inactive
```

This command renamed the storage pool `STGPool_DS3500-2` to `STGPool_DS3500-2_new` as shown.

> **Changing the storage pool:** The **chmdiskgrp** command specifies the new name first. You can use letters A to Z, a to z, numbers 0 - 9, the dash (-), and the underscore (_). The new name can be between 1 and 63 characters in length. However, the new name cannot start with a number, dash, or the word "`mdiskgrp`" (because this prefix is reserved for SAN Volume Controller assignment only).

9.3.14 Deleting a storage pool

Use the **rmmdiskgrp** command to remove a storage pool from the SAN Volume Controller system configuration (Example 9-33).

Example 9-33 rmmdiskgrp

```
IBM_2145:ITSO_SVC1:admin>rmmdiskgrp STGPool_DS3500-2_new
```

This command removes storage pool `STGPool_DS3500-2_new` from the SAN Volume Controller system configuration.

> **Removing a storage pool from the SAN Volume Controller system configuration:** If there are MDisks within the storage pool, you must use the **-force** flag to remove the storage pool from the SAN Volume Controller system configuration, for example:
>
> `rmmdiskgrp STGPool_DS3500-2_new -force`
>
> Ensure that you definitely want to use this flag because it destroys all mapping information and data held on the volumes, which cannot be recovered.

9.3.15 Removing MDisks from a storage pool

Use the **rmmdisk** command to remove an MDisk from a storage pool (Example 9-34).

Example 9-34 rmmdisk command

```
IBM_2145:ITSO_SVC1:admin>rmmdisk -mdisk 8 -force 2
```

This command removes the MDisk with ID 8 from the storage pool with ID 2. The **-force** flag is set because there are volumes using this storage pool.

Sufficient space: The removal only takes place if there is sufficient space to migrate the volume's data to other extents on other MDisks that remain in the storage pool. After you remove the MDisk from the storage pool, it takes time to change the mode from managed to unmanaged depending on the size of the MDisk that you are removing.

9.4 Working with hosts

In this section, we explain the tasks that you can perform at a host level. When we create a host in our SAN Volume Controller system, we need to define the connection method. Starting with SAN Volume Controller 5.1, we can now define our host as iSCSI-attached or FC-attached.

9.4.1 Creating a Fibre Channel-attached host

In the following sections, we illustrate how to create an FC-attached host under various circumstances.

Host is powered on, connected, and zoned to the SAN Volume Controller

When you create your host on the SAN Volume Controller, it is good practice to check whether the host bus adapter (HBA) worldwide port names (WWPNs) of the server are visible to the SAN Volume Controller. By checking, you ensure that zoning is done and that the correct WWPN will be used. Issue the **lshbaportcandidate** command, as shown in Example 9-35.

Example 9-35 lshbaportcandidate command

```
IBM_2145:ITSO_SVC1:admin>lshbaportcandidate
id
210000E08B89C1CD
210000E08B054CAA
```

After you know that the WWPNs that are displayed match your host (use host or SAN switch utilities to verify), use the **mkhost** command to create your host.

Name: If you do not provide the **-name** parameter, the SAN Volume Controller automatically generates the name host*x* (where *x* is the ID sequence number that is assigned by the SAN Volume Controller internally).

You can use the letters A to Z and a to z, the numbers 0 - 9, the dash (-), and the underscore (_). The name can be between 1 and 63 characters in length. However, the name cannot start with a number, dash, or the word "host" (because this prefix is reserved for SAN Volume Controller assignment only).

The command to create a host is shown in Example 9-36.

Example 9-36 mkhost

```
IBM_2145:ITSO_SVC1:admin>mkhost -name Almaden -hbawwpn
210000E08B89C1CD:210000E08B054CAA
Host, id [2], successfully created
```

This command creates a host called Almaden using WWPN 21:00:00:E0:8B:89:C1:CD and 21:00:00:E0:8B:05:4C:AA.

Ports: You can define from one up to eight ports per host, or you can use the **addport** command, which we show in 9.4.5, "Adding ports to a defined host" on page 188.

Host is not powered on or not connected to the SAN

If you want to create a host on the SAN Volume Controller without seeing your target WWPN by using the **lshbaportcandidate** command, add the **-force** flag to your **mkhost** command, as shown in Example 9-37. This option is more open to human error than if you choose the WWPN from a list, but it is typically used when many host definitions are created at the same time, such as through a script.

In this case, you can type the WWPN of your HBA or HBAs and use the **-force** flag to create the host, regardless of whether they are connected, as shown in Example 9-37.

Example 9-37 mkhost -force

```
IBM_2145:ITSO_SVC1:admin>mkhost -name Almaden -hbawwpn
210000E08B89C1CD:210000E08B054CAA -force
Host, id [2], successfully created
```

This command forces the creation of a host called Almaden using WWPN 210000E08B89C1CD:210000E08B054CAA.

WWPNs: WWPNs are *not* case-sensitive in the CLI.

9.4.2 Creating an iSCSI-attached host

Now, we can create host definitions to a host that is not connected to the SAN but that has LAN access to our SAN Volume Controller nodes. Before we create the host definition, we configure our SAN Volume Controller systems to use the new iSCSI connection method. We describe additional information about configuring your nodes to use iSCSI in 9.9.3, "iSCSI configuration" on page 223.

The iSCSI functionality allows the host to access volumes through the SAN Volume Controller without being attached to the SAN. Back-end storage and node-to-node communication still need the FC network to communicate, but the host does not necessarily need to be connected to the SAN.

When we create a host that is going to use iSCSI as a communication method, iSCSI initiator software must be installed on the host to initiate the communication between the SAN Volume Controller and the host. This installation creates an iSCSI qualified name (IQN) identifier that is needed before we create our host.

Before we start, we check our server's IQN address. We are running Windows Server 2008. We select **Start** → **Programs** → **Administrative tools**, and we select **iSCSI initiator**. In our example, our IQN, as shown in Figure 9-1, is:

```
iqn.1991-05.com.microsoft:st-2k8hv-004.englab.brocade.com
```

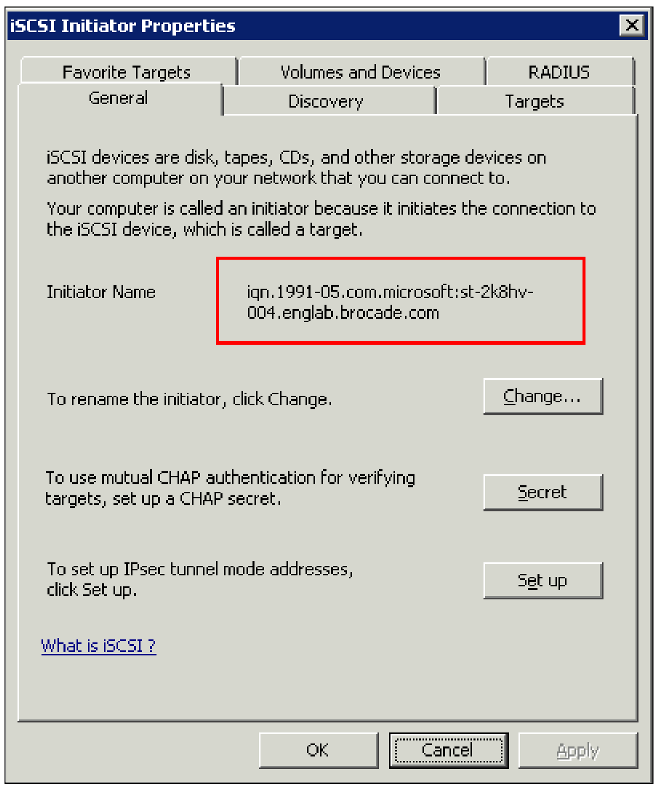


Figure 9-1 IQN from the iSCSI initiator tool

We create the host by issuing the **mkhost** command, as shown in Example 9-38. When the command completes successfully, we display our newly created host.

Example 9-38 mkhost command

```
IBM_2145:ITSO_SVC1:admin>mkhost -name Baldur -iogrp 0 -iscsiname
iqn.1991-05.com.microsoft:st-2k8hv-004.englab.brocade.com
Host, id [4], successfully created
IBM_2145:ITSO_SVC1:admin>lshost 4
id 4
name Baldur
port_count 1
type generic
mask 1111
```

```
iogrp_count 1
iscsiname iqn.1991-05.com.microsoft:st-2k8hv-004.englab.brocade.com
node_logged_in_count 0
state offline
```

It is important to know that when the host is initially configured, the default authentication method is set to no authentication and no Challenge Handshake Authentication Protocol (CHAP) secret is set. To set a CHAP secret for authenticating the iSCSI host with the SAN Volume Controller system, use the **chhost** command with the **chapsecret** parameter.

We have now created our host definition. We map a volume to our new iSCSI server, as shown in Example 9-39. We have already created the volume, as shown in 9.6.1, "Creating a volume" on page 191. In our scenario, our volume's ID is `21` and the host name is `Baldur`. We map it to our iSCSI host.

Example 9-39 Mapping a volume to the iSCSI host

```
IBM_2145:ITSO_SVC1:admin>mkvdiskhostmap -host Baldur 21
Virtual Disk to Host map, id [0], successfully created
```

After the volume has been mapped to the host, we display the host information again, as shown in Example 9-40.

Example 9-40 lshost

```
IBM_2145:ITSO_SVC1:admin>lshost 4
id 4
name Baldur
port_count 1
type generic
mask 1111
iogrp_count 1
iscsiname iqn.1991-05.com.microsoft:st-2k8hv-004.englab.brocade.com
node_logged_in_count 1
state online
```

Tip: FC hosts and iSCSI hosts are handled in the same way operationally after they have been created.

If you need to display a CHAP secret for an already defined server, use the **lsiscsiauth** command. The **lsiscsiauth** command lists the CHAP secret configured for authenticating an entity to the SAN Volume Controller system.

9.4.3 Modifying a host

Use the **chhost** command to change the name of a host. To verify the change, run the **lshost** command. Example 9-41 shows both of these commands.

Example 9-41 chhost command

```
IBM_2145:ITSO_SVC1:admin>chhost -name Angola Guinea

IBM_2145:ITSO_SVC1:admin>lshost
id              name            port_count     iogrp_count
0               Palau           2              4
```

```
1              Nile          2           1
2              Kanaga        2           1
3              Siam          2           2
4              Angola        1           4
```

This command renamed the host from `Guinea` to `Angola`.

Host name: The **`chhost`** command specifies the *new name first.* You can use letters A to Z and a to z, numbers 0 - 9, the dash (-), and the underscore (_). The new name can be between 1 and 63 characters in length. However, it cannot start with a number, dash, or the word “`host`” (because this prefix is reserved for SAN Volume Controller assignment only).

Hosts that require the -type parameter: If you use Hewlett-Packard UNIX (HP-UX), you use the **`-type`** option. See *IBM System Storage Open Software Family SAN Volume Controller: Host Attachment Guide*, SC26-7563, for more information about the hosts that require the **`-type`** parameter.

9.4.4 Deleting a host

Use the **`rmhost`** command to delete a host from the SAN Volume Controller configuration. If your host is still mapped to volumes and you use the **`-force`** flag, the host and all the mappings with it are deleted. The volumes are not deleted, only the mappings to them.

The command that is shown in Example 9-42 deletes the host called `Angola` from the SAN Volume Controller configuration.

Example 9-42 rmhost Angola

```
IBM_2145:ITSO_SVC1:admin>rmhost Angola
```

Deleting a host: If there are any volumes that are assigned to the host, you must use the **`-force`** flag, for example, **`rmhost -force Angola`**.

9.4.5 Adding ports to a defined host

If you add an HBA or a network interface controller (NIC) to a server that is already defined within the SAN Volume Controller, you can use the **`addhostport`** command to add the new port definitions to your host configuration.

If your host is currently connected through SAN with FC and if the WWPN is already zoned to the SAN Volume Controller system, issue the **`lshbaportcandidate`** command, as shown in Example 9-43, to compare with the information that you have from the server administrator.

Example 9-43 lshbaportcandidate

```
IBM_2145:ITSO_SVC1:admin>lshbaportcandidate
id
210000E08B054CAA
```

If the WWPN matches your information (use host or SAN switch utilities to verify), use the **`addhostport`** command to add the port to the host.

Example 9-44 shows the command to add a host port.

Example 9-44 addhostport

```
IBM_2145:ITSO_SVC1:admin>addhostport -hbawwpn 210000E08B054CAA Palau
```

This command adds the WWPN of `210000E08B054CAA` to the `Palau` host.

> **Adding multiple ports:** You can add multiple ports all at one time by using the separator or colon (:) between WWPNs, for example:
>
> `addhostport -hbawwpn 210000E08B054CAA:210000E08B89C1CD Palau`

If the new HBA is not connected or zoned, the **`lshbaportcandidate`** command does not display your WWPN. In this case, you can manually type the WWPN of your HBA or HBAs and use the **`-force`** flag to create the host, as shown in Example 9-45.

Example 9-45 addhostport

```
IBM_2145:ITSO_SVC1:admin>addhostport -hbawwpn 210000E08B054CAA -force Palau
```

This command forces the addition of the WWPN named `210000E08B054CAA` to the host called `Palau`.

> **WWPNs:** WWPNs are *not* case-sensitive within the CLI.

If you run the **`lshost`** command again, you see your host with an updated port count of `2` in Example 9-46.

Example 9-46 lshost command: Port count

```
IBM_2145:ITSO_SVC1:admin>lshost
id              name            port_count     iogrp_count
0               Palau           2              4
1               ITSO_W2008      1              4
2               Thor            3              1
3               Frigg           1              1
4               Baldur          1              1
```

If your host currently uses iSCSI as a connection method, you must have the new iSCSI IQN ID before you add the port. Unlike FC-attached hosts, you cannot check for available candidates with iSCSI.

After you have acquired the additional iSCSI IQN, use the **`addhostport`** command, as shown in Example 9-47.

Example 9-47 Adding an iSCSI port to an already configured host

```
IBM_2145:ITSO_SVC1:admin>addhostport -iscsiname iqn.1991-05.com.microsoft:baldur 4
```

9.4.6 Deleting ports

If you make a mistake when adding a port, or if you remove an HBA from a server that is already defined within the SAN Volume Controller, you can use the **`rmhostport`** command to remove WWPN definitions from an existing host.

Before you remove the WWPN, be sure that it is the correct WWPN by issuing the **lshost** command, as shown in Example 9-48.

Example 9-48 lshost command

```
IBM_2145:ITSO_SVC1:admin>lshost Palau
id 0
name Palau
port_count 2
type generic
mask 1111
iogrp_count 4
WWPN 210000E08B054CAA
node_logged_in_count 2
state active
WWPN 210000E08B89C1CD
node_logged_in_count 2
state offline
```

When you know the WWPN or iSCSI IQN, use the **rmhostport** command to delete a host port, as shown in Example 9-49.

Example 9-49 rmhostport

```
For removing WWPN
IBM_2145:ITSO_SVC1:admin>rmhostport -hbawwpn 210000E08B89C1CD Palau

and for removing iSCSI IQN

IBM_2145:ITSO_SVC1:admin>rmhostport -iscsiname iqn.1991-05.com.microsoft:baldur
Baldur
```

This command removes the WWPN of `210000E08B89C1CD` from the `Palau` host and the iSCSI IQN `iqn.1991-05.com.microsoft:baldur` from the `Baldur` host.

Removing multiple ports: You can remove multiple ports at one time by using the separator or colon (:) between the port names, for example:

```
rmhostport -hbawwpn 210000E08B054CAA:210000E08B892BCD Angola
```

9.5 Working with the Ethernet port for iSCSI

In this section, we describe the commands that are useful for setting, changing, and displaying the SAN Volume Controller Ethernet port for iSCSI configuration.

Example 9-50 shows the **lsportip** command listing the iSCSI IP addresses assigned for each port on each node in the system.

Example 9-50 lsportip command

```
IBM_2145:ITSO_SVC1:admin>lsportip
id               node_id       node_name        IP_address        mask
gateway           IP_address_6      prefix_6       gateway_6         MAC
duplex            state         speed            failover
```

```
1        1       node1
00:1a:64:95:2f:cc Full             unconfigured   1Gb/s            no
1        1       node1
00:1a:64:95:2f:cc Full             unconfigured   1Gb/s            yes
2              1            node1              10.44.36.64      255.255.255.0
10.44.36.254                                          00:1a:64:95:2f:ce Full
online        1Gb/s           no
2        1       node1
00:1a:64:95:2f:ce Full             online         1Gb/s            yes
1        2       node2
00:1a:64:95:3f:4c Full             unconfigured   1Gb/s            no
1        2       node2
00:1a:64:95:3f:4c Full             unconfigured   1Gb/s            yes
2              2            node2              10.44.36.65      255.255.255.0
10.44.36.254                                          00:1a:64:95:3f:4e Full
online        1Gb/s           no
2        2       node2
00:1a:64:95:3f:4e Full             online         1Gb/s            yes
1        3       node3
00:21:5e:41:53:18 Full             unconfigured   1Gb/s            no
1        3       node3
00:21:5e:41:53:18 Full             unconfigured   1Gb/s            yes
2              3            node3              10.44.36.60      255.255.255.0
10.44.36.254                                                   00:21:5e:41:53:1a
Full            online        1Gb/s            no
2        3       node3
00:21:5e:41:53:1a Full             online         1Gb/s            yes
1        4       node4
00:21:5e:41:56:8c Full             unconfigured   1Gb/s            no
1        4       node4
00:21:5e:41:56:8c Full             unconfigured   1Gb/s            yes
2              4            node4              10.44.36.63      255.255.255.0
10.44.36.254                                          00:21:5e:41:56:8e Full
online        1Gb/s           no
2         4        node4
00:21:5e:41:56:8e Full             online         1Gb/s            yes
```

Example 9-51 shows how the **cfgportip** command assigns an IP address to each node Ethernet port for iSCSI I/O.

Example 9-51 cfgportip command

```
IBM_2145:ITSO_SVC1:admin>cfgportip -node 4 -ip 10.44.36.63 -gw 10.44.36.254 -mask 255.255.255.0 2
IBM_2145:ITSO_SVC1:admin>cfgportip -node 1 -ip 10.44.36.64 -gw 10.44.36.254 -mask 255.255.255.0 2
IBM_2145:ITSO_SVC1:admin>cfgportip -node 2 -ip 10.44.36.65 -gw 10.44.36.254 -mask 255.255.255.0 2
```

9.6 Working with volumes

In this section, we describe the various configuration and administrative tasks that can be performed on the volume within the SAN Volume Controller environment.

9.6.1 Creating a volume

The **mkvdisk** command creates sequential, striped, or image mode volume objects. When they are mapped to a host object, these objects are seen as disk drives with which the host can perform I/O operations.

When creating a volume, you must enter several parameters at the CLI. There are both mandatory and optional parameters.

See the full command string and detailed information in *Command-Line Interface User's Guide*, SC27-2287.

Creating an image mode disk: If you do not specify the **-size** parameter when you create an image mode disk, the entire MDisk capacity is used.

When you are ready to create a volume, you must know the following information before you start creating the volume:

- In which storage pool the volume is going to have its extents
- From which I/O Group the volume will be accessed
- Which SAN Volume Controller node will be the preferred node for the volume
- Size of the volume
- Name of the volume
- Type of the volume
- Whether this volume will be managed by Easy Tier to optimize its performance

When you are ready to create your striped volume, use the **mkvdisk** command (we discuss sequential and image mode volumes later). In Example 9-52, this command creates a 10 GB striped volume with volume ID 20 within the storage pool `STGPool_DS3500-2` and assigns it to the `iogrp_0` I/O Group. Its preferred node will be node `1`.

Example 9-52 mkvdisk command

```
IBM_2145:ITSO_SVC1:admin>mkvdisk -mdiskgrp STGPool_DS3500-2 -iogrp io_grp0 -node 1
-size 10 -unit gb -name Tiger
Virtual Disk, id [20], successfully created
```

To verify the results, use the **lsvdisk** command, as shown in Example 9-53.

Example 9-53 lsvdisk command

```
IBM_2145:ITSO_SVC1:admin>lsvdisk 20
id 20
name Tiger
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 1
mdisk_grp_name STGPool_DS3500-2
capacity 10.00GB
type striped
formatted no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076801AF813F1000000000000016
throttling 0
preferred_node_id 1
fast_write_state empty
cache readwrite
udid
```

```
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
filesystem
mirror_write_priority latency

copy_id 0
status online
sync yes
primary yes
mdisk_grp_id 1
mdisk_grp_name STGPool_DS3500-2
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status inactive
tier generic_ssd
tier_capacity 0.00MB
tier generic_hdd
tier_capacity 10.00GB
```

At this point, you have completed the required tasks to create a volume.

9.6.2 Volume information

Use the **lsvdisk** command to display summary information about all volumes that are defined within the SAN Volume Controller environment. To display more detailed information about a specific volume, run the command again and append the volume name parameter or the volume ID.

Example 9-54 shows both of these commands.

Example 9-54 lsvdisk command

```
IBM_2145:ITSO_SVC1:admin>lsvdisk
id name     IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name capacity type
FC_id FC_name RC_id RC_name vdisk_UID                        fc_map_count copy_count
fast_write_state se_copy_count RC_change
0  Volume_A 0        io_grp0     online 0         Pool_DS3500-1  10.00GB  striped
0    GMREL1  6005076801AF813F1000000000000031 0         1        empty         0
0                    no
1  Volume_B 0        io_grp0     online 0         Pool_DS3500-1  10.00GB  striped
1    GMREL2  6005076801AF813F1000000000000032 0         1        empty         0
0                    no
2  Volume_C 0        io_grp0     online 0         Pool_DS3500-1  10.00GB  striped
2    GMREL3  6005076801AF813F1000000000000033 0         1        empty         0
0                    no
IBM_2145:ITSO_SVC1:admin>lsvdisk Volume_A
```

```
id 0
name Volume_A
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name Pool_DS3500-1
capacity 10.00GB
type striped
formatted no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id 0
RC_name GMREL1
vdisk_UID 6005076801AF813F1000000000000031
throttling 0
preferred_node_id 2
fast_write_state empty
cache readwrite
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
filesystem
mirror_write_priority latency

copy_id 0
status online
sync yes
primary yes
mdisk_grp_id 0
mdisk_grp_name Pool_DS3500-1
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status inactive
tier generic_ssd
tier_capacity 0.00MB
tier generic_hdd
tier_capacity 10.00GB
```

9.6.3 Creating a thin-provisioned volume

Example 9-55 shows how to create a thin-provisioned volume. In addition to the normal parameters, you *must* use the following parameters:

-rsize This parameter makes the volume a thin-provisioned volume; otherwise, the volume is fully allocated.

-autoexpand This parameter specifies that thin-provisioned volume copies automatically expand their real capacities by allocating new extents from their storage pool.

-grainsize This parameter sets the grain size (in KB) for a thin-provisioned volume.

Example 9-55 Usage of the command mkvdisk

```
IBM_2145:ITSO_SVC1:admin>mkvdisk -mdiskgrp STGPool_DS3500-2 -iogrp 0 -vtype striped -size
10 -unit gb -rsize 50% -autoexpand -grainsize 32
Virtual Disk, id [21], successfully
created
```

This command creates a space-efficient 10 GB volume. The volume belongs to the storage pool named `STGPool_DS3500-2` and is owned by the `io_grp1` I/O Group. The real capacity automatically expands until the volume size of 10 GB is reached. The grain size is set to 32 K, which is the default.

Disk size: When using the **-rsize** parameter, you have the following options: *disk_size*, *disk_size_percentage*, and **auto**:

- Specify the *disk_size_percentage* value using an integer, or an integer immediately followed by the percent (%) symbol.
- Specify the units for a *disk_size* integer using the **-unit** parameter; the default is MB. The **-rsize** value can be greater than, equal to, or less than the size of the volume.
- The **auto** option creates a volume copy that uses the entire size of the MDisk. If you specify the **-rsize auto** option, you must also specify the **-vtype image** option.

An entry of 1 GB uses 1024 MB.

9.6.4 Creating a volume in image mode

This virtualization type allows an image mode volume to be created when an MDisk already has data on it, perhaps from a pre-virtualized subsystem. When an image mode volume is created, it directly corresponds to the previously unmanaged MDisk from which it was created. Therefore, with the exception of thin-provisioned image mode volume, the volume’s logical block address (LBA) *x* equals MDisk LBA *x*.

You can use this command to bring a non-virtualized disk under the control of the clustered system. After it is under the control of the clustered system, you can migrate the volume from the single managed disk.

As soon as the first MDisk extent has been migrated, the volume is no longer an image mode volume. You can add an image mode volume to an already populated storage pool with other types of volume, such as a striped or sequential volume.

> **Size:** An image mode volume must be at least 512 bytes (the capacity cannot be 0). That is, the minimum size that can be specified for an image mode volume must be the same as the storage pool extent size to which it is added, with a minimum of 16 MB.

You must use the **-mdisk** parameter to specify an MDisk that has a mode of unmanaged. The **-fmtdisk** parameter *cannot* be used to create an image mode volume.

> **Capacity:** If you create a mirrored volume from two image mode MDisks without specifying a **-capacity** value, the capacity of the resulting volume is the smaller of the two MDisks, and the remaining space on the larger MDisk is inaccessible.
>
> If you do not specify the **-size** parameter when you create an image mode disk, the entire MDisk capacity is used.

Use the **mkvdisk** command to create an image mode volume, as shown in Example 9-56.

Example 9-56 mkvdisk (image mode)

```
IBM_2145:ITSO_SVC1:admin>mkvdisk -mdiskgrp STGPool_DS3500-1  -iogrp 0 -mdisk mdisk10 -vtype
image -name Image_Volume_A
Virtual Disk, id [22], successfully
created
```

This command creates an image mode volume called `Image_Volume_A` using the `mdisk10` MDisk. The volume belongs to the storage pool `STGPool_DS3500-1` and is owned by the `io_grp0` I/O Group.

If we run the **lsvdisk** command again, notice that the volume named `Image_Volume_A` has a status of `image`, as shown in Example 9-57.

Example 9-57 lsvdisk

```
IBM_2145:ITSO_SVC1:admin>lsvdisk -filtervalue type=image
id name           IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name   capacity
type  FC_id FC_name RC_id RC_name vdisk_UID                        fc_map_count copy_count
fast_write_state se_copy_count RC_change
22 Image_Volume_A 0           io_grp0       online 0            STGPool_DS3500-1 10.00GB
image                             6005076801AF813F1000000000000018 0            1
empty            0  no
```

9.6.5 Adding a mirrored volume copy

You can create a mirrored copy of a volume, which keeps a volume accessible even when the MDisk on which it depends has become unavailable. You can create a copy of a volume either on separate storage pools or by creating an image mode copy of the volume. Copies increase the availability of data; however, they are not separate objects. You can only create or change mirrored copies from the volume.

In addition, you can use volume mirroring as an alternative method of migrating volumes between storage pools.

For example, if you have a non-mirrored volume in one storage pool and want to migrate that volume to another storage pool, you can add a new copy of the volume and specify the second storage pool. After the copies are synchronized, you can delete the copy on the first storage pool. The volume is copied to the second storage pool while remaining online during the copy.

To create a mirrored copy of a volume, use the **addvdiskcopy** command. This command adds a copy of the chosen volume to the selected storage pool, which changes a non-mirrored volume into a mirrored volume.

In the following scenario, we show creating a mirrored volume from one storage pool to another storage pool.

As you can see in Example 9-58, the volume has a copy with copy_id 0.

Example 9-58 lsvdisk

```
IBM_2145:ITSO_SVC1:admin>lsvdisk Volume_no_mirror
id 23
name Volume_no_mirror
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name STGPool_DS3500-1
capacity 1.00GB
type striped
formatted no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076801AF813F1000000000000019
throttling 0
preferred_node_id 1
fast_write_state empty
cache readwrite
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
filesystem
mirror_write_priority latency

copy_id 0
status online
sync yes
primary yes
mdisk_grp_id 0
mdisk_grp_name STGPool_DS3500-1
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 1.00GB
real_capacity 1.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
```

```
grainsize
se_copy no
easy_tier on
easy_tier_status inactive
tier generic_ssd
tier_capacity 0.00MB
tier generic_hdd
tier_capacity
1.00GB
```

In Example 9-59, we add the volume copy mirror by using the **addvdiskcopy** command.

Example 9-59 addvdiskcopy

```
IBM_2145:ITSO_SVC1:admin>addvdiskcopy -mdiskgrp STGPool_DS5000-1 -vtype striped -unit gb
Volume_no_mirror
Vdisk [23] copy [1] successfully created
```

During the synchronization process, you can see the status by using the **lsvdisksyncprogress** command. As shown in Example 9-60, the first time that the status is checked, the synchronization progress is at 48%, and the estimated completion time is 11:09:26. The second time that the command is run, the progress status is at 100%, and the synchronization is complete.

Example 9-60 Synchronization

```
IBM_2145:ITSO_SVC1:admin>lsvdisksyncprogress
vdisk_id vdisk_name       copy_id progress estimated_completion_time
23       Volume_no_mirror 1       48       110926203918
IBM_2145:ITSO_SVC1:admin>lsvdisksyncprogress
vdisk_id vdisk_name       copy_id progress estimated_completion_time
23       Volume_no_mirror 1       100
```

As you can see in Example 9-61, the new mirrored volume copy (copy_id 1) has been added and can be seen by using the **lsvdisk** command.

Example 9-61 lsvdisk

```
IBM_2145:ITSO_SVC1:admin>lsvdisk 23
id 23
name Volume_no_mirror
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id many
mdisk_grp_name many
capacity 1.00GB
type many
formatted no
mdisk_id many
mdisk_name many
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076801AF813F1000000000000019
throttling 0
preferred_node_id 1
```

```
fast_write_state empty
cache readwrite
udid
fc_map_count 0
sync_rate 50
copy_count 2
se_copy_count 0
filesystem
mirror_write_priority latency

copy_id 0
status online
sync yes
primary yes
mdisk_grp_id 0
mdisk_grp_name STGPool_DS3500-1
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 1.00GB
real_capacity 1.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status inactive
tier generic_ssd
tier_capacity 0.00MB
tier generic_hdd
tier_capacity 1.00GB

copy_id 1
status online
sync yes
primary no
mdisk_grp_id 2
mdisk_grp_name STGPool_DS5000-1
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 1.00GB
real_capacity 1.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status inactive
tier generic_ssd
tier_capacity 0.00MB
tier generic_hdd
```

```
tier_capacity
1.00GB
```

While adding a volume copy mirror, you can define a mirror with different parameters to the volume copy. Therefore, you can define a thin-provisioned volume copy for a non-volume copy volume and vice versa, which is one way to migrate a non-thin-provisioned volume to a thin-provisioned volume.

Volume copy mirror parameters: To change the parameters of a volume copy mirror, you must delete the volume copy and redefine it with the new values.

Now, we can change the name of the volume just mirrored from `Volume_no_mirror to Volume_mirrored`, as shown in Example 9-62.

Example 9-62 Volume name changes

```
IBM_2145:ITSO_SVC1:admin>chvdisk -name Volume_mirrored Volume_no_mirror
```

9.6.6 Splitting a mirrored volume

The **splitvdiskcopy** command creates a new volume in the specified I/O Group from a copy of the specified volume. If the copy that you are splitting is not synchronized, you must use the **-force** parameter. The command fails if you are attempting to remove the only synchronized copy. To avoid this failure, wait for the copy to synchronize, or split the unsynchronized copy from the volume by using the **-force** parameter. You can run the command when either volume copy is offline.

Example 9-63 shows the **splitvdiskcopy** command, which is used to split a mirrored volume. It creates a new volume, `Volume_new` from `Volume_mirrored`.

Example 9-63 Split volume

```
IBM_2145:ITSO_SVC1:admin>splitvdiskcopy -copy 1 -iogrp 0 -name Volume_new Volume_mirrored
Virtual Disk, id [24], successfully created
```

As you can see in Example 9-64, the new volume named `Volume_new` has been created as an independent volume.

Example 9-64 lsvdisk

```
IBM_2145:ITSO_SVC1:admin>lsvdisk Volume_new
id 24
name Volume_new
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 2
mdisk_grp_name STGPool_DS5000-1
capacity 1.00GB
type striped
formatted no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
```

```
RC_name
vdisk_UID 6005076801AF813F100000000000001A
throttling 0
preferred_node_id 2
fast_write_state empty
cache readwrite
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
filesystem
mirror_write_priority latency

copy_id 0
status online
sync yes
primary yes
mdisk_grp_id 2
mdisk_grp_name STGPool_DS5000-1
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 1.00GB
real_capacity 1.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
easy_tier_status inactive
tier generic_ssd
tier_capacity 0.00MB
tier generic_hdd
tier_capacity
1.00GB
```

By issuing the command in Example 9-63 on page 200, `Volume_mirrored` will no longer have its mirrored copy and a new volume will be created automatically.

9.6.7 Modifying a volume

Executing the **`chvdisk`** command will modify a *single* property of a volume. Only *one* property can be modified at a time. So, changing the name and modifying the I/O Group require two invocations of the command.

You can specify a new name or label. The new name can be used subsequently to reference the volume. The I/O Group with which this volume is associated can be changed. Note that changing the I/O Group with which this volume is associated requires a flush of the cache within the nodes in the current I/O Group to ensure that all data is written to disk. I/O must be suspended at the host level before performing this operation.

Tips: If the volume has a mapping to any hosts, it is not possible to move the volume to an I/O Group that does not include any of those hosts.

This operation will fail if there is not enough space to allocate bitmaps for a mirrored volume in the target I/O Group.

If the **-force** parameter is used and the system is unable to destage all write data from the cache, the contents of the volume are corrupted by the loss of the cached data.

If the **-force** parameter is used to move a volume that has out-of-sync copies, a full resynchronization is required.

9.6.8 I/O governing

You can set a limit on the number of I/O operations accepted for a volume. The limit is set in terms of I/Os per second or MB per second. By default, no I/O governing rate is set when a volume is created.

Base the choice between I/O and MB as the I/O governing throttle on the disk access profile of the application. Database applications generally issue large amounts of I/O, but they only transfer a relatively small amount of data. In this case, setting an I/O governing throttle based on MB per second does not achieve much. It is better to use an I/Os per second as a second throttle.

At the other extreme, a streaming video application generally issues a small amount of I/O, but it transfers large amounts of data. In contrast to the database example, setting an I/O governing throttle based on I/Os per second does not achieve much, so it is better to use an MB per second throttle.

I/O governing rate: An I/O governing rate of 0 (displayed as *throttling* in the CLI output of the **lsvdisk** command) does not mean that zero I/Os per second (or MB per second) can be achieved. It means that no throttle is set.

An example of the **chvdisk** command is shown in Example 9-65.

Example 9-65 chvdisk

```
IBM_2145:ITSO_SVC1:admin>chvdisk -rate 20 -unitmb volume_7
IBM_2145:ITSO_SVC1:admin>chvdisk -warning 85% volume_7
```

New name first: The **chvdisk** command specifies the new name first. The name can consist of letters A to Z and a to z, numbers 0 - 9, the dash (-), and the underscore (_). It can be between 1 and 63 characters in length. However, it cannot start with a number, the dash, or the word “vdisk” (because this prefix is reserved for SAN Volume Controller assignment only).

The first command changes the volume throttling of volume_7 to 20 MBps. The second command changes the thin-provisioned volume warning to 85%. To verify the changes, issue the **lsvdisk** command, as shown in Example 9-66 on page 203.

Example 9-66 lsvdisk command: Verifying throttling

```
IBM_2145:ITSO_SVC1:admin>lsvdisk volume_7
id 1
name volume_7
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name STGPool_DS3500-1
capacity 10.00GB
type striped
formatted no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076801AF813F100000000000001F
virtual_disk_throttling (MB) 20
preferred_node_id 2
fast_write_state empty
cache readwrite
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 1
filesystem
mirror_write_priority latency

copy_id 0
status online
sync yes
primary yes
mdisk_grp_id 0
mdisk_grp_name STGPool_DS3500-1
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 0.41MB
real_capacity 2.02GB
free_capacity 2.02GB
overallocation 496
autoexpand on
warning 85
grainsize 32
se_copy yes
easy_tier on
easy_tier_status inactive
tier generic_ssd
tier_capacity 0.00MB
tier generic_hdd
tier_capacity
2.02GB
```

9.6.9 Deleting a volume

When executing this command on an existing *fully managed mode* volume, any data that remained on it will be lost. The extents that made up this volume will be returned to the pool of free extents available in the storage pool.

If any remote copy, FlashCopy, or host mappings still exist for this volume, the delete fails unless the **-force** flag is specified. This flag ensures the deletion of the volume and any volume to host mappings and copy mappings.

If the volume is currently the subject of a migrate to image mode, the delete fails unless the **-force** flag is specified. This flag halts the migration and then deletes the volume.

If the command succeeds (without the **-force** flag) for an image mode volume, the underlying back-end controller logical unit will be consistent with the data that a host might previously have read from the image mode volume. That is, all fast write data has been flushed to the underlying LUN. If the **-force** flag is used, there is no guarantee.

If there is any non-destaged data in the fast write cache for this volume, the deletion of the volume fails unless the **-force** flag is specified. Now, any non-destaged data in the fast write cache is deleted.

Use the **rmvdisk** command to delete a volume from your SAN Volume Controller configuration, as shown in Example 9-67.

Example 9-67 rmvdisk

```
IBM_2145:ITSO_SVC1:admin>rmvdisk volume_A
```

This command deletes the `volume_A` volume from the SAN Volume Controller configuration. If the volume is assigned to a host, you need to use the **-force** flag to delete the volume (Example 9-68).

Example 9-68 rmvdisk (-force)

```
IBM_2145:ITSO_SVC1:admin>rmvdisk -force volume_A
```

9.6.10 Expanding a volume

Expanding a volume presents a larger capacity disk to your operating system. Although this expansion can be easily performed using the SAN Volume Controller, you must ensure that your operating systems support expansion before using this function.

Assuming that your operating systems support it, you can use the **expandvdisksize** command to increase the capacity of a given volume.

Example 9-69 shows a sample of this command.

Example 9-69 expandvdisksize

```
IBM_2145:ITSO_SVC1:admin>expandvdisksize -size 5 -unit gb volume_C
```

This command expands the `volume_C` volume, which was `35` GB before, by another 5 GB to give it a total size of `40` GB.

To expand a thin-provisioned volume, you can use the **-rsize** option, as shown in Example 9-70. This command changes the real size of the volume_B volume to a real capacity of 55 GB. The capacity of the volume remains unchanged.

Example 9-70 lsvdisk

```
IBM_2145:ITSO_SVC1:admin>lsvdisk volume_B
id 26
capacity 100.00GB
type striped
.
.
copy_id 0
status online
used_capacity 0.41MB
real_capacity 50.02GB
free_capacity 50.02GB
overallocation 199
autoexpand on
warning 80
grainsize 32
se_copy yes
IBM_2145:ITSO_SVC1:admin>expandvdisksize -rsize 5 -unit gb volume_B
IBM_2145:ITSO_SVC1:admin>lsvdisk volume_B
id 26
name volume_B
capacity 100.00GB
type striped
.
.
copy_id 0
status online
used_capacity 0.41MB
real_capacity 55.02GB
free_capacity 55.02GB
overallocation 181
autoexpand on
warning 80
grainsize 32
se_copy yes
```

Important: If a volume is expanded, its type will become striped even if it was previously sequential or in image mode. If there are not enough extents to expand your volume to the specified size, you receive the following error message:

`CMMVC5860E Ic_failed_vg_insufficient_virtual_extents`

9.6.11 Assigning a volume to a host

Use the **mkvdiskhostmap** command to map a volume to a host. When executed, this command creates a new mapping between the volume and the specified host, which essentially presents this volume to the host as though the disk was directly attached to the host. It is only after this command is executed that the host can perform I/O to the volume. Optionally, a SCSI LUN ID can be assigned to the mapping.

When the HBA on the host scans for devices that are attached to it, it discovers all of the volumes that are mapped to its FC ports. When the devices are found, each one is allocated an identifier (SCSI LUN ID).

For example, the first disk found is generally SCSI LUN 1, and so on. You can control the order in which the HBA discovers volumes by assigning the SCSI LUN ID as required. If you do not specify a SCSI LUN ID, the system automatically assigns the next available SCSI LUN ID, given any mappings that already exist with that host.

Using the volume and host definition that we created in the previous sections, we assign volumes to hosts that are ready for their use. We use the **mkvdiskhostmap** command (see Example 9-71).

Example 9-71 mkvdiskhostmap

```
IBM_2145:ITSO_SVC1:admin>mkvdiskhostmap -host Almaden  volume_B
Virtual Disk to Host map, id [0], successfully created
IBM_2145:ITSO_SVC1:admin>mkvdiskhostmap -host Almaden volume_C
Virtual Disk to Host map, id [1], successfully
created
```

This command displays `volume_B` and `volume_C` that are assigned to host `Almaden`, as shown in Example 9-72.

Example 9-72 lshostvdiskmap -delim command

```
IBM_2145:ITSO_SVC1:admin>lshostvdiskmap -delim :
id:name:SCSI_id:vdisk_id:vdisk_name:vdisk_UID
2:Almaden:0:26:volume_B:6005076801AF813F1000000000000020
2:Almaden:1:27:volume_C:6005076801AF813F1000000000000021
```

Assigning a specific LUN ID to a volume: The optional **-scsi scsi_num** parameter can help assign a specific LUN ID to a volume that is to be associated with a given host. The default (if nothing is specified) is to increment based on what is already assigned to the host.

Be aware that certain HBA device drivers stop when they find a gap in the SCSI LUN IDs, for example:

- Volume 1 is mapped to Host 1 with SCSI LUN ID 1.
- Volume 2 is mapped to Host 1 with SCSI LUN ID 2.
- Volume 3 is mapped to Host 1 with SCSI LUN ID 4.

When the device driver scans the HBA, it might stop after discovering Volumes 1 and 2 because there is no SCSI LUN mapped with ID 3.

Important: Ensure that the SCSI LUN ID allocation is contiguous.

It is not possible to map a volume to a host more than one time at separate LUNs (Example 9-73).

Example 9-73 mkvdiskhostmap

```
IBM_2145:ITSO_SVC1:admin>mkvdiskhostmap -host Siam volume_A
Virtual Disk to Host map, id [0], successfully created
```

This command maps the volume called `volume_A` to the host called `Siam`.

At this point, you have completed all tasks that are required to assign a volume to an attached host.

9.6.12 Showing volumes to host mapping

Use the **lshostvdiskmap** command to show which volumes are assigned to a specific host (Example 9-74).

Example 9-74 lshostvdiskmap

```
IBM_2145:ITSO_SVC1:admin>lshostvdiskmap -delim , Siam
id,name,SCSI_id,vdisk_id,vdisk_name,wwpn,vdisk_UID
3,Siam,0,0,volume_A,210000E08B18FF8A,60050768018301BF280000000000000C
```

From this command, you can see that the host `Siam` has only one assigned volume called `volume_A`. The SCSI LUN ID is also shown, which is the ID by which the volume is presented to the host. If no host is specified, all defined host to volume mappings will be returned.

Specifying the flag before the host name: Although the **-delim** flag normally comes at the end of the command string, in this case, you must specify this flag *before* the host name. Otherwise, it returns the following message:

```
CMMVC6070E An invalid or duplicated parameter, unaccompanied argument, or
incorrect argument sequence has been detected. Ensure that the input is as per
the help.
```

9.6.13 Deleting a volume to host mapping

When deleting a volume mapping, you are not deleting the volume itself, only the connection from the host to the volume. If you mapped a volume to a host by mistake, or you simply want to reassign the volume to another host, use the **rmvdiskhostmap** command to unmap a volume from a host (Example 9-75).

Example 9-75 rmvdiskhostmap

```
IBM_2145:ITSO_SVC1:admin>rmvdiskhostmap -host Tiger volume_D
```

This command unmaps the volume called `volume_D` from the host called `Tiger`.

9.6.14 Migrating a volume

From time to time, you might want to migrate volumes from one set of MDisks to another set of MDisks to decommission an old disk subsystem, to have better balanced performance across your virtualized environment, or simply to migrate data into the SAN Volume Controller environment transparently using image mode.

Important: After migration is started, it continues until completion unless it is stopped or suspended by an error condition or unless the volume being migrated is deleted.

As you can see from the parameters that are shown in Example 9-76 on page 208, before you can migrate your volume, you must know the name of the volume that you want to migrate and the name of the storage pool to which you want to migrate. To discover the names, run the **lsvdisk** and **lsmdiskgrp** commands.

After you know these details, you can issue the **migratevdisk** command, as shown in Example 9-76.

Example 9-76 migratevdisk

```
IBM_2145:ITSO_SVC1:admin>migratevdisk -mdiskgrp STGPool_DS5000-1 -vdisk volume_C
```

This command moves `volume_C` to the storage pool named `STGPool_DS5000-1`.

> **Tips:** If insufficient extents are available within your target storage pool, you receive an error message. Make sure that the source MDisk group and target MDisk group have the same extent size.
>
> The optional threads parameter allows you to assign a priority to the migration process. The default is `4`, which is the highest priority setting. However, if you want the process to take a lower priority over other types of I/O, you can specify `3`, `2`, or `1`.

You can run the **lsmigrate** command at any time to see the status of the migration process (Example 9-77).

Example 9-77 lsmigrate command

```
IBM_2145:ITSO_SVC1:admin>lsmigrate
migrate_type MDisk_Group_Migration
progress 0
migrate_source_vdisk_index 27
migrate_target_mdisk_grp 2
max_thread_count 4
migrate_source_vdisk_copy_id 0

IBM_2145:ITSO_SVC1:admin>lsmigrate
migrate_type MDisk_Group_Migration
progress 76
migrate_source_vdisk_index 27
migrate_target_mdisk_grp 2
max_thread_count 4
migrate_source_vdisk_copy_id
0
```

> **Progress:** The progress is given as percent complete. If you receive no more replies, it means that the process has finished.

9.6.15 Migrating a fully managed volume to an image mode volume

Migrating a fully managed volume to an image mode volume allows the SAN Volume Controller to be removed from the data path, which might be useful where the SAN Volume Controller is used as a data mover appliance. You can use the **migratetoimage** command.

To migrate a fully managed volume to an image mode volume, the following rules apply:

- The destination MDisk must be greater than or equal to the size of the volume.
- The MDisk that is specified as the target must be in an unmanaged state.
- Regardless of the mode in which the volume starts, it is reported as a managed mode during the migration.

- Both of the MDisks that are involved are reported as being image mode volumes during the migration.
- If the migration is interrupted by a system recovery or a cache problem, the migration resumes after the recovery completes.

Example 9-78 shows an example of the command.

Example 9-78 migratetoimage

```
IBM_2145:ITSO_SVC1:admin>migratetoimage -vdisk volume_A -mdisk mdisk10 -mdiskgrp
STGPool_IMAGE
```

In this example, you migrate the data from `volume_A` onto `mdisk10`, and the MDisk must be put into the `STGPool_IMAGE` storage pool.

9.6.16 Shrinking a volume

The **shrinkvdisksize** command reduces the capacity that is allocated to the particular volume by the amount that you specify. You cannot shrink the real size of a thin-provisioned volume to less than its used size. All capacities, including changes, must be in multiples of 512 bytes. An entire extent is reserved even if it is only partially used. The default capacity units are MBs.

You can use this command to shrink the physical capacity that is allocated to a particular volume by the specified amount. You also can use this command to shrink the virtual capacity of a thin-provisioned volume without altering the physical capacity that is assigned to the volume:

- For a non-thin-provisioned volume, use the **-size** parameter.
- For a thin-provisioned volume's *real* capacity, use the **-rsize** parameter.
- For the thin-provisioned volume's *virtual* capacity, use the **-size** parameter.

When the virtual capacity of a thin-provisioned volume is changed, the warning threshold is automatically scaled to match. The new threshold is stored as a percentage.

The system arbitrarily reduces the capacity of the volume by removing a partial extent, one extent, or multiple extents from those extents that are allocated to the volume. You cannot control which extents are removed, and therefore you *cannot* assume that it is unused space that is removed.

Image mode volumes cannot be reduced in size. Instead, they must first be migrated to fully Managed Mode. To run the **shrinkvdisksize** command on a mirrored volume, all copies of the volume must be synchronized.

Important:

- If the volume contains data, do *not* shrink the disk.
- Certain operating systems or file systems use the outer edge of the disk for performance reasons. This command can shrink a FlashCopy target volume to the same capacity as the source.
- Before you shrink a volume, validate that the volume is not mapped to any host objects. If the volume is mapped, data is displayed. You can determine the exact capacity of the source or master volume by issuing the **svcinfo lsvdisk -bytes vdiskname** command. Shrink the volume by the required amount by issuing the **shrinkvdisksize -size disk_size -unit b | kb | mb | gb | tb | pb vdisk_name | vdisk_id** command.

Assuming that your operating system supports it, you can use the **shrinkvdisksize** command to decrease the capacity of a given volume.

Example 9-79 shows an example of this command.

Example 9-79 shrinkvdisksize

```
IBM_2145:ITSO_SVC1:admin>shrinkvdisksize -size 44 -unit gb volume_D
```

This command shrinks a volume called `volume_D` from a total size of 80 GB, by `44` GB, to a new total size of 36 GB.

9.6.17 Showing a volume on an MDisk

Use the **lsmdiskmember** command to display information about the volume that is using space on a specific MDisk, as shown in Example 9-80.

Example 9-80 lsmdiskmember command

```
IBM_2145:ITSO_SVC1:admin>lsmdiskmember mdisk8
id copy_id
24 0
27 0
```

This command displays a list of all of the volume IDs that correspond to the volume copies that use `mdisk8`.

To correlate the IDs that are displayed in this output to volume names, we can run the **lsvdisk** command, which we discuss in more detail in 9.6, "Working with volumes" on page 191.

9.6.18 Showing which volumes are using a storage pool

Use the **lsvdisk -filtervalue** command, as shown in Example 9-81, to see which volumes are part of a specific storage pool. This command shows all of the volumes that are part of the storage pool named `STGPool_DS3500_2`.

Example 9-81 lsvdisk -filtervalue: VDisks in the MDG

```
IBM_2145:ITSO_SVC1:admin>lsvdisk -filtervalue mdisk_grp_name=STGPool_DS3500-2 -delim ,
id,name,IO_group_id,IO_group_name,status,mdisk_grp_id,mdisk_grp_name,capacity,type,FC_id,FC
_name,RC_id,RC_name,vdisk_UID,fc_map_count,copy_count,fast_write_state,se_copy_count,RC_cha
nge
7,W2K3_SRV2_VOL01,0,io_grp0,online,1,STGPool_DS3500-2,10.00GB,striped,,,,,6005076801AF813F1
000000000000008,0,1,empty,0,0,no
8,W2K3_SRV2_VOL02,0,io_grp0,online,1,STGPool_DS3500-2,10.00GB,striped,,,,,6005076801AF813F1
000000000000009,0,1,empty,0,0,no
9,W2K3_SRV2_VOL03,0,io_grp0,online,1,STGPool_DS3500-2,10.00GB,striped,,,,,6005076801AF813F1
00000000000000A,0,1,empty,0,0,no
10,W2K3_SRV2_VOL04,0,io_grp0,online,1,STGPool_DS3500-2,10.00GB,striped,,,,,6005076801AF813F
100000000000000B,0,1,empty,0,0,no
11,W2K3_SRV2_VOL05,0,io_grp0,online,1,STGPool_DS3500-2,10.00GB,striped,,,,,6005076801AF813F
100000000000000C,0,1,empty,0,0,no
12,W2K3_SRV2_VOL06,0,io_grp0,online,1,STGPool_DS3500-2,10.00GB,striped,,,,,6005076801AF813F
100000000000000D,0,1,empty,0,0,no
16,AIX_SRV2_VOL01,0,io_grp0,online,1,STGPool_DS3500-2,20.00GB,striped,,,,,6005076801AF813F1
000000000000011,0,1,empty,0,0,no
```

9.6.19 Showing which MDisks are used by a specific volume

Use the **lsvdiskmember** command, as shown in Example 9-82, to show from which MDisks a specific volume's extents came.

Example 9-82 lsvdiskmember command

```
IBM_2145:ITSO_SVC1:admin>lsvdiskmember 0
id
4
5
6
7
```

If you want to know more about these MDisks, you can run the **lsmdisk** command, as explained in 9.2, "New commands" on page 168 (using the ID that is displayed in Example 9-82 rather than the name).

9.6.20 Showing from which storage pool a volume has its extents

Use the **lsvdisk** command, as shown in Example 9-83, to show to which storage pool a specific volume belongs.

Example 9-83 lsvdisk command: Storage pool name

```
IBM_2145:ITSO_SVC1:admin>lsvdisk Volume_D
id 25
name Volume_D
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 0
mdisk_grp_name STGPool_DS3500-1
capacity 10.00GB
type striped
formatted no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 6005076801AF813F100000000000001E
throttling 0
preferred_node_id 1
fast_write_state empty
cache readwrite
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 1
filesystem
mirror_write_priority latency

copy_id 0
status online
sync yes
primary yes
```

```
mdisk_grp_id 0
mdisk_grp_name STGPool_DS3500-1
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 0.41MB
real_capacity 2.02GB
free_capacity 2.02GB
overallocation 496
autoexpand on
warning 80
grainsize 32
se_copy yes
easy_tier on
easy_tier_status inactive
tier generic_ssd
tier_capacity 0.00MB
tier generic_hdd
tier_capacity 2.02GB
```

To learn more about these storage pools, you can run the **lsmdiskgrp** command, as explained in 9.3.10, “Working with a storage pool” on page 180.

9.6.21 Showing the host to which the volume is mapped

To show the hosts to which a specific volume has been assigned, run the **lsvdiskhostmap** command, as shown in Example 9-84.

Example 9-84 lsvdiskhostmap command

```
IBM_2145:ITSO_SVC1:admin>lsvdiskhostmap -delim , volume_B
id,name,SCSI_id,host_id,host_name,vdisk_UID
26,volume_B,0,2,Almaden,6005076801AF813F1000000000000020
```

This command shows the host or hosts to which the `volume_B` volume was mapped. It is normal for you to see duplicate entries because there are more paths between the clustered system and the host. To be sure that the operating system on the host sees the disk only one time, you must install and configure a multipath software application, such as the IBM Subsystem Driver (SDD).

Specifying the -delim flag: Although the optional **-delim** flag normally comes at the end of the command string, in this case, you must specify this flag *before* the volume name. Otherwise, the command does not return any data.

9.6.22 Showing the volume to which the host is mapped

To show the volume to which a specific host has been assigned, run the **lshostvdiskmap** command, as shown in Example 9-85 on page 213.

Example 9-85 lshostvdiskmap command example

```
IBM_2145:ITSO_SVC1:admin>lshostvdiskmap -delim , Almaden
id,name,SCSI_id,vdisk_id,vdisk_name,vdisk_UID
2,Almaden,0,26,volume_B,60050768018301BF2800000000000005
2,Almaden,1,27,volume_A,60050768018301BF2800000000000004
```

This command shows which volumes are mapped to the host called `Almaden`.

> **Specifying the -delim flag:** Although the optional **`-delim`** flag normally comes at the end of the command string, in this case, you must specify this flag *before* the volume name. Otherwise, the command does not return any data.

9.6.23 Tracing a volume from a host back to its physical disk

In many cases, you must verify exactly which physical disk is presented to the host; for example, from which storage pool a specific volume comes. However, from the host side, it is not possible for the server administrator using the GUI to see on which physical disks the volumes are running.

Instead, you must enter the command (listed in Example 9-86) from your multipath command prompt. Follow these steps:

1. On your host, run the **`datapath query device`** command. You see a long disk serial number for each vpath device, as shown in Example 9-86.

Example 9-86 datapath query device

```
DEV#:   0  DEVICE NAME: Disk1 Part0  TYPE: 2145       POLICY: OPTIMIZED
SERIAL: 60050768018301BF2800000000000005
============================================================================
Path#            Adapter/Hard Disk        State  Mode       Select     Errors
    0     Scsi Port2 Bus0/Disk1 Part0     OPEN   NORMAL         20          0
    1     Scsi Port3 Bus0/Disk1 Part0     OPEN   NORMAL       2343          0

DEV#:   1  DEVICE NAME: Disk2 Part0  TYPE: 2145       POLICY: OPTIMIZED
SERIAL: 60050768018301BF2800000000000004
============================================================================
Path#            Adapter/Hard Disk        State  Mode       Select     Errors
    0     Scsi Port2 Bus0/Disk2 Part0     OPEN   NORMAL       2335          0
    1     Scsi Port3 Bus0/Disk2 Part0     OPEN   NORMAL          0          0

DEV#:   2  DEVICE NAME: Disk3 Part0  TYPE: 2145       POLICY: OPTIMIZED
SERIAL: 60050768018301BF2800000000000006
============================================================================
Path#            Adapter/Hard Disk        State  Mode       Select     Errors
    0     Scsi Port2 Bus0/Disk3 Part0     OPEN   NORMAL       2331          0
    1     Scsi Port3 Bus0/Disk3 Part0     OPEN   NORMAL          0          0
```

> **State:** In Example 9-86, the state of each path is `OPEN`. Sometimes, you will see the state `CLOSED`. This state does not necessarily indicate a problem because it might be a result of the path's processing stage.

2. Run the **lshostvdiskmap** command to return a list of all assigned volumes (Example 9-87).

Example 9-87 lshostvdiskmap

```
IBM_2145:ITSO_SVC1:admin>lshostvdiskmap -delim , Almaden
id,name,SCSI_id,vdisk_id,vdisk_name,vdisk_UID
2,Almaden,0,26,volume_B,60050768018301BF2800000000000005
2,Almaden,1,27,volume_A,60050768018301BF2800000000000004
2,Almaden,2,28,volume_C,60050768018301BF2800000000000006
```

Look for the disk serial number that matches your **datapath query device** output. This host was defined in our SAN Volume Controller as `Almaden`.

3. Run the **lsvdiskmember vdiskname** command for a list of the MDisk or MDisks that make up the specified volume (Example 9-88).

Example 9-88 lsvdiskmember

```
IBM_2145:ITSO_SVC1:admin>lsvdiskmember volume_E
id
0
1
2
3
4
10
11
13
15
16
17
```

4. Query the MDisks with the **lsmdisk mdiskID** to find their controller and LUN number information, as shown in Example 9-89. The output displays the controller name and the controller LUN ID to help you to track back to a LUN within the disk subsystem (if you gave your controller a unique name, such as a serial number). See Example 9-89.

Example 9-89 lsmdisk command

```
IBM_2145:ITSO_SVC1:admin>lsmdisk 0
id 0
name mdisk0
status online
mode managed
mdisk_grp_id 0
mdisk_grp_name STGPool_DS3500-1
capacity 128.0GB
quorum_index 1
block_size 512
controller_name ITSO-DS3500
ctrl_type 4
ctrl_WWNN 20080080E51B09E8
controller_id 2
path_count 4
max_path_count 4
ctrl_LUN_# 0000000000000000
UID 60080e50001b0b62000007b04e731e4d00000000000000000000000000000000
preferred_WWPN 20580080E51B09E8
active_WWPN 20580080E51B09E8
fast_write_state empty
raid_status
raid_level
```

```
redundancy
strip_size
spare_goal
spare_protection_min
balanced
tier generic_hdd
```

9.7 Scripting under the CLI for SAN Volume Controller task automation

> **Command prefix changes:** The **svctask** and **svcinfo** command prefixes are no longer necessary when issuing a command. If you have existing scripts that use those prefixes, they will continue to function. You do not need to change the scripts.

Using scripting constructs works better for the automation of regular operational jobs. You can use available shells to develop scripts. Scripting enhances the productivity of SAN Volume Controller administrators and the integration of their storage virtualization environment. You can create your own customized scripts to automate many tasks for completion at various times and run them through the CLI.

We suggest that you keep the scripting as simple as possible in large SAN environments where scripting commands are used. It is harder to manage fallback, documentation, and the verification of a successful script before execution in a large SAN environment.

In this section, we present an overview of how to automate various tasks by creating scripts using the SAN Volume Controller CLI.

9.7.1 Scripting structure

When creating scripts to automate the tasks on the SAN Volume Controller, use the structure that is illustrated in Figure 9-2 on page 216.

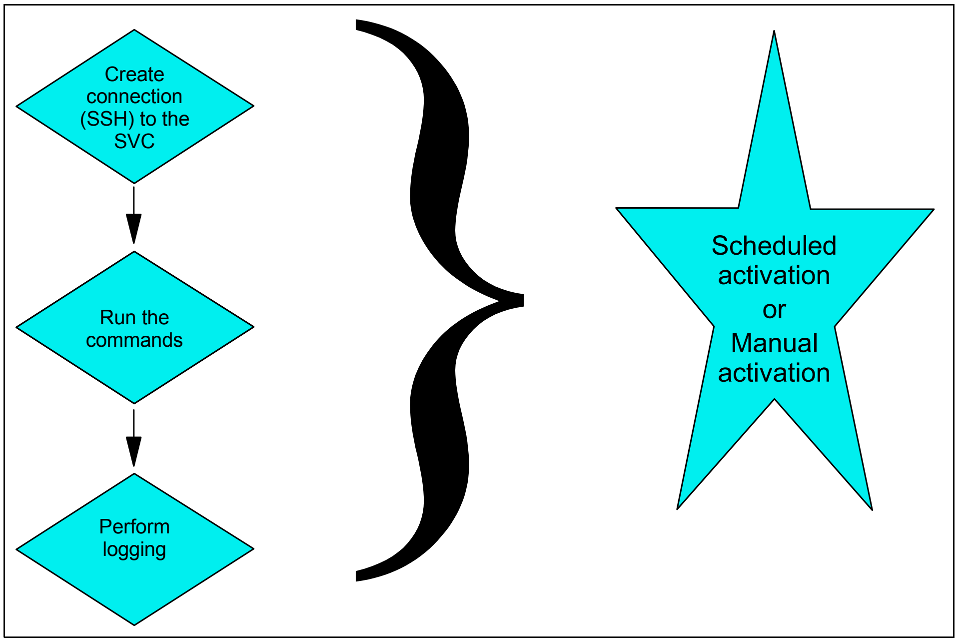


Figure 9-2 Scripting structure for SAN Volume Controller task automation

Creating a Secure Shell connection to the SAN Volume Controller

Secure Shell Key: Starting with SAN Volume Controller 6.3, using a Secure Shell (SSH) Key is optional. You can use a user ID and password to access the system. However, for security reasons, we suggest the use of SSH Key. We provide a sample of its use.

When creating a connection to the SAN Volume Controller, if you are running the script, you must have access to a public key that corresponds to a public key that has been previously uploaded to the SAN Volume Controller.

The key is used to establish the SSH connection that is needed to use the CLI on the SAN Volume Controller. If the SSH keypair is generated without a passphrase, you can connect without the need of special scripting to parse in the passphrase.

On UNIX systems, you can use the **`ssh`** command to create an SSH connection with the SAN Volume Controller. On Windows systems, you can use a utility called **`plink.exe`**, which is provided with the PuTTY tool, to create an SSH connection with the SAN Volume Controller. In the following examples, we use **`plink`** to create the SSH connection to the SAN Volume Controller.

Executing the commands

When using the CLI, see the *IBM System Storage SAN Volume Controller Command-Line Interface User's Guide,* GC27-2287 to obtain the correct syntax and a detailed explanation of each command. You can download this guide for each SAN Volume Controller code level from the SAN Volume Controller documentation page at this website:

`http://ibm.co/1wHFfIw`

When using the CLI, not all commands provide a response to determine the status of the invoked command. Therefore, always create checks that can be logged for monitoring and troubleshooting purposes.

Connecting to the SVC using a predefined SSH connection

The easiest way to create an SSH connection to the SAN Volume Controller is when **`plink`** can call a predefined PuTTY session.

Define a session, including this information:

- The *auto-login user name* and setting the auto-login user name to your SAN Volume Controller admin user name (for example, `admin`). Set this parameter under the **Connection** → **Data** category, as shown in Figure 9-3.

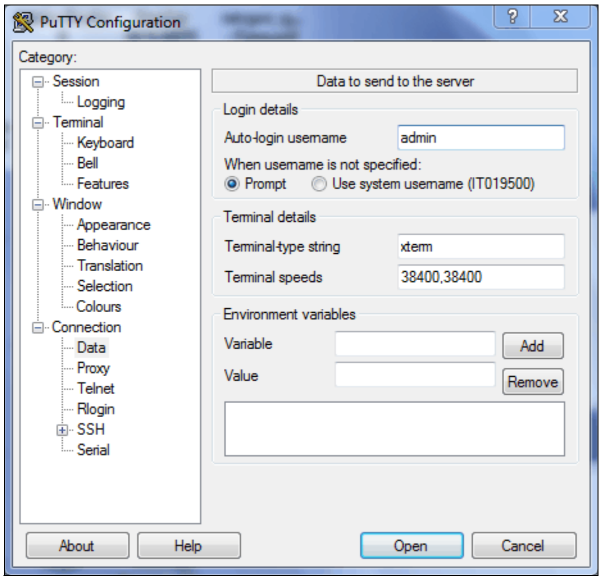


Figure 9-3 Auto-login configuration

- The *private key for authentication* (for example, `icat.ppk`). This key is the private key that you have already created. Set this parameter under the **Connection** → **Session** → **Auth** category, as shown in Figure 9-4.

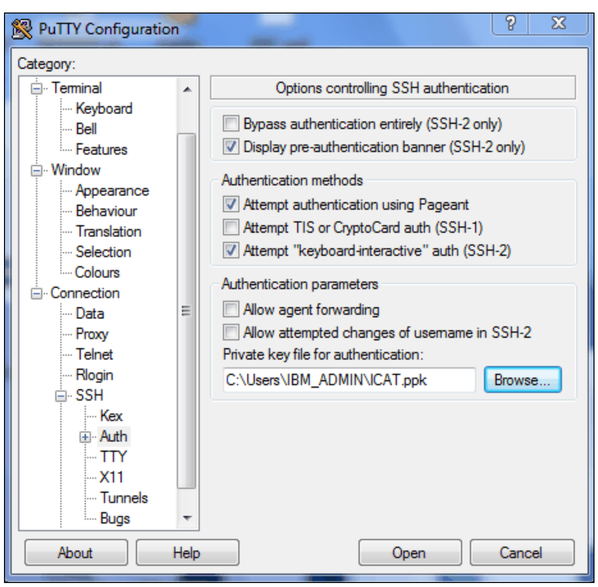


Figure 9-4 An ssh private key configuration

- The IP address of the SAN Volume Controller clustered system. Set this parameter under the **Session** category, as shown in Figure 9-5.

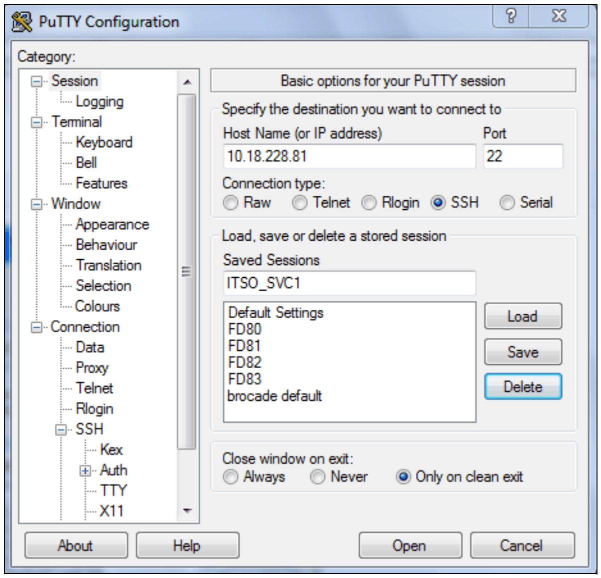


Figure 9-5 IP address

Enter this information:

– A session name. Our example uses `ITSO_SVC1`.

– Our PuTTy version is `0.63`.

– To use this predefined PuTTY session, use the following syntax:

```
plink ITSO_SVC1
```

– If a predefined PuTTY session is not used, use this syntax:

```
plink admin@<your cluster ip add> -i "C:\DirectoryPath\KeyName.PPK"
```

IBM provides a suite of scripting tools based on Perl. You can download these scripting tools from this website:

http://www.alphaworks.ibm.com/tech/svctools

9.8 SAN Volume Controller advanced operations using the CLI

In the following sections, we describe the commands that we think best represent advanced operational commands.

Important command prefix changes: The **svctask** and **svcinfo** command prefixes are no longer necessary when issuing a command. If you have existing scripts that use those prefixes, they will continue to function. You do not need to change the scripts.

9.8.1 Command syntax

Two major command sets are available:

- The **svcinfo** command set allows you to query the various components within the SAN Volume Controller environment.
- The **svctask** command set allows you to make changes to the various components within the SAN Volume Controller.

When the command syntax is shown, you see several parameters in square brackets, for example, `[parameter]`, which indicate that the parameter is optional in most if not all instances. Any parameter that is not in square brackets is required information. You can view the syntax of a command by entering one of the following commands:

svcinfo -?	Shows a complete list of information commands.
svctask -?	Shows a complete list of task commands.
svcinfo commandname -?	Shows the syntax of information commands.
svctask commandname -?	Shows the syntax of task commands.
svcinfo commandname -filtervalue?	Shows which filters you can use to reduce the output of the information commands.

Help: You can also use **-h** instead of **-?**, for example, **svcinfo -h** or **svctask commandname -h**.

If you look at the syntax of the command by typing **svcinfo command name -?**, you often see **-filter** listed as a parameter. Be aware that the correct parameter is **-filtervalue**.

Tip: You can use the up and down arrow keys on your keyboard to recall commands that were issued recently. Then, you can use the left and right, Backspace, and Delete keys to edit commands before you resubmit them.

9.8.2 Organizing on window content

Sometimes the output of a command can be long and difficult to read in the window. In cases where you need information about a subset of the total number of available items, you can use filtering to reduce the output to a more manageable size.

Filtering

To reduce the output that is displayed by a command, you can specify a number of filters, depending on which command you are running. To see which filters are available, type the command followed by the **-filtervalue?** flag, as shown in Example 9-90.

Example 9-90 lsvdisk -filtervalue? command

```
IBM_2145:ITSO_SVC1:admin>lsvdisk -filtervalue?

Filters for this view are :
     name
     id
     IO_group_id
     IO_group_name
     status
     mdisk_grp_name
     mdisk_grp_id
     capacity
     type
```

```
FC_id
FC_name
RC_id
RC_name
vdisk_name
vdisk_id

vdisk_UID
 fc_map_count
copy_count
fast_write_state
se_copy_count
filesystem
preferred_node_id
mirror_write_priority
  RC_flash
```

When you know the filters, you can be more selective in generating output:

- Multiple filters can be combined to create specific searches.
- You can use an asterisk (*) as a wildcard when using names.
- When capacity is used, the units must also be specified using **-u b** | **kb** | **mb** | **gb** | **tb** | **pb**.

For example, if we issue the **lsvdisk** command with no filters but with the **-delim** parameter, we see the output that is shown in Example 9-91.

Example 9-91 lsvdisk command: No filters

```
IBM_2145:ITSO_SVC1:admin>lsvdisk -delim ,
id,name,IO_group_id,IO_group_name,status,mdisk_grp_id,mdisk_grp_name,capacity,type,FC_id,FC
_name,RC_id,RC_name,vdisk_UID,fc_map_count,copy_count,fast_write_state,se_copy_count,RC_cha
nge
0,ESXI_SRV1_VOL01,1,io_grp1,online,many,many,100.00GB,many,,,,,6005076801AF813F100000000000
0014,0,2,empty,0,no
1,volume_7,0,io_grp0,online,0,STGPool_DS3500-1,10.00GB,striped,,,,,6005076801AF813F10000000
0000001F,0,1,empty,1,no
2,W2K3_SRV1_VOL02,1,io_grp1,online,0,STGPool_DS3500-1,10.00GB,striped,,,,,6005076801AF813F1
000000000000003,0,1,empty,0,no
3,W2K3_SRV1_VOL03,1,io_grp1,online,0,STGPool_DS3500-1,10.00GB,striped,,,,,6005076801AF813F1
000000000000004,0,1,empty,0,no
4,W2K3_SRV1_VOL04,1,io_grp1,online,0,STGPool_DS3500-1,10.00GB,striped,,,,,6005076801AF813F1
000000000000005,0,1,empty,0,no
5,W2K3_SRV1_VOL05,1,io_grp1,online,0,STGPool_DS3500-1,10.00GB,striped,,,,,6005076801AF813F1
000000000000006,0,1,empty,0,no
6,W2K3_SRV1_VOL06,1,io_grp1,online,0,STGPool_DS3500-1,10.00GB,striped,,,,,6005076801AF813F1
000000000000007,0,1,empty,0,no
7,W2K3_SRV2_VOL01,0,io_grp0,online,1,STGPool_DS3500-2,10.00GB,striped,,,,,6005076801AF813F1
000000000000008,0,1,empty,0,no
8,W2K3_SRV2_VOL02,0,io_grp0,online,1,STGPool_DS3500-2,10.00GB,striped,,,,,6005076801AF813F1
000000000000009,0,1,empty,0,no
```

> **Tip:** The **-delim** parameter truncates the content in the window and separates data fields with colons as opposed to wrapping text over multiple lines. This parameter is normally used in cases where you need to get reports during script execution.

If we now add a filter to our **lsvdisk** command (**mdisk_grp_name**) we can reduce the output, as shown in Example 9-92 on page 221.

Example 9-92 lsvdisk command: With a filter

```
IBM_2145:ITSO_SVC1:admin>lsvdisk -filtervalue mdisk_grp_name=STGPool_DS3500-2
id name            IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name   capacity
type    FC_id FC_name RC_id RC_name vdisk_UID                        fc_map_count
copy_count fast_write_state se_copy_count RC_change
7,W2K3_SRV2_VOL01,0,io_grp0,online,1,STGPool_DS3500-2,10.00GB,striped,,,,,6005076801AF813F1
000000000000008,0,1,empty,0,no
8,W2K3_SRV2_VOL02,0,io_grp0,online,1,STGPool_DS3500-2,10.00GB,striped,,,,,6005076801AF813F1
000000000000009,0,1,empty,0,no
```

9.9 Managing the clustered system using the CLI

In these sections, we demonstrate how to perform system administration.

9.9.1 Viewing clustered system properties

Changes since SAN Volume Controller 6.3:

- The **svcinfo lscluster** command is changed to **lssystem**.
- The **svctask chcluster** command is changed to **chsystem**, and several optional parameters have moved to new commands. For example, to change the IP address of the system, you can now use the **chsystemip** command. All the old commands are maintained for compatibility reasons.

Use the **lssystem** command to display summary information about the clustered system, as shown in Example 9-93.

Example 9-93 lssystem command

```
IBM_2145:ITSO_SVC1:admin>lssystem
id 000002006BE04FC4
name ITSO_SVC1
location local
partnership
bandwidth
total_mdisk_capacity 836.5GB
space_in_mdisk_grps 786.5GB
space_allocated_to_vdisks 434.02GB
total_free_space 402.5GB
total_vdiskcopy_capacity 442.00GB
total_used_capacity 432.00GB
total_overallocation 52
total_vdisk_capacity 341.00GB
total_allocated_extent_capacity 435.75GB
statistics_status on
statistics_frequency 15
cluster_locale en_US
time_zone 520 US/Pacific
code_level 6.3.0.0 (build 54.0.1109090000)
console_IP 10.18.228.81:443
id_alias 000002006BE04FC4
gm_link_tolerance 300
gm_inter_cluster_delay_simulation 0
gm_intra_cluster_delay_simulation 0
```

```
gm_max_host_delay 5
email_reply
email_contact
email_contact_primary
email_contact_alternate
email_contact_location
email_contact2
email_contact2_primary
email_contact2_alternate
email_state stopped
inventory_mail_interval 0
cluster_ntp_IP_address 69.50.219.51
cluster_isns_IP_address
iscsi_auth_method none
iscsi_chap_secret
auth_service_configured no
auth_service_enabled no
auth_service_url
auth_service_user_name
auth_service_pwd_set no
auth_service_cert_set no
auth_service_type tip
relationship_bandwidth_limit 25
tier generic_ssd
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier generic_hdd
tier_capacity 786.50GB
tier_free_capacity 352.25GB
has_nas_key no
layer appliance
```

Use the **lssystemstats** command to display the most recent values of all node statistics across all nodes in a clustered system, as shown in Example 9-94.

Example 9-94 lssystemstats command

```
IBM_2145:ITSO_SVC1:admin>lssystemstats
stat_name          stat_current stat_peak stat_peak_time
cpu_pc             1            1         110927162859
fc_mb              0            0         110927162859
fc_io              7091         7314      110927162524
sas_mb             0            0         110927162859
sas_io             0            0         110927162859
iscsi_mb           0            0         110927162859
iscsi_io           0            0         110927162859
write_cache_pc     0            0         110927162859
total_cache_pc     0            0         110927162859
vdisk_mb           0            0         110927162859
vdisk_io           0            0         110927162859
vdisk_ms           0            0         110927162859
mdisk_mb           0            0         110927162859
mdisk_io           0            0         110927162859
mdisk_ms           0            0         110927162859
drive_mb           0            0         110927162859
drive_io           0            0         110927162859
drive_ms           0            0         110927162859
vdisk_r_mb         0            0         110927162859
vdisk_r_io         0            0         110927162859
vdisk_r_ms         0            0         110927162859
```

```
vdisk_w_mb        0         0         110927162859
vdisk_w_io        0         0         110927162859
vdisk_w_ms        0         0         110927162859
mdisk_r_mb        0         0         110927162859
mdisk_r_io        0         0         110927162859
mdisk_r_ms        0         0         110927162859
mdisk_w_mb        0         0         110927162859
mdisk_w_io        0         0         110927162859
mdisk_w_ms        0         0         110927162859
drive_r_mb        0         0         110927162859
drive_r_io        0         0         110927162859
drive_r_ms        0         0         110927162859
drive_w_mb        0         0         110927162859
drive_w_io        0         0         110927162859
drive_w_ms        0         0         110927162859
```

9.9.2 Changing system settings

Use the **chsystem** command to change the settings of the system. This command modifies specific features of a clustered system. You can change multiple features by issuing a single command.

All command parameters are optional; however, you must specify at least one parameter.

Important considerations:

- Starting with SAN Volume Controller 6.3, the **svctask chcluster** command is changed to **chsystem**, and several optional parameters have moved to new commands. For example, to change the IP address of the system, you can now use the **chsystemip** command. All the old commands are maintained for script compatibility reasons.
- Changing the speed on a running system breaks I/O service to the attached hosts. Before changing the fabric speed, stop the I/O from the active hosts and force these hosts to flush any cached data by unmounting volumes (for UNIX host types) or by removing drive letters (for Windows host types). You might need to reboot specific hosts to detect the new fabric speed.

Example 9-95 shows configuring the Network Time Protocol (NTP) IP address.

Example 9-95 chsystem command

```
IBM_2145:ITSO_SVC1:admin>chsystem -ntpip 10.200.80.1
```

9.9.3 iSCSI configuration

SAN Volume Controller 5.1 introduced the IP-based Small Computer System Interface (iSCSI) as a supported method of communication between the SAN Volume Controller and hosts. All back-end storage and intracluster communication still use FC and the SAN, so iSCSI cannot be used for that type of communication.

We configured our nodes to use the primary and secondary Ethernet ports for iSCSI and to contain the clustered system IP. When we configured our nodes to be used with iSCSI, we did not affect our clustered system IP. The clustered system IP is changed, as shown in 9.9.2, “Changing system settings” on page 223.

It is important to know that you can have more than a one IP address-to-one physical connection relationship. We have the capability to have a four-to-one relationship (4:1), consisting of two IPv4 plus two IPv6 addresses (four total) to one physical connection per port per node.

Tip: When reconfiguring IP ports, be aware that already configured iSCSI connections will need to reconnect if changes are made to the IP addresses of the nodes.

There are two ways to perform iSCSI authentication or Challenge Handshake Authentication Protocol (CHAP), either for the whole clustered system or per host connection. Example 9-96 shows configuring CHAP for the whole clustered system.

Example 9-96 Setting a CHAP secret for the entire clustered system to "passw0rd"

```
IBM_2145:ITSO_SVC1:admin>chsystem -iscsiauthmethod chap -chapsecret passw0rd
```

In our scenario, we have a clustered system IP of `9.64.210.64`, which is not affected during our configuration of the node's IP addresses.

We start by listing our ports using the **`lsportip`** command (not shown). We see that we have two ports per node with which to work. Both ports can have two IP addresses that can be used for iSCSI.

We configure the secondary port in both nodes in our I/O Group, as shown in Example 9-97.

Example 9-97 Configuring secondary Ethernet port on SAN Volume Controller nodes

```
IBM_2145:ITSO_SVC1:admin>cfgportip -node 1 -ip 9.8.7.1 -gw 9.0.0.1 -mask 255.255.255.0 2
IBM_2145:ITSO_SVC1:admin>cfgportip -node 2 -ip 9.8.7.3 -gw 9.0.0.1 -mask 255.255.255.0 2
```

While both nodes are online, each node will be available to iSCSI hosts on the IP address that we have configured. Note that iSCSI failover between nodes is enabled automatically. Therefore, if a node goes offline for any reason, its partner node in the I/O Group will become available on the failed node's port IP address. This design ensures that hosts can continue to perform I/O. The **`lsportip`** command displays the port IP addresses that are currently active on each node.

9.9.4 Modifying IP addresses

We can use both IP ports of the nodes. However, all IP information is required the first time that you configure a second port because port 1 on the system must always have one stack fully configured.

There are now two active system ports on the configuration node. If the system IP address is changed, the open command-line shell closes during the processing of the command. You must reconnect to the new IP address if connected through that port.

If the clustered system IP address is changed, the open command-line shell closes during the processing of the command and you must reconnect to the new IP address. If this node cannot rejoin the clustered system, you can bring up the node in service mode. In this mode, the node can be accessed as a stand-alone node using the service IP address.

We discuss the service IP address in more detail in 9.20, "Working with the Service Assistant menu" on page 323.

List the IP addresses of the clustered system by issuing the **lssystemip** command, as shown in Example 9-98.

Example 9-98 lssystemip command

```
IBM_2145:ITSO_SVC1:admin>lssystemip
cluster_id       cluster_name location port_id IP_address   subnet_mask   gateway
IP_address_6                             prefix_6 gateway_6
000002006BE04FC4 ITSO_SVC1    local    1       10.18.228.81 255.255.255.0 10.18.228.1
fd09:5030:beef:cafe:0000:0000:0000:0083 64       fd09:5030:beef:cafe:0000:0000:0000:0001
000002006BE04FC4 ITSO_SVC1   local   2
000002006AC03A42 ITSO_SVC2  remote 1   10.18.228.82 255.255.255.0 10.18.228.1
000002006AC03A42 ITSO_SVC2   remote  2
0000020060A06FB8 ITSO_SVC3    remote   1       10.18.228.83 255.255.255.0 10.18.228.83
fdee:beeb:beeb:0000:0000:0000:0000:0083 48       fdee:beeb:beeb:0000:0000:0000:0000:0083
0000020060A06FB8 ITSO_SVC3   remote  2
```

Modify the IP address by issuing the **chsystemip** command. You can either specify a static IP address or have the system assign a dynamic IP address, as shown in Example 9-99.

Example 9-99 chsystemip -systemip

```
IBM_2145:ITSO_SVC1:admin>chsystemip -systemip 10.20.133.5 -gw 10.20.135.1 -mask
255.255.255.0 -port 1
```

This command changes the current IP address of the clustered system to `10.20.133.5`.

Important: If you specify a new system IP address, the existing communication with the system through the CLI is broken and the PuTTY application automatically closes. You must relaunch the PuTTY application and point to the new IP address, but your SSH key still works.

List the IP service addresses of the clustered system by issuing the **lsserviceip** command.

9.9.5 Supported IP address formats

Table 9-1 lists the IP address formats.

Table 9-1 ip_address_list formats

IP type	ip_address_list format
IPv4 (no port set, SAN Volume Controller uses default)	1.2.3.4
IPv4 with specific port	1.2.3.4:22
Full IPv6, default port	1234:1234:0001:0123:1234:1234:1234:1234
Full IPv6, default port, leading zeros suppressed	1234:1234:1:123:1234:1234:1234:1234
Full IPv6 with port	[2002:914:fc12:848:209:6bff:fe8c:4ff6]:23
Zero-compressed IPv6, default port	2002::4ff6
Zero-compressed IPv6 with port	[2002::4ff6]:23

At this point, we have completed the tasks that are required to change the IP addresses of the clustered system.

9.9.6 Setting the clustered system time zone and time

Use the **-timezone** parameter to specify the numeric ID of the time zone that you want to set. Issue the **lstimezones** command to list the time zones that are available on the system; this command displays a list of valid time zone settings.

> **Tip:** If you have changed the time zone, you must clear the event log dump directory before you can view the event log through the web application.

Setting the clustered system time zone

Perform the following steps to set the clustered system time zone and time:

1. Find out for which time zone your system is currently configured. Enter the **showtimezone** command, as shown in Example 9-100.

   Example 9-100 showtimezone command

   ```
   IBM_2145:ITSO_SVC1:admin>showtimezone
   id              timezone
   522             UTC
   ```

2. To find the time zone code that is associated with your time zone, enter the **lstimezones** command, as shown in Example 9-101. A truncated list is provided for this example. If this setting is correct (for example, `522 UTC`), go to Step 4. If not, continue with Step 3.

   Example 9-101 lstimezones command

   ```
   IBM_2145:ITSO_SVC1:admin>lstimezones
   id              timezone
   .
   .
   507             Turkey
   508             UCT
   509             Universal
   510             US/Alaska
   511             US/Aleutian
   512             US/Arizona
   513             US/Central
   514             US/Eastern
   515             US/East-Indiana
   516             US/Hawaii
   517             US/Indiana-Starke
   518             US/Michigan
   519             US/Mountain
   520             US/Pacific
   521             US/Samoa
   522             UTC
   .
   .
   ```

3. Now that you know which time zone code is correct for you, set the time zone by issuing the **settimezone** (Example 9-102 on page 227) command.

Example 9-102 settimezone command

```
IBM_2145:ITSO_SVC1:admin>settimezone -timezone 520
```

4. Set the system time by issuing the **setclustertime** command (Example 9-103).

Example 9-103 setclustertime command

```
IBM_2145:ITSO_SVC1:admin>setclustertime -time 061718402008
```

The format of the time is *MMDDHHmmYYYY*.

You have completed the necessary tasks to set the clustered system time zone and time.

9.9.7 Starting statistics collection

Statistics are collected at the end of each sampling period (as specified by the **-interval** parameter). These statistics are written to a file. A new file is created at the end of each sampling period. Separate files are created for MDisks, volumes, and node statistics.

Use the **startstats** command to start the collection of statistics, as shown in Example 9-104.

Example 9-104 startstats command

```
IBM_2145:ITSO_SVC1:admin>startstats -interval 15
```

The interval that we specify (minimum 1, maximum 60) is in minutes. This command starts statistics collection and gathers data at 15-minute intervals.

Statistics collection: To verify that the statistics collection is set, display the system properties again, as shown in Example 9-105.

Example 9-105 Statistics collection status and frequency

```
IBM_2145:ITSO_SVC1:admin>lssystem
statistics_status on
statistics_frequency 15
-- Note that the output has been shortened for easier reading. --
```

SAN Volume Controller 6.3: Starting with SAN Volume Controller 6.3, the command **svctask stopstats** has been removed. You cannot disable the statistics collection.

At this point, we have completed the required tasks to start the statistics collection on the clustered system.

9.9.8 Determining the status of a copy operation

Use the **lscopystatus** command, as shown in Example 9-106 on page 228, to determine if a file copy operation is in progress. Only one file copy operation can be performed at a time. The output of this command is a status of active or inactive.

Example 9-106 lscopystatus command

```
IBM_2145:ITSO_SVC1:admin>lscopystatus
status
inactive
```

9.9.9 Shutting down a clustered system

If all input power to a SAN Volume Controller system is to be removed for more than a few minutes (for example, if the machine room power is to be shut down for maintenance), it is important to shut down the clustered system *before* removing the power. If the input power is removed from the uninterruptible power supply units without first shutting down the system and the uninterruptible power supply units, the uninterruptible power supply units remain operational and eventually become drained of power.

When input power is restored to the uninterruptible power supply units, they start to recharge. However, the SAN Volume Controller does not permit any I/O activity to be performed to the volumes until the uninterruptible power supply units are charged enough to enable all of the data on the SAN Volume Controller nodes to be destaged in the event of a subsequent unexpected power loss. Recharging the uninterruptible power supply can take as long as two hours.

Shutting down the clustered system before removing input power to the uninterruptible power supply units prevents the battery power from being drained. It also makes it possible for I/O activity to be resumed as soon as input power is restored.

You can use the following procedure to shut down the system:

1. Use the **stopsystem** command to shut down your SAN Volume Controller system (Example 9-107).

   Example 9-107 stopsystem command

   ```
   IBM_2145:ITSO_SVC1:admin>stopsystem
   Are you sure that you want to continue with the shut down?
   ```

   This command shuts down the SAN Volume Controller clustered system. All data is flushed to disk before the power is removed. At this point, you lose administrative contact with your system, and the PuTTY application automatically closes.

2. You will be presented with the following message:

   Warning: `Are you sure that you want to continue with the shut down?`

   Ensure that you have stopped all FlashCopy mappings, Metro Mirror (remote copy) relationships, data migration operations, and forced deletions before continuing. Entering `y` to this message will execute the command. No feedback is then displayed. Entering anything other than `y(es)` or `Y(ES)` will result in the command not executing. No feedback is displayed.

   Important: Before shutting down a clustered system, ensure that all I/O operations are stopped that are destined for this system because you will lose *all* access to *all* volumes being provided by this system. Failure to do so can result in failed I/O operations being reported to the host operating systems.

   Begin the process of quiescing all I/O to the system by stopping the applications on the hosts that are using the volumes provided by the clustered system.

3. We have completed the tasks that are required to shut down the system. To shut down the uninterruptible power supply units, press the power-on button on the front panel of each uninterruptible power supply unit.

Restarting the system: To restart the clustered system, you must first restart the uninterruptible power supply units by pressing the power button on their front panels. Then, press the power-on button on the service panel of one of the nodes within the system. After the node is fully booted (for example, displaying `Cluster:` on line 1 and the cluster name on line 2 of the panel), you can start the other nodes in the same way.

As soon as all of the nodes are fully booted, you can reestablish administrative contact using PuTTY, and your system will be fully operational again.

9.10 Nodes

In this section, we describe the tasks that can be performed at an individual node level.

9.10.1 Viewing node details

Use the **lsnode** command to view the summary information about the nodes that are defined within the SAN Volume Controller environment. To view more details about a specific node, append the node name (for example, `SVC1N1`) to the command.

Example 9-108 shows both of these commands.

Tip: The **-delim** parameter truncates the content in the window and separates data fields with colons (:) as opposed to wrapping text over multiple lines.

Example 9-108 lsnode command

```
IBM_2145:ITSO_SVC1:admin>lsnode -delim ,
id,name,UPS_serial_number,WWNN,status,IO_group_id,IO_group_name,config_node,UPS_unique_id,h
ardware,iscsi_name,iscsi_alias,panel_name,enclosure_id,canister_id,enclosure_serial_number
1,SVC1N1,1000739004,50050768010027E2,online,0,io_grp0,no,10000000000027E2,8G4,iqn.1986-03.c
om.ibm:2145.itsosvc1.svc1n1,,108283,,,
2,SVC1N2,1000739005,5005076801005034,online,0,io_grp0,yes,1000000000005034,8G4,iqn.1986-03.
com.ibm:2145.itsosvc1.svc1n2,,110711,,,
3,SVC1N4,1000739006,500507680100505C,online,1,io_grp1,no,20400001C3240004,8G4,iqn.1986-03.c
om.ibm:2145.itsosvc1.svc1n4,,110775,,,
4,SVC1N3,1000739007,50050768010037E5,online,1,io_grp1,no,10000000000037E5,8G4,iqn.1986-03.c
om.ibm:2145.itsosvc1.svc1n3,,104643,,,
IBM_2145:ITSO_SVC1:admin>lsnode SVC1N1
id 1
name SVC1N1
UPS_serial_number 1000739004
WWNN 50050768010027E2
status online
IO_group_id 0
IO_group_name io_grp0
partner_node_id 2
partner_node_name SVC1N2
config_node no
UPS_unique_id 10000000000027E2
port_id 50050768014027E2
```

```
port_status active
port_speed 2Gb
port_id 50050768013027E2
port_status active
port_speed 2Gb
port_id 50050768011027E2
port_status active
port_speed 2Gb
port_id 50050768012027E2
port_status active
port_speed 2Gb
hardware 8G4
iscsi_name iqn.1986-03.com.ibm:2145.itsosvc1.svc1n1
iscsi_alias
failover_active no
failover_name SVC1N2
failover_iscsi_name iqn.1986-03.com.ibm:2145.itsosvc1.svc1n2
failover_iscsi_alias
panel_name 108283
enclosure_id
canister_id
enclosure_serial_number
service_IP_address 10.18.228.101
service_gateway 10.18.228.1
service_subnet_mask 255.255.255.0
service_IP_address_6
service_gateway_6
service_prefix_6
```

9.10.2 Adding a node

After clustered system creation is completed through the service panel (the front panel of one of the SAN Volume Controller nodes) and the system web interface, only one node (the configuration node) is set up.

To have a fully functional SAN Volume Controller system, you must add a second node to the configuration. To add a node to a clustered system, gather the necessary information, as explained in these steps:

1. Before you can add a node, you must know which unconfigured nodes are available as "candidates". Issue the **lsnodecandidate** command (Example 9-109).

   Example 9-109 lsnodecandidate command

   ```
   IBM_2145:ITSO_SVC1:admin>lsnodecandidate
   id               panel_name UPS_serial_number UPS_unique_id    hardware
   50050768010037E5 104643 1000739007  10000000000037E5
   8G4
   ```

2. You must specify to which I/O Group you are adding the node. If you enter the **lsnode** command, you can easily identify the I/O Group ID of the group to which you are adding your node, as shown in Example 9-110 on page 231.

   Tip: The node that you want to add must have a separate uninterruptible power supply unit serial number from the uninterruptible power supply unit on the first node.

Example 9-110 lsnode command

```
IBM_2145:ITSO_SVC1:admin>lsnode -delim ,
id,name,UPS_serial_number,WWNN,status,IO_group_id,IO_group_name,config_node,UPS
_unique_id,hardware,iscsi_name,iscsi_alias,panel_name,enclosure_id,canister_id,
enclosure_serial_number
4,SVC1N3,1000739007,50050768010037E5,online,1,io_grp1,no,10000000000037E5,8G4,i
qn.1986-03.com.ibm:2145.itsosvc1.svc1n3,,104643,,,
```

3. Now that we know the available nodes, we can use the **addnode** command to add the node to the SAN Volume Controller clustered system configuration. Example 9-111 shows the command to add a node to the SAN Volume Controller system.

Example 9-111 addnode (wwnodename) command

```
IBM_2145:ITSO_SVC1:admin>addnode -wwnodename 50050768010037E5 -iogrp io_grp1
Node, id [5], successfully added
```

This command adds the candidate node with the wwnodename of 50050768010037E5 to the I/O Group called io_grp1.

We used the **-wwnodename** parameter (50050768010037E5). However, we can also use the **-panelname** parameter (104643) instead, as shown in Example 9-112. If standing in front of the node, it is easier to read the panel name than it is to get the worldwide node name (WWNN).

Example 9-112 addnode (panelname) command

```
IBM_2145:ITSO_SVC1:admin>addnode -panelname 104643 -name SVC1N3 -iogrp io_grp1
```

We also used the optional **-name** parameter (SVC1N3). If you do not provide the **-name** parameter, the SAN Volume Controller automatically generates the name node*x* (where *x* is the ID sequence number that is assigned internally by the SAN Volume Controller).

> **Name:** If you want to provide a name, you can use letters A to Z and a to z, numbers 0 - 9, the dash (-), and the underscore (_). The name can be between 1 and 63 characters in length. However, the name cannot start with a number, dash, or the word "node" (because this prefix is reserved for SAN Volume Controller assignment only).

4. If the **addnode** command returns no information, your second node is powered on, and the zones are correctly defined, the preexisting system configuration data can be stored in the node. If you are sure that this node is not part of another active SAN Volume Controller system, you can use the service panel to delete the existing system information. After this action is complete, reissue the **lsnodecandidate** command and you will see it listed.

9.10.3 Renaming a node

Use the **chnode** command to rename a node within the SAN Volume Controller system configuration, as shown in Example 9-113.

Example 9-113 chnode -name command

```
IBM_2145:ITSO_SVC1:admin>chnode -name ITSO_SVC1_SVC1N3 4
```

This command renames node ID 4 to ITSO_SVC1_SVC1N3 4.

Name: The **chnode** command specifies the new name first. You can use letters A to Z and a to z, numbers 0 - 9, the dash (-), and the underscore (_). The name can be between 1 and 63 characters in length. However, the name cannot start with a number, dash, or the word "node" (because this prefix is reserved for SAN Volume Controller assignment only).

9.10.4 Deleting a node

Use the **rmnode** command to remove a node from the SAN Volume Controller clustered system configuration (Example 9-114).

Example 9-114 rmnode command

```
IBM_2145:ITSO_SVC1:admin>rmnode SVC1N2
```

This command removes SVC1N2 from the SAN Volume Controller clustered system.

Because SVC1N2 was also the configuration node, the SAN Volume Controller transfers the configuration node responsibilities to a surviving node, within the I/O Group. Unfortunately, the PuTTY session cannot be dynamically passed to the surviving node. Therefore, the PuTTY application loses communication and closes automatically.

We must restart the PuTTY application to establish a secure session with the new configuration node.

Important: If this node is the last node in an I/O Group, and there are volumes still assigned to the I/O Group, the node is *not* deleted from the clustered system.

If this node is the last node in the system, and the I/O Group has *no* volumes remaining, the clustered system is destroyed and all virtualization information is lost. Any data that is still required must be backed up or migrated before destroying the system.

9.10.5 Shutting down a node

On occasion, it can be necessary to shut down a single node within the clustered system to perform tasks, such as scheduled maintenance, while leaving the SAN Volume Controller environment up and running.

Use the **stopcluster -node** command, as shown in Example 9-115, to shut down a single node.

Example 9-115 stopcluster -node command

```
IBM_2145:ITSO_SVC1:admin>stopcluster -node SVC1N3
Are you sure that you want to continue with the shut down?
```

This command shuts down node SVC1N3 in a graceful manner. When this node has been shut down, the other node in the I/O Group will destage the contents of its cache and will go into write-through mode until the node is powered up and rejoins the clustered system.

Important: There is no need to stop FlashCopy mappings, remote copy relationships, and data migration operations. The other node will handle these activities, but be aware that the system has a single point of failure now.

If this node is the last node in an I/O Group, all access to the volumes in the I/O Group will be lost. Verify that you want to shut down this node before executing this command. You must specify the **-force** flag.

By reissuing the **lsnode** command (Example 9-116), we can see that the node is now offline.

Example 9-116 lsnode command

```
IBM_2145:ITSO_SVC1:admin>lsnode -delim ,
id,name,UPS_serial_number,WWNN,status,IO_group_id,IO_group_name,config_node,UPS_unique_id,h
ardware,iscsi_name,iscsi_alias,panel_name,enclosure_id,canister_id,enclosure_serial_number
1,SVC1N1,1000739004,50050768010027E2,online,0,io_grp0,no,10000000000027E2,8G4,iqn.1986-03.c
om.ibm:2145.itsosvc1.svc1n1,,108283,,,
2,SVC1N2,1000739005,5005076801005034,online,0,io_grp0,yes,1000000000005034,8G4,iqn.1986-03.
com.ibm:2145.itsosvc1.svc1n2,,110711,,,
3,SVC1N4,1000739006,500507680100505C,online,1,io_grp1,no,20400001C3240004,8G4,iqn.1986-03.c
om.ibm:2145.itsosvc1.svc1n4,,110775,,,
4,SVC1N3,1000739007,50050768010037E5,offline,1,io_grp1,no,10000000000037E5,8G4,iqn.1986-03.
com.ibm:2145.itsosvc1.svc1n3,,104643,,,
IBM_2145:ITSO_SVC1:admin>lsnode SVC1N3
CMMVC5782E The object specified is offline.
```

Restart: To restart the node manually, press the power-on button that is on the service panel of the node.

At this point, we have completed the tasks that are required to view, add, delete, rename, and shut down a node within a SAN Volume Controller environment.

9.11 I/O Groups

In this section, we explain the tasks that you can perform at an I/O Group level.

9.11.1 Viewing I/O Group details

Use the **lsiogrp** command, as shown in Example 9-117, to view information about the I/O Groups that are defined within the SAN Volume Controller environment.

Example 9-117 I/O Group details

```
IBM_2145:ITSO_SVC1:admin>lsiogrp
id name            node_count vdisk_count host_count
0  io_grp0         2          24          9
1  io_grp1         2          22          9
2  io_grp2         0          0           1
3  io_grp3         0          0           1
4  recovery_io_grp 0          0           0
```

As shown, the SAN Volume Controller predefines five I/O Groups. In a four-node clustered system (similar to our example), only two I/O Groups are actually in use. The other I/O Groups (io_grp2 and io_grp3) are for a six-node or eight-node clustered system.

The recovery I/O Group is a temporary home for volumes when all nodes in the I/O Group that normally owns them have suffered multiple failures. This design allows us to move the volumes to the recovery I/O Group and then into a working I/O Group. Note that while temporarily assigned to the recovery I/O Group, I/O access is not possible.

9.11.2 Renaming an I/O Group

Use the **chiogrp** command to rename an I/O Group (Example 9-118).

Example 9-118 chiogrp command

```
IBM_2145:ITSO_SVC1:admin>chiogrp -name io_grpA io_grp1
```

This command renames the I/O Group `io_grp1` to `io_grpA`.

> **Name:** The **chiogrp** command specifies the new name first.
>
> If you want to provide a name, you can use letters A to Z, letters a to z, numbers 0 - 9, the dash (-), and the underscore (_). The name can be between 1 and 63 characters in length. However, the name cannot start with a number, dash, or the word "`iogrp`" (because this prefix is reserved for SAN Volume Controller assignment only).

To see whether the renaming was successful, issue the **lsiogrp** command again to see the change.

At this point, we have completed the tasks that are required to rename an I/O Group.

9.11.3 Adding and removing hostiogrp

To map or unmap a specific host object to a specific I/O Group to reach the maximum number of hosts supported by a SAN Volume Controller clustered system, use the **addhostiogrp** command to map a specific host to a specific I/O Group, as shown in Example 9-119.

Example 9-119 addhostiogrp command

```
IBM_2145:ITSO_SVC1:admin>addhostiogrp -iogrp 1 Kanaga
```

The **addhostiogrp** command uses these parameters:

- **-iogrp** ***iogrp_list*** **-iogrpall**

  Specify a list of one or more I/O Groups that must be mapped to the host. This parameter is mutually exclusive with the **-iogrpall** option. The **-iogrpall** option specifies that all the I/O Groups must be mapped to the specified host. This parameter is mutually exclusive with **-iogrp**.

- **-host** ***host_id_or_name***

  Identify the host either by ID or name to which the I/O Groups must be mapped.

Use the **rmhostiogrp** command to unmap a specific host to a specific I/O Group, as shown in Example 9-120.

Example 9-120 rmhostiogrp command

```
IBM_2145:ITSO_SVC1:admin>rmhostiogrp -iogrp 0 Kanaga
```

The **rmhostiogrp** command uses these parameters:

- **-iogrp** ***iogrp_list*** **-iogrpall**

  Specify a list of one or more I/O Groups that must be unmapped to the host. This parameter is mutually exclusive with the **-iogrpall** option. The **-iogrpall** option specifies that all of the I/O Groups must be unmapped to the specified host. This parameter is mutually exclusive with **-iogrp**.

- **-force**

  If the removal of a host to I/O Group mapping will result in the loss of volume to host mappings, the command fails if the **-force** flag is not used. The **-force** flag, however, overrides this behavior and forces the deletion of the host to I/O Group mapping.

- ***host_id_or_name***

  Identify the host either by the ID or name to which the I/O Groups must be unmapped.

9.11.4 Listing I/O Groups

To list all of the I/O Groups that are mapped to the specified host and vice versa, use the **lshostiogrp** command, specifying the host name `Kanaga`, as shown in Example 9-121.

Example 9-121 lshostiogrp command

```
IBM_2145:ITSO_SVC1:admin>lshostiogrp Kanaga
id              name
1               io_grp1
```

To list all of the host objects that are mapped to the specified I/O Group, use the **lsiogrphost** command, as shown in Example 9-122.

Example 9-122 lsiogrphost command

```
IBM_2145:ITSO_SVC1:admin> lsiogrphost io_grp1
id              name
1               Nile
2               Kanaga
3               Siam
```

In Example 9-122, `io_grp1` is the I/O Group name.

9.12 Managing authentication

In the following sections, we illustrate authentication administration.

9.12.1 Managing users using the CLI

Here, we demonstrate how to operate and manage authentication by using the CLI. All users must now be a member of a predefined user group. You can list those groups by using the **lsusergrp** command, as shown in Example 9-123 on page 236.

Example 9-123 lsusergrp command

```
IBM_2145:ITSO_SVC1:admin>lsusergrp
id name          role          remote
0  SecurityAdmin SecurityAdmin no
1  Administrator Administrator no
2  CopyOperator  CopyOperator  no
3  Service       Service       no
4  Monitor       Monitor       no
```

Example 9-124 is a simple example of creating a user. User `John` is added to the user group `Monitor` with the password `m0nitor`.

Example 9-124 mkuser called John with password m0nitor

```
IBM_2145:ITSO_SVC1:admin>mkuser -name John -usergrp Monitor -password m0nitor
User, id [6], successfully created
```

Local users are users that are not authenticated by a remote authentication server. *Remote* users are users that are authenticated by a remote central registry server.

The user groups already have a defined authority role, as listed in Table 9-2.

Table 9-2 Authority roles

User group	Role	User
Security admin	All commands	Superusers
Administrator	All commands *except*: chauthservice, mkuser, rmuser, chuser, mkusergrp, rmusergrp, chusergrp, and setpwdreset	Administrators that control the SAN Volume Controller
Copy operator	All **`display`** commands and the following commands: prestartfcconsistgrp, startfcconsistgrp, stopfcconsistgrp, chfcconsistgrp, prestartfcmap, startfcmap, stopfcmap, chfcmap, startrcconsistgrp, stoprcconsistgrp, switchrcconsistgrp, chrcconsistgrp, startrcrelationship, stoprcrelationship, switchrcrelationship, chrcrelationship, and chpartnership	For users that control all of the copy functionality of the cluster

User group	Role	User
Service	All **display** commands and the following commands: applysoftware, setlocale, addnode, rmnode, cherrstate, writesernum, detectmdisk, includemdisk, clearerrlog, cleardumps, settimezone, stopcluster, startstats, stopstats, and settime	For users that perform service maintenance and other hardware tasks on the system
Monitor	All **display** commands and the following commands: finderr, dumperrlog, dumpinternallog, and chcurrentuser And svcconfig: backup	For users that only need view access

9.12.2 Managing user roles and groups

Role-based security commands are used to restrict the administrative abilities of a user. We cannot create new user roles, but we can create new user groups and assign a predefined role to our group.

As of SAN Volume Controller 6.3, you can connect to the clustered system using the same user name with which you log in to a SAN Volume Controller GUI.

To view the user roles on your system, use the **lsusergrp** command, as shown in Example 9-125, to list all users.

Example 9-125 lsusergrp command

```
IBM_2145:ITSO_SVC1:admin>lsusergrp
id name          role          remote
0  SecurityAdmin SecurityAdmin no
1  Administrator Administrator no
2  CopyOperator  CopyOperator  no
3  Service       Service       no
4  Monitor       Monitor       no
```

To view our currently defined users and the user groups to which they belong, we use the **lsuser** command, as shown in Example 9-126.

Example 9-126 lsuser command

```
IBM_2145:ITSO_SVC1:admin>lsuser -delim ,
id,name,password,ssh_key,remote,usergrp_id,usergrp_name
0,superuser,yes,no,no,0,SecurityAdmin
1,admin,yes,yes,no,0,SecurityAdmin
2,Torben,yes,no,no,0,SecurityAdmin
3,Massimo,yes,no,no,1,Administrator
4,Christian,yes,no,no,1,Administrator
5,Alejandro,yes,no,no,1,Administrator
6,John,yes,no,no,4,Monitor
```

9.12.3 Changing a user

To change user passwords, issue the **chuser** command.

The **chuser** command allows you to modify a user that is already created. You can rename a user, assign a new password (if you are logged on with administrative privileges), and move a user from one user group to another user group. Be aware, however, that a member can only be a member of one group at a time.

9.12.4 Audit log command

The audit log is extremely helpful in showing which commands have been entered on a system. Most action commands that are issued by the old or new CLI are recorded in the audit log:

- The native GUI performs actions by using the CLI programs.
- The SAN Volume Controller Console performs actions by issuing Common Information Model (CIM) commands to the CIM object manager (CIMOM), which then runs the CLI programs.

Actions performed by using both the native GUI and the SAN Volume Controller Console are recorded in the audit log.

Certain commands are not audited:

- `dumpconfig`
- `cpdumps`
- `cleardumps`
- `finderr`
- `dumperrlog`
- `dumpinternallog`
- `svcservicetask dumperrlog`
- `svcservicetask finderror`

The audit log contains approximately 1 MB of data, which can contain about 6000 average length commands. When this log is full, the system copies it to a new file in the `/dumps/audit` directory on the config node and resets the in-memory audit log.

To display entries from the audit log, use the **catauditlog -first 5** command to return a list of five in-memory audit log entries, as shown in Example 9-127.

Example 9-127 catauditlog command

```
IBM_2145:ITSO_SVC1:admin>catauditlog -first 5
audit_seq_no timestamp    cluster_user ssh_ip_address result res_obj_id action_cmd
459          110928150506 admin        10.18.228.173  0      6          svctask mkuser
-name John -usergrp Monitor -password '######'
460          110928160353 admin        10.18.228.173  0      7          svctask mkmdiskgrp
-name DS5000-2 -ext 256
461          110928160535 admin        10.18.228.173  0      1          svctask mkhost
-name hostone -hbawwpn 210100E08B251DD4 -force -mask 1001
462          110928160755 admin        10.18.228.173  0      1          svctask mkvdisk
-iogrp 0 -mdiskgrp 3 -size 10 -unit gb -vtype striped -autoexpand -grainsize 32 -rsize 20%
463          110928160817 admin        10.18.228.173  0                 svctask rmvdisk 1
```

If you need to dump the contents of the in-memory audit log to a file on the current configuration node, use the **dumpauditlog** command. This command does not provide any feedback; it only provides the prompt. To obtain a list of the audit log dumps, use the **lsdumps** command, as shown in Example 9-128.

Example 9-128 lsdumps command

```
IBM_2145:ITSO_SVC1:admin>lsdumps
id filename
0  dump.110711.110914.182844
1  svc.config.cron.bak_108283
2  sel.110711.trc
3  endd.trc
4  rtc.race_mq_log.txt.110711.trc
5  dump.110711.110920.102530
6  ethernet.110711.trc
7  svc.config.cron.bak_110711
8  svc.config.cron.xml_110711
9  svc.config.cron.log_110711
10 svc.config.cron.sh_110711
11 110711.trc
```

9.13 Managing Copy Services

In the following sections, we illustrate how to manage Copy Services.

9.13.1 FlashCopy operations

In this section, we use a scenario to illustrate how to use commands with PuTTY to perform FlashCopy. See the *IBM System Storage Open Software Family SAN Volume Controller: Command-Line Interface User's Guide*, GC27-2287, for information about other commands.

Scenario description

We use the following scenario in both the CLI section and the GUI section. In the following scenario, we want to FlashCopy the following volumes:

DB_Source Database files
Log_Source Database log files
App_Source Application files

We create Consistency Groups to handle the FlashCopy of `DB_Source` and `Log_Source` because data integrity must be kept on `DB_Source` and `Log_Source`.

In our scenario, the application files are independent of the database, so we create a single FlashCopy mapping for `App_Source`. We will make two FlashCopy targets for `DB_Source` and `Log_Source` and therefore, two Consistency Groups. Figure 9-6 on page 240 shows the scenario.

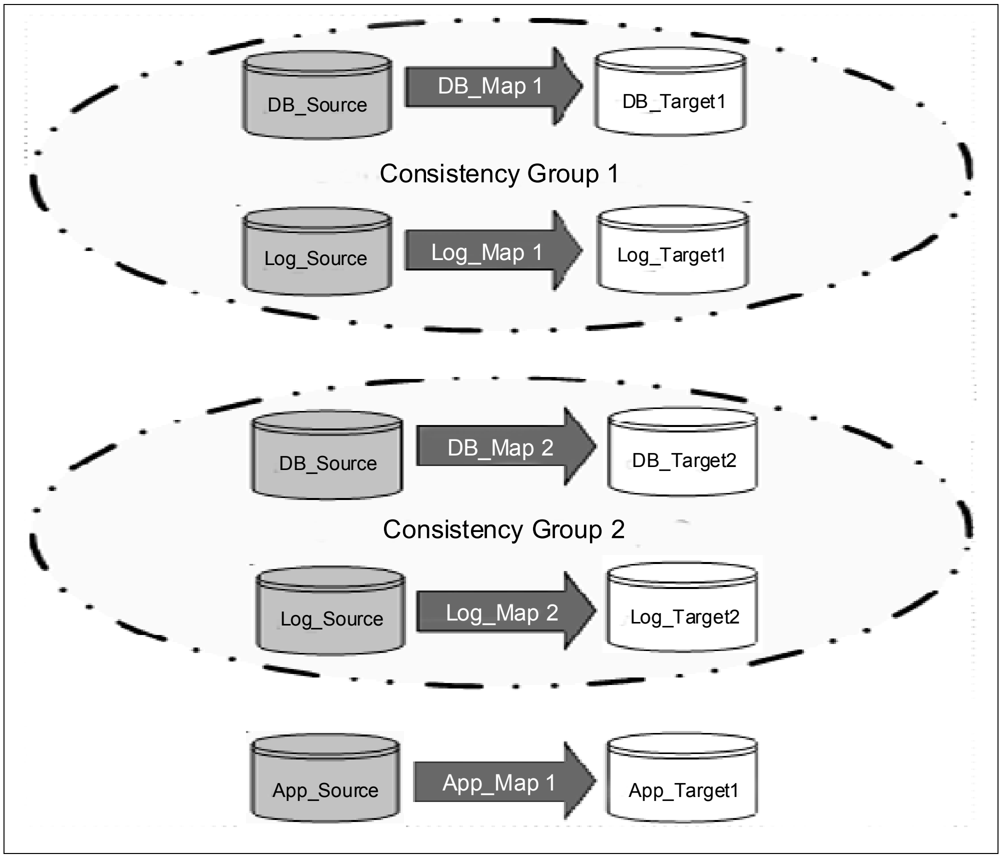


Figure 9-6 FlashCopy scenario

9.13.2 Setting up FlashCopy

We have already created the source and target volumes. The source and target volumes are identical in size, which is a requirement of the FlashCopy function:

- `DB_Source`, `DB_Target1`, and `DB_Target2`
- `Log_Source`, `Log_Target1`, and `Log_Target2`
- `App_Source` and `App_Target1`

To set up the FlashCopy, we perform the following steps:

1. Create two FlashCopy Consistency Groups:
   - `FCCG1`
   - `FCCG2`
2. Create FlashCopy mappings for Source volumes:
   - `DB_Source` FlashCopy to `DB_Target1`, the mapping name is `DB_Map1`
   - `DB_Source` FlashCopy to `DB_Target2`, the mapping name is `DB_Map2`
   - `Log_Source` FlashCopy to `Log_Target1`, the mapping name is `Log_Map1`
   - `Log_Source` FlashCopy to `Log_Target2`, the mapping name is `Log_Map2`
   - `App_Source` FlashCopy to `App_Target1`, the mapping name is `App_Map1`
   - Copyrate 50

9.13.3 Creating a FlashCopy Consistency Group

To create a FlashCopy Consistency Group, we use the command **mkfcconsistgrp** to create a new Consistency Group. The ID of the new group is returned. If you created several FlashCopy mappings for a group of volumes that contain elements of data for the same application, it might be convenient to assign these mappings to a single FlashCopy Consistency Group. Then, you can issue a single **prepare** or **start** command for the whole group so that, for example, all files for a particular database are copied at the same time.

In Example 9-129, the `FCCG1` and `FCCG2` Consistency Groups are created to hold the FlashCopy maps of `DB` and `Log`. This step is extremely important for FlashCopy on database applications because it helps to maintain data integrity during FlashCopy.

Example 9-129 Creating two FlashCopy Consistency Groups

```
IBM_2145:ITSO_SVC3:admin>mkfcconsistgrp -name FCCG1
FlashCopy Consistency Group, id [1], successfully created

IBM_2145:ITSO_SVC3:admin>mkfcconsistgrp -name FCCG2
FlashCopy Consistency Group, id [2], successfully created
```

In Example 9-130, we checked the status of the Consistency Groups. Each Consistency Group has a status of empty.

Example 9-130 Checking the status

```
IBM_2145:ITSO_SVC3:admin>lsfcconsistgrp
id name  status
1  FCCG1 empty
2  FCCG2 empty
```

If you want to change the name of a Consistency Group, you can use the **chfcconsistgrp** command. Type **chfcconsistgrp -h** for help with this command.

9.13.4 Creating a FlashCopy mapping

To create a FlashCopy mapping, we use the **mkfcmap** command. This command creates a new FlashCopy mapping, which maps a source volume to a target volume to prepare for subsequent copying.

When executed, this command creates a new FlashCopy mapping logical object. This mapping persists until it is deleted. The mapping specifies the source and destination volumes. The destination must be identical in size to the source or the mapping will fail. Issue the **lsvdisk -bytes** command to find the exact size of the source volume for which you want to create a target disk of the same size.

In a single mapping, source and destination cannot be on the same volume. A mapping is triggered at the point in time when the copy is required. The mapping can optionally be given a name and assigned to a Consistency Group. These groups of mappings can be triggered at the same time, enabling multiple volumes to be copied at the same time, which creates a consistent copy of multiple disks. A consistent copy of multiple disks is required for database products in which the database and log files reside on separate disks.

If no Consistency Group is defined, the mapping is assigned to the default group `0`, which is a special group that cannot be started as a whole. Mappings in this group can only be started on an individual basis.

The *background copy rate* specifies the priority that must be given to completing the copy. If `0` is specified, the copy will not proceed in the background. The default is `50`.

> **Tip:** There is a parameter to delete FlashCopy mappings automatically after the completion of a background copy (when the mapping gets to the idle_or_copied state). Use the command:
>
> `mkfcmap -autodelete`
>
> This command does not delete mappings in cascade with dependent mappings because it cannot get to the idle_or_copied state in this situation.

In Example 9-131, the first FlashCopy mapping for `DB_Source`, `Log_Source`, and `App_Source` is created.

Example 9-131 Create the first FlashCopy mapping for DB_Source, Log_Source, and App_Source

```
IBM_2145:ITSO_SVC3:admin>mkfcmap -source DB_Source -target DB_Target1 -name
DB_Map1 -consistgrp FCCG1
FlashCopy Mapping, id [0], successfully created

IBM_2145:ITSO_SVC3:admin>mkfcmap -source Log_Source -target Log_Target1 -name
Log_Map1 -consistgrp FCCG1
FlashCopy Mapping, id [1], successfully created

IBM_2145:ITSO_SVC3:admin>mkfcmap -source App_Source -target App_Target1 -name
App_Map1
FlashCopy Mapping, id [2], successfully created
```

Example 9-132 shows the command to create a second FlashCopy mapping for volume DB_Source and Log_Source.

Example 9-132 Create additional FlashCopy mappings

```
IBM_2145:ITSO_SVC3:admin>mkfcmap -source DB_Source -target DB_Target2 -name
DB_Map2 -consistgrp FCCG2
FlashCopy Mapping, id [3], successfully created

IBM_2145:ITSO_SVC3:admin>mkfcmap -source Log_Source -target Log_Target2 -name
Log_Map2 -consistgrp FCCG2
FlashCopy Mapping, id [4], successfully created
```

Example 9-133 shows the result of these FlashCopy mappings. The status of the mapping is idle_or_copied.

Example 9-133 Check the result of Multiple Target FlashCopy mappings

```
IBM_2145:ITSO_SVC3:admin>lsfcmap
id name     source_vdisk_id source_vdisk_name target_vdisk_id target_vdisk_name
group_id group_name status         progress copy_rate clean_progress incremental
partner_FC_id partner_FC_name restoring start_time rc_controlled
0  DB_Map1  3               DB_Source        4              DB_Target1       1
FCCG1    idle_or_copied 0      50      100         off
no                    no
1  Log_Map1 6               Log_Source       7              Log_Target1      1
FCCG1    idle_or_copied 0      50      100         off
no                    no
```

```
2  App_Map1 9           App_Source      10          App_Target1
idle_or_copied 0       50       100          off
no                   no
3  DB_Map2  3            DB_Source        5              DB_Target2       2
FCCG2    idle_or_copied 0      50      100         off
no                   no
4  Log_Map2 6            Log_Source       8              Log_Target2      2
FCCG2    idle_or_copied 0      50      100         off
no                   no
IBM_2145:ITSO_SVC3:admin>lsfcconsistgrp
id name  status
1  FCCG1 idle_or_copied
2  FCCG2 idle_or_copied
```

If you want to change the FlashCopy mapping, you can use the **chfcmap** command. Type **chfcmap -h** to get help with this command.

9.13.5 Preparing (pre-triggering) the FlashCopy mapping

At this point, the mapping has been created, but the cache still accepts data for the source volumes. You can only trigger the mapping when the cache does not contain any data for FlashCopy source volumes. You must issue a **prestartfcmap** command to prepare a FlashCopy mapping to start. This command tells the SAN Volume Controller to flush the cache of any content for the source volume and to pass through any further write data for this volume.

When the **prestartfcmap** command is executed, the mapping enters the Preparing state. After the preparation is complete, it changes to the Prepared state. At this point, the mapping is ready for triggering. Preparing and the subsequent triggering are usually performed on a Consistency Group basis. Only mappings belonging to Consistency Group 0 can be prepared on their own because Consistency Group 0 is a special group that contains the FlashCopy mappings that do not belong to any Consistency Group. A FlashCopy must be prepared before it can be triggered.

In our scenario, App_Map1 is not in a Consistency Group. In Example 9-134, we show how to initialize the preparation for App_Map1.

Another option is that you add the **-prep** parameter to the **startfcmap** command, which first prepares the mapping and then starts the FlashCopy.

In the example, we also show how to check the status of the current FlashCopy mapping. The status of App_Map1 is prepared.

Example 9-134 Prepare a FlashCopy without a Consistency Group

```
IBM_2145:ITSO_SVC3:admin>prestartfcmap App_Map1

IBM_2145:ITSO_SVC3:admin>lsfcmap App_Map1
id 2
name App_Map1
source_vdisk_id 9
source_vdisk_name App_Source
target_vdisk_id 10
target_vdisk_name App_Target1
group_id
group_name
status prepared
```

```
progress 0
copy_rate 50
start_time
dependent_mappings 0
autodelete off
clean_progress 0
clean_rate 50
incremental off
difference 0
grain_size 256
IO_group_id 0
IO_group_name io_grp0
partner_FC_id
partner_FC_name
restoring no
rc_controlled no
```

9.13.6 Preparing (pre-triggering) the FlashCopy Consistency Group

We use the **prestartfcconsistsgrp** command to prepare a FlashCopy Consistency Group. As with 9.13.5, “Preparing (pre-triggering) the FlashCopy mapping” on page 243, this command flushes the cache of any data that is destined for the source volume and forces the cache into the write-through mode until the mapping is started. The difference is that this command prepares a group of mappings (at a Consistency Group level) instead of one mapping.

When you have assigned several mappings to a FlashCopy Consistency Group, you only have to issue a single **prepare** command for the whole group to prepare all of the mappings at one time.

Example 9-135 shows how we prepare the Consistency Groups for DB and Log and check the result. After the command has executed all of the FlashCopy maps that we have, all of them are in the prepared status and all the Consistency Groups are in the prepared status, too. Now, we are ready to start the FlashCopy.

Example 9-135 Prepare a FlashCopy Consistency Group

```
IBM_2145:ITSO_SVC3:admin>prestartfcconsistgrp FCCG1
IBM_2145:ITSO_SVC3:admin>prestartfcconsistgrp FCCG2

IBM_2145:ITSO_SVC3:admin>lsfcconsistgrp FCCG1
id 1
name FCCG1
status prepared
autodelete off
FC_mapping_id 0
FC_mapping_name DB_Map1
FC_mapping_id 1
FC_mapping_name Log_Map1

IBM_2145:ITSO_SVC3:admin>lsfcconsistgrp
id name  status
1  FCCG1 prepared
2  FCCG2 prepared
```

9.13.7 Starting (triggering) FlashCopy mappings

The **startfcmap** command is used to start a single FlashCopy mapping. When invoked, a point-in-time copy of the source volume is created on the target volume.

When the FlashCopy mapping is triggered, it enters the Copying state. The way that the copy proceeds depends on the background copy rate attribute of the mapping. If the mapping is set to 0 (NOCOPY), only data that is subsequently updated on the source will be copied to the destination. We suggest that you use this scenario as a backup copy while the mapping exists in the Copying state. If the copy is stopped, the destination is unusable.

If you want to end up with a duplicate copy of the source at the destination, set the background copy rate greater than 0. This way, the system copies all of the data (even unchanged data) to the destination and eventually reaches the idle_or_copied state. After this data is copied, you can delete the mapping and have a usable point-in-time copy of the source at the destination.

In Example 9-136, after the FlashCopy is started, App_Map1 changes to copying status.

Example 9-136 Start App_Map1

```
IBM_2145:ITSO_SVC3:admin>startfcmap App_Map1
IBM_2145:ITSO_SVC3:admin>lsfcmap
id name     source_vdisk_id source_vdisk_name target_vdisk_id target_vdisk_name group_id
group_name status   progress copy_rate clean_progress incremental partner_FC_id
partner_FC_name restoring start_time   rc_controlled
0  DB_Map1  3               DB_Source         4               DB_Target1        1
FCCG1      prepared 0       50       0             off
no                    no
1  Log_Map1 6               Log_Source        7               Log_Target1       1
FCCG1      prepared 0       50       0             off
no                    no
2  App_Map1 9             App_Source       10           App_Target1
copying  0       50       100           off                                    no
110929113407 no
3  DB_Map2  3               DB_Source         5               DB_Target2        2
FCCG2      prepared 0       50       0             off
no                    no
4  Log_Map2 6               Log_Source        8               Log_Target2       2
FCCG2      prepared 0       50       0             off
no                    no
IBM_2145:ITSO_SVC3:admin>lsfcmap App_Map1
id 2
name App_Map1
source_vdisk_id 9
source_vdisk_name App_Source
target_vdisk_id 10
target_vdisk_name App_Target1
group_id
group_name
status copying
progress 0
copy_rate 50
start_time 110929113407
dependent_mappings 0
autodelete off
clean_progress 100
clean_rate 50
incremental off
```

```
difference 0
grain_size 256
IO_group_id 0
IO_group_name io_grp0
partner_FC_id
partner_FC_name
restoring no
rc_controlled no
```

9.13.8 Starting (triggering) FlashCopy Consistency Group

We execute the **startfcconsistgrp** command, as shown in Example 9-137, and afterward the database can be resumed. We created two point-in-time consistent copies of the DB and Log volumes. After the execution, the Consistency Group and the FlashCopy maps are all in the copying status.

Example 9-137 Start FlashCopy Consistency Group

```
IBM_2145:ITSO_SVC3:admin>startfcconsistgrp FCCG1
IBM_2145:ITSO_SVC3:admin>startfcconsistgrp FCCG2
IBM_2145:ITSO_SVC3:admin>lsfcconsistgrp FCCG1
id 1
name FCCG1
status copying
autodelete off
FC_mapping_id 0
FC_mapping_name DB_Map1
FC_mapping_id 1
FC_mapping_name Log_Map1
IBM_2145:ITSO_SVC3:admin>
IBM_2145:ITSO_SVC3:admin>lsfcconsistgrp
id name  status
1  FCCG1 copying
2  FCCG2 copying
```

9.13.9 Monitoring the FlashCopy progress

To monitor the background copy progress of the FlashCopy mappings, we issue the **lsfcmapprogress** command for each FlashCopy mapping.

Alternatively, you can also query the copy progress by using the **lsfcmap** command. As shown in Example 9-138, both DB_Map1 returns information that the background copy is 23% completed and Log_Map1 returns information that the background copy is 41% completed. DB_Map2 returns information that the background copy is 5% completed and Log_Map2 returns information that the background copy is 4% completed.

Example 9-138 Monitoring the background copy progress

```
IBM_2145:ITSO_SVC3:admin>lsfcmapprogress DB_Map1
id progress
0  23
IBM_2145:ITSO_SVC3:admin>lsfcmapprogress Log_Map1
id progress
1  41
IBM_2145:ITSO_SVC3:admin>lsfcmapprogress Log_Map2
```

```
id progress
4  4
IBM_2145:ITSO_SVC3:admin>lsfcmapprogress DB_Map2
id progress
3  5
IBM_2145:ITSO_SVC3:admin>lsfcmapprogress App_Map1
id progress
2  10
```

When the background copy has completed, the FlashCopy mapping enters the idle_or_copied state. When all FlashCopy mappings in a Consistency Group enter this status, the Consistency Group will be at idle_or_copied status.

When in this state, the FlashCopy mapping can be deleted and the target disk can be used independently if, for example, another target disk is to be used for the next FlashCopy of the particular source volume.

9.13.10 Stopping the FlashCopy mapping

The **stopfcmap** command is used to stop a FlashCopy mapping. This command allows you to stop an active (copying) or suspended mapping. When executed, this command stops a single FlashCopy mapping.

> **Tip:** In a Multiple Target FlashCopy environment, if you want to stop a mapping or group, consider whether you want to keep any of the dependent mappings. If not, issue the **stop** command with the **-force** parameter, which will stop all of the dependent maps and negate the need for the stopping copy process to run.

When a FlashCopy mapping is stopped, the target volume becomes invalid and is set offline by the SAN Volume Controller. The FlashCopy mapping needs to be prepared again or retriggered to bring the target volume online again.

> **Important:** Only stop a FlashCopy mapping when the data on the target volume is not in use, or when you want to modify the FlashCopy mapping. When a FlashCopy mapping is stopped, the target volume becomes invalid and is set offline by the SAN Volume Controller, if the mapping is in the Copying state and progress=100.

Example 9-139 shows how to stop the App_Map1 FlashCopy. The status of App_Map1 has changed to idle_or_copied.

Example 9-139 Stop App_Map1 FlashCopy

```
IBM_2145:ITSO_SVC3:admin>stopfcmap App_Map1

IBM_2145:ITSO_SVC3:admin>lsfcmap App_Map1
id 2
name App_Map1
source_vdisk_id 9
source_vdisk_name App_Source
target_vdisk_id 10
target_vdisk_name App_Target1
group_id
group_name
status idle_or_copied
progress 100
```

```
copy_rate 50
start_time 110929113407
dependent_mappings 0
autodelete off
clean_progress 100
clean_rate 50
incremental off
difference 100
grain_size 256
IO_group_id 0
IO_group_name io_grp0
partner_FC_id
partner_FC_name
restoring no
rc_controlled no
```

9.13.11 Stopping the FlashCopy Consistency Group

The **stopfcconsistgrp** command is used to stop any active FlashCopy Consistency Group. It stops all mappings in a Consistency Group. When a FlashCopy Consistency Group is stopped for all mappings that are not 100% copied, the target volumes become invalid and are set offline by the SAN Volume Controller. The FlashCopy Consistency Group needs to be prepared again and restarted to bring the target volumes online again.

> **Important:** Only stop a FlashCopy mapping when the data on the target volume is not in use, or when you want to modify the FlashCopy Consistency Group. When a Consistency Group is stopped, the target volume might become invalid and set offline by the SAN Volume Controller, depending on the state of the mapping.

As shown in Example 9-140, we stop the FCCG1 and FCCG2 Consistency Groups. The status of the two Consistency Groups has changed to stopped. Most of the FlashCopy mapping relationships now have the status of stopped. As you can see, several of them have already completed the copy operation and are now in a status of idle_or_copied.

Example 9-140 Stop FCCG1 and FCCG2 Consistency Groups

```
IBM_2145:ITSO_SVC3:admin>stopfcconsistgrp FCCG1

IBM_2145:ITSO_SVC3:admin>stopfcconsistgrp FCCG2

IBM_2145:ITSO_SVC3:admin>lsfcconsistgrp
id name  status
1  FCCG1 idle_or_copied
2  FCCG2 idle_or_copied

IBM_2145:ITSO_SVC3:admin>lsfcmap -delim ,
id,name,source_vdisk_id,source_vdisk_name,target_vdisk_id,target_vdisk_name,group_id,group_
name,status,progress,copy_rate,clean_progress,incremental,partner_FC_id,partner_FC_name,res
toring,start_time,rc_controlled
0,DB_Map1,3,DB_Source,4,DB_Target1,1,FCCG1,idle_or_copied,100,50,100,off,,,no,110929113806,
no
1,Log_Map1,6,Log_Source,7,Log_Target1,1,FCCG1,idle_or_copied,100,50,100,off,,,no,1109291138
06,no
2,App_Map1,9,App_Source,10,App_Target1,,,idle_or_copied,100,50,100,off,,,no,110929113407,no
3,DB_Map2,3,DB_Source,5,DB_Target2,2,FCCG2,idle_or_copied,100,50,100,off,,,no,110929113806,
no
```

```
4,Log_Map2,6,Log_Source,8,Log_Target2,2,FCCG2,idle_or_copied,100,50,100,off,,,no,1109291138
06,no
```

9.13.12 Deleting the FlashCopy mapping

To delete a FlashCopy mapping, use the **rmfcmap** command. When the command is executed, it attempts to delete the specified FlashCopy mapping. If the FlashCopy mapping is stopped, the command fails unless the **-force** flag is specified. If the mapping is active (copying), it must first be stopped before it can be deleted.

Deleting a mapping only deletes the logical relationship between the two volumes. However, when issued on an active FlashCopy mapping using the **-force** flag, the delete renders the data on the FlashCopy mapping target volume as inconsistent.

> **Tip:** If you want to use the target volume as a normal volume, monitor the background copy progress until it is complete (100% copied) and, then, delete the FlashCopy mapping. Another option is to set the **-autodelete** option when creating the FlashCopy mapping.

As shown in Example 9-141, we delete App_Map1.

Example 9-141 Delete App_Map1

```
IBM_2145:ITSO_SVC3:admin>rmfcmap App_Map1
```

9.13.13 Deleting the FlashCopy Consistency Group

The **rmfcconsistgrp** command is used to delete a FlashCopy Consistency Group. When executed, this command deletes the specified Consistency Group. If there are mappings that are members of the group, the command fails unless the **-force** flag is specified.

If you want to delete all of the mappings in the Consistency Group as well, first delete the mappings and then delete the Consistency Group.

As shown in Example 9-142, we delete all of the maps and Consistency Groups and then check the result.

Example 9-142 Remove fcmaps and fcconsistgrp

```
IBM_2145:ITSO_SVC3:admin>rmfcmap DB_Map1
IBM_2145:ITSO_SVC3:admin>rmfcmap DB_Map2
IBM_2145:ITSO_SVC3:admin>rmfcmap Log_Map1
IBM_2145:ITSO_SVC3:admin>rmfcmap Log_Map2
IBM_2145:ITSO_SVC3:admin>rmfcconsistgrp FCCG1
IBM_2145:ITSO_SVC3:admin>rmfcconsistgrp FCCG2
IBM_2145:ITSO_SVC3:admin>lsfcconsistgrp
IBM_2145:ITSO_SVC3:admin>lsfcmap
IBM_2145:ITSO_SVC3:admin>
```

9.13.14 Migrating a volume to a thin-provisioned volume

Use the following scenario to migrate a volume to a thin-provisioned volume:

1. Create a thin-provisioned space-efficient target volume with exactly the same size as the volume that you want to migrate.

   Example 9-143 shows the details of a volume with ID `11`. It has been created as a thin-provisioned volume with the same size as the `App_Source` volume.

   Example 9-143 lsvdisk 11 command

```
IBM_2145:ITSO_SVC3:admin>lsvdisk 11
id 11
name App_Source_SE
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 1
mdisk_grp_name Multi_Tier_Pool
capacity 10.00GB
type striped
formatted no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 60050768018281BEE00000000000000B
throttling 0
preferred_node_id 1
fast_write_state empty
cache readwrite
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 1
filesystem
mirror_write_priority latency

copy_id 0
status online
sync yes
primary yes
mdisk_grp_id 1
mdisk_grp_name Multi_Tier_Pool
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 0.41MB
real_capacity 221.17MB
free_capacity 220.77MB
overallocation 4629
autoexpand on
warning 80
grainsize 32
se_copy yes
easy_tier on
```

```
easy_tier_status active
tier generic_ssd
tier_capacity 0.00MB
tier generic_hdd
tier_capacity 221.17MB
```

2. Define a FlashCopy mapping in which the non-thin-provisioned volume is the source and the thin-provisioned volume is the target. Specify a copy rate as high as possible and activate the **-autodelete** option for the mapping. See Example 9-144.

Example 9-144 mkfcmap

```
IBM_2145:ITSO_SVC3:admin>mkfcmap -source App_Source -target App_Source_SE -name
MigrtoThinProv -copyrate 100 -autodelete
FlashCopy Mapping, id [0], successfully created
IBM_2145:ITSO_SVC3:admin>lsfcmap 0
id 0
name MigrtoThinProv
source_vdisk_id 9
source_vdisk_name App_Source
target_vdisk_id 11
target_vdisk_name App_Source_SE
group_id
group_name
status idle_or_copied
progress 0
copy_rate 100
start_time
dependent_mappings 0
autodelete on
clean_progress 100
clean_rate 50
incremental off
difference 100
grain_size 256
IO_group_id 0
IO_group_name io_grp0
partner_FC_id
partner_FC_name
restoring no
rc_controlled no
```

3. Run the **prestartfcmap** command and the **lsfcmap MigrtoThinProv** command, as shown in Example 9-145.

Example 9-145 prestartfcmap

```
IBM_2145:ITSO_SVC3:admin>prestartfcmap MigrtoThinProv
IBM_2145:ITSO_SVC3:admin>lsfcmap MigrtoThinProv
id 0
name MigrtoThinProv
source_vdisk_id 9
source_vdisk_name App_Source
target_vdisk_id 11
target_vdisk_name App_Source_SE
group_id
group_name
status prepared
progress 0
copy_rate 100
```

```
start_time
dependent_mappings 0
autodelete on
clean_progress 0
clean_rate 50
incremental off
difference 100
grain_size 256
IO_group_id 0
IO_group_name io_grp0
partner_FC_id
partner_FC_name
restoring no
rc_controlled no
```

4. Run the **startfcmap** command, as shown in Example 9-146.

Example 9-146 startfcmap command

```
IBM_2145:ITSO_SVC3:admin>startfcmap MigrtoThinProv
```

5. Monitor the copy process using the **lsfcmapprogress** command, as shown in Example 9-147.

Example 9-147 lsfcmapprogress command

```
IBM_2145:ITSO_SVC3:admin>lsfcmapprogress MigrtoThinProv
id progress
0  67
```

6. The FlashCopy mapping has been deleted automatically, as shown in Example 9-148.

Example 9-148 lsfcmap command

```
IBM_2145:ITSO_SVC3:admin>lsfcmap MigrtoThinProv
id 0
name MigrtoThinProv
source_vdisk_id 9
source_vdisk_name App_Source
target_vdisk_id 11
target_vdisk_name App_Source_SE
group_id
group_name
status copying
progress 67
copy_rate 100
start_time 110929135848
dependent_mappings 0
autodelete on
clean_progress 100
clean_rate 50
incremental off
difference 100
grain_size 256
IO_group_id 0
IO_group_name io_grp0
partner_FC_id
partner_FC_name
restoring no
rc_controlled no
```

```
IBM_2145:ITSO_SVC3:admin>lsfcmapprogress MigrtoThinProv
CMMVC5804E The action failed because an object that was specified in the command does
not exist.
IBM_2145:ITSO_SVC3:admin>
```

An independent copy of the source volume (`App_Source`) has been created. The migration has completed, as shown in Example 9-149.

Example 9-149 lsvdisk App_Source

```
IBM_2145:ITSO_SVC3:admin>lsvdisk App_Source
id 9
name App_Source
IO_group_id 0
IO_group_name io_grp0
status online
mdisk_grp_id 1
mdisk_grp_name Multi_Tier_Pool
capacity 10.00GB
type striped
formatted no
mdisk_id
mdisk_name
FC_id
FC_name
RC_id
RC_name
vdisk_UID 60050768018281BEE000000000000009
throttling 0
preferred_node_id 1
fast_write_state empty
cache readwrite
udid
fc_map_count 0
sync_rate 50
copy_count 1
se_copy_count 0
filesystem
mirror_write_priority latency

copy_id 0
status online
sync yes
primary yes
mdisk_grp_id 1
mdisk_grp_name Multi_Tier_Pool
type striped
mdisk_id
mdisk_name
fast_write_state empty
used_capacity 10.00GB
real_capacity 10.00GB
free_capacity 0.00MB
overallocation 100
autoexpand
warning
grainsize
se_copy no
easy_tier on
```

```
easy_tier_status active
tier generic_ssd
tier_capacity 0.00MB
tier generic_hdd
tier_capacity 10.00GB
```

Real size: Independently of what you defined as the real size of the target thin-provisioned volume, the real size will be at *least* the capacity of the source volume.

To migrate a thin-provisioned volume to a fully allocated volume, you can follow the same scenario.

9.13.15 Reverse FlashCopy

You can also have a reverse FlashCopy mapping without having to remove the original FlashCopy mapping, and without restarting a FlashCopy mapping from the beginning.

In Example 9-150, `FCMAP_1` is the forward FlashCopy mapping, and `FCMAP_rev_1` is a reverse FlashCopy mapping. We have also a cascade `FCMAP_2` where its source is `FCMAP_1`'s target volume, and its target is a separate volume named `Volume_FC_T1`.

In our example, after creating the environment, we started the `FCMAP_1` and later `FCMAP_2`.

As an example, we started `FCMAP_rev_1` without specifying the **`-restore`** parameter to show why we have to use it, and to show the message that is issued if you do not use it:

```
CMMVC6298E The command failed because a target VDisk has dependent FlashCopy
mappings.
```

When starting a reverse FlashCopy mapping, you must use the **`-restore`** option to indicate that you want to overwrite the data on the source disk of the forward mapping.

Example 9-150 Reverse FlashCopy

```
IBM_2145:ITSO_SVC3:admin>lsvdisk
id name           IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name  capacity
type    FC_id FC_name RC_id RC_name vdisk_UID                        fc_map_count
copy_count fast_write_state se_copy_count RC_change
3  Volume_FC_S    0           io_grp0       online 1            Multi_Tier_Pool 10.00GB
striped                            60050768018281BEE000000000000003 0           1
empty           0             0                     no
4  Volume_FC_T_S1 0           io_grp0       online 1            Multi_Tier_Pool 10.00GB
striped                            60050768018281BEE000000000000004 0           1
empty           0             0                     no
5  Volume_FC_T1   0           io_grp0       online 1            Multi_Tier_Pool 10.00GB
striped                            60050768018281BEE000000000000005 0           1
empty           0             0                     no

IBM_2145:ITSO_SVC3:admin>mkfcmap -source Volume_FC_S -target Volume_FC_T_S1 -name FCMAP_1
-copyrate 50
FlashCopy Mapping, id [0], successfully created

IBM_2145:ITSO_SVC3:admin>mkfcmap -source Volume_FC_T_S1 -target Volume_FC_S -name
FCMAP_rev_1 -copyrate 50
FlashCopy Mapping, id [1], successfully created

IBM_2145:ITSO_SVC3:admin>mkfcmap -source Volume_FC_T_S1 -target Volume_FC_T1 -name FCMAP_2
-copyrate 50
```

```
FlashCopy Mapping, id [2], successfully created

IBM_2145:ITSO_SVC3:admin>lsfcmap
id name       source_vdisk_id source_vdisk_name target_vdisk_id target_vdisk_name group_id
group_name status         progress copy_rate clean_progress incremental partner_FC_id
partner_FC_name restoring start_time rc_controlled
0  FCMAP_1    3            Volume_FC_S      4            Volume_FC_T_S1
idle_or_copied 0        50        100            off         1            FCMAP_rev_1
no                 no
1  FCMAP_rev_1 4            Volume_FC_T_S1   3            Volume_FC_S
idle_or_copied 0        50        100            off         0            FCMAP_1
no                 no
2  FCMAP_2    4            Volume_FC_T_S1   5            Volume_FC_T1
idle_or_copied 0        50        100            off
no                 no

IBM_2145:ITSO_SVC3:admin>startfcmap -prep FCMAP_1

IBM_2145:ITSO_SVC3:admin>startfcmap -prep FCMAP_2

IBM_2145:ITSO_SVC3:admin>lsfcmap
id name       source_vdisk_id source_vdisk_name target_vdisk_id target_vdisk_name group_id
group_name status         progress copy_rate clean_progress incremental partner_FC_id
partner_FC_name restoring start_time   rc_controlled
0  FCMAP_1    3             Volume_FC_S      4             Volume_FC_T_S1 copying 0
50        100           off         1            FCMAP_rev_1     no
no
1  FCMAP_rev_1 4            Volume_FC_T_S1   3            Volume_FC_S
idle_or_copied 0        50        100            off         0            FCMAP_1
no                    no
2  FCMAP_2    4            Volume_FC_T_S1   5            Volume_FC_T1
copying       4        50        100            off
no        110929143739 no

IBM_2145:ITSO_SVC3:admin>startfcmap -prep FCMAP_rev_1
CMMVC6298E The command failed because a target VDisk has dependent FlashCopy mappings.

IBM_2145:ITSO_SVC3:admin>startfcmap -prep -restore FCMAP_rev_1

IBM_2145:ITSO_SVC3:admin>lsfcmap
id name       source_vdisk_id source_vdisk_name target_vdisk_id target_vdisk_name group_id
group_name status  progress copy_rate clean_progress incremental partner_FC_id
partner_FC_name restoring start_time   rc_controlled
0  FCMAP_1    3            Volume_FC_S      4            Volume_FC_T_S1
copying 43      100       56            off         1            FCMAP_rev_1     no
110929151911 no
1  FCMAP_rev_1 4            Volume_FC_T_S1   3            Volume_FC_S
copying 56      100       43            off         0            FCMAP_1         yes
110929152030 no
2  FCMAP_2    4            Volume_FC_T_S1   5            Volume_FC_T1
copying 37      100       100           off                                      no
110929151926 no
```

As you can see in Example 9-150 on page 254, `FCMAP_rev_1` shows a restoring value of `yes` while the FlashCopy mapping is copying. After it has finished copying, the restoring value field will change to `no`.

9.13.16 Split-stopping of FlashCopy maps

The **stopfcmap** command has a **-split** option. This option allows the source target of a map, which is 100% complete, to be removed from the head of a cascade when the map is stopped.

For example, if we have four volumes in a cascade (A → B → C → D), and the map A → B is 100% complete, using the **stopfcmap -split mapAB** command results in mapAB becoming idle_copied and the remaining cascade becoming B → C → D.

Without the **-split** option, volume A remains at the head of the cascade (A → C → D). Consider this sequence of steps:

1. User takes a backup using the mapping A → B. A is the production volume; B is a backup.
2. At a later point, the user experiences corruption on A and so reverses the mapping to B → A.
3. The user then takes another backup from the production disk A, resulting in the cascade B → A → C.

Stopping A → B without the **-split** option results in the cascade B → C. Note that the backup disk B is now at the head of this cascade.

When the user next wants to take a backup to B, the user can still start mapping A → B (using the **-restore** flag), but the user cannot then reverse the mapping to A (B → A or C → A).

Stopping A → B with the **-split** option results in the cascade A → C. This action does not result in the same problem, because the production disk A is at the head of the cascade instead of the backup disk B.

9.14 Metro Mirror operation

Intercluster example: This example is for intercluster operations only.

If you want to set up intracluster operations, we highlight those parts of the following procedure that you do *not* need to perform.

In the following scenario, we set up an intercluster Metro Mirror relationship between the SAN Volume Controller system ITSO_SVC1 primary site and the SAN Volume Controller system ITSO_SVC4 at the secondary site. Table 9-3 shows the details of the volumes.

Table 9-3 Volume details

Content of volume	Volumes at primary site	Volumes at secondary site
Database files	MM_DB_Pri	MM_DB_Sec
Database log files	MM_DBLog_Pri	MM_DBLog_Sec
Application files	MM_App_Pri	MM_App_Sec

Because data consistency is needed across the MM_DB_Pri and MM_DBLog_Pri volumes, a CG_WIN2K3_MM Consistency Group is created to handle Metro Mirror relationships for them.

Because, in this scenario, application files are independent of the database, a stand-alone Metro Mirror relationship is created for the MM_App_Pri volume. Figure 9-7 illustrates the Metro Mirror setup.

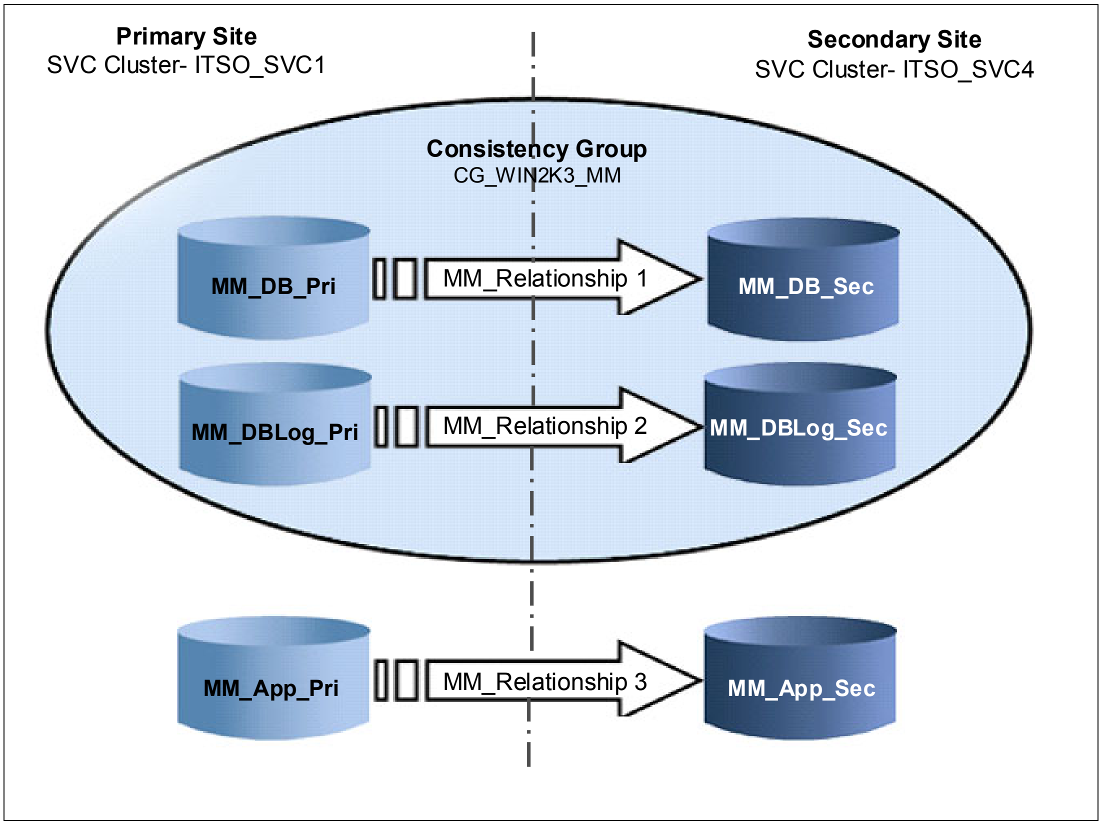


Figure 9-7 Metro Mirror scenario

9.14.1 Setting up Metro Mirror

In the following section, we assume that the source and target volumes have already been created and that the inter-switch links (ISLs) and zoning are in place, enabling the SAN Volume Controller clustered systems to communicate.

To set up the Metro Mirror, perform the following steps:

1. Create a SAN Volume Controller partnership between ITSO_SVC1 and ITSO_SVC4 on both of the SAN Volume Controller clustered systems.
2. Create a Metro Mirror Consistency Group:

   Name: CG_W2K3_MM
3. Create the Metro Mirror relationship for MM_DB_Pri:
   - Master: MM_DB_Pri
   - Auxiliary: MM_DB_Sec
   - Auxiliary SAN Volume Controller system: ITSO_SVC4
   - Name: MMREL1
   - Consistency Group: CG_W2K3_MM
4. Create the Metro Mirror relationship for MM_DBLog_Pri:
   - Master: MM_DBLog_Pri
   - Auxiliary: MM_DBLog_Sec

– Auxiliary SAN Volume Controller system: ITSO_SVC4
– Name: MMREL2
– Consistency Group: CG_W2K3_MM

5. Create the Metro Mirror relationship for MM_App_Pri:
   – Master: MM_App_Pri
   – Auxiliary: MM_App_Sec
   – Auxiliary SAN Volume Controller system: ITSO_SVC4
   – Name: MMREL3

In the following section, we perform each step by using the CLI.

9.14.2 Creating a SAN Volume Controller partnership between ITSO_SVC1 and ITSO_SVC4

We create the SAN Volume Controller partnership on both systems.

> **Intracluster Metro Mirror:** If you are creating an intracluster Metro Mirror, do *not* perform the next step; instead, go to 9.14.3, "Creating a Metro Mirror Consistency Group" on page 261.

Preverification

To verify that both systems can communicate with each other, use the **lspartnershipcandidate** command.

As shown in Example 9-151, ITSO_SVC4 is an eligible SAN Volume Controller system candidate at ITSO_SVC1 for the SAN Volume Controller system partnership, and vice versa. Therefore, both systems communicate with each other.

Example 9-151 Listing the available SAN Volume Controller systems for partnership

```
IBM_2145:ITSO_SVC1:admin>lspartnershipcandidate
id               configured name
0000020061C06FCA no         ITSO_SVC4
000002006AC03A42 no         ITSO_SVC2
0000020060A06FB8 no         ITSO_SVC3
00000200A0C006B2 no         ITSO-Storwize-V7000-2

IBM_2145:ITSO_SVC4:admin>lspartnershipcandidate
id               configured name
000002006AC03A42 no         ITSO_SVC2
0000020060A06FB8 no         ITSO_SVC3
00000200A0C006B2 no         ITSO-Storwize-V7000-2
000002006BE04FC4 no         ITSO_SVC1
```

Example 9-152 on page 259 shows the output of the **lspartnership** and **lssystem** commands before setting up the Metro Mirror relationship. We show them so that you can compare with the same relationship *after* setting up the Metro Mirror relationship.

As of SAN Volume Controller 6.3, you can create a partnership between the SAN Volume Controller system and the IBM Storwize V7000 system. Be aware that to create this partnership, you need to change the layer parameter on the IBM Storwize V7000 system. It must be changed from storage to replication with the **chsystem** command.

This parameter cannot be changed on the SAN Volume Controller system. It is fixed to the value of `appliance`, as shown in Example 9-152.

Example 9-152 Pre-verification of system configuration

```
IBM_2145:ITSO_SVC1:admin>lspartnership
id               name      location partnership bandwidth
000002006BE04FC4 ITSO_SVC1 local

IBM_2145:ITSO_SVC4:admin>lspartnership
id               name      location partnership bandwidth
0000020061C06FCA ITSO_SVC4 local

IBM_2145:ITSO_SVC1:admin>lssystem
id 000002006BE04FC4
name ITSO_SVC1
location local
partnership
bandwidth
total_mdisk_capacity 766.5GB
space_in_mdisk_grps 766.5GB
space_allocated_to_vdisks 0.00MB
total_free_space 766.5GB
total_vdiskcopy_capacity 0.00MB
total_used_capacity 0.00MB
total_overallocation 0
total_vdisk_capacity 0.00MB
total_allocated_extent_capacity 1.50GB
statistics_status on
statistics_frequency 15
cluster_locale en_US
time_zone 520 US/Pacific
code_level 6.3.0.0 (build 54.0.1109090000)
console_IP 10.18.228.81:443
id_alias 000002006BE04FC4
gm_link_tolerance 300
gm_inter_cluster_delay_simulation 0
gm_intra_cluster_delay_simulation 0
gm_max_host_delay 5
email_reply
email_contact
email_contact_primary
email_contact_alternate
email_contact_location
email_contact2
email_contact2_primary
email_contact2_alternate
email_state stopped
inventory_mail_interval 0
cluster_ntp_IP_address
cluster_isns_IP_address
iscsi_auth_method chap
iscsi_chap_secret passw0rd
auth_service_configured no
auth_service_enabled no
auth_service_url
auth_service_user_name
auth_service_pwd_set no
auth_service_cert_set no
auth_service_type tip
relationship_bandwidth_limit 25
```

```
tier generic_ssd
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier generic_hdd
tier_capacity 766.50GB
tier_free_capacity 766.50GB
has_nas_key no
layer appliance

IBM_2145:ITSO_SVC4:admin>lssystem
id 0000020061C06FCA
name ITSO_SVC4
location local
partnership
bandwidth
total_mdisk_capacity 768.0GB
space_in_mdisk_grps 0
space_allocated_to_vdisks 0.00MB
total_free_space 768.0GB
total_vdiskcopy_capacity 0.00MB
total_used_capacity 0.00MB
total_overallocation 0
total_vdisk_capacity 0.00MB
total_allocated_extent_capacity 0.00MB
statistics_status on
statistics_frequency 15
cluster_locale en_US
time_zone 520 US/Pacific
code_level 6.3.0.0 (build 54.0.1109090000)
console_IP 10.18.228.84:443
id_alias 0000020061C06FCA
gm_link_tolerance 300
gm_inter_cluster_delay_simulation 0
gm_intra_cluster_delay_simulation 0
gm_max_host_delay 5
email_reply
email_contact
email_contact_primary
email_contact_alternate
email_contact_location
email_contact2
email_contact2_primary
email_contact2_alternate
email_state stopped
inventory_mail_interval 0
cluster_ntp_IP_address
cluster_isns_IP_address
iscsi_auth_method none
iscsi_chap_secret
auth_service_configured no
auth_service_enabled no
auth_service_url
auth_service_user_name
auth_service_pwd_set no
auth_service_cert_set no
auth_service_type tip
relationship_bandwidth_limit 25
tier generic_ssd
tier_capacity 0.00MB
tier_free_capacity 0.00MB
```

```
tier generic_hdd
tier_capacity 0.00MB
tier_free_capacity 0.00MB
has_nas_key no
layer appliance
```

Partnership between clustered systems

In Example 9-153, a partnership is created between `ITSO_SVC1` and `ITSO_SVC4`, specifying 50 MBps bandwidth to be used for the background copy.

To check the status of the newly created partnership, issue the **`lspartnership`** command. Also, notice that the new partnership is only *partially* configured. It remains partially configured until the Metro Mirror relationship is created on the other node.

Example 9-153 Creating the partnership from ITSO_SVC1 to ITSO_SVC4 and verifying it

```
IBM_2145:ITSO_SVC1:admin>mkpartnership -bandwidth 50 ITSO_SVC4
IBM_2145:ITSO_SVC1:admin>lspartnership
id               name      location partnership               bandwidth
000002006BE04FC4 ITSO_SVC1 local
0000020061C06FCA ITSO_SVC4 remote   partially_configured_local 50
```

In Example 9-154, the partnership is created between `ITSO_SVC4` back to `ITSO_SVC1`, specifying the bandwidth to be used for a background copy of 50 MBps.

After creating the partnership, verify that the partnership is *fully* configured on both systems by reissuing the **`lspartnership`** command.

Example 9-154 Creating the partnership from ITSO_SVC4 to ITSO_SVC1 and verifying it

```
IBM_2145:ITSO_SVC4:admin>mkpartnership -bandwidth 50 ITSO_SVC1
IBM_2145:ITSO_SVC4:admin>lspartnership
id               name      location partnership      bandwidth
0000020061C06FCA ITSO_SVC4 local
000002006BE04FC4 ITSO_SVC1 remote   fully_configured 50
```

9.14.3 Creating a Metro Mirror Consistency Group

In Example 9-155, we create the Metro Mirror Consistency Group using the **`mkrcconsistgrp`** command. This Consistency Group will be used for the Metro Mirror relationships of the database volumes named `MM_DB_Pri` and `MM_DBLog_Pri`. The Consistency Group is named `CG_W2K3_MM`.

Example 9-155 Creating the Metro Mirror Consistency Group CG_W2K3_MM

```
IBM_2145:ITSO_SVC1:admin>mkrcconsistgrp -cluster ITSO_SVC4 -name CG_W2K3_MM
RC Consistency Group, id [0], successfully created

IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp
id name       master_cluster_id master_cluster_name aux_cluster_id
aux_cluster_name primary state relationship_count copy_type   cycling_mode
0  CG_W2K3_MM 000002006BE04FC4  ITSO_SVC1       0000020061C06FCA ITSO_SVC4
empty 0                  empty_group none
```

9.14.4 Creating the Metro Mirror relationships

In Example 9-156, we create the Metro Mirror relationships MMREL1 and MMREL2 for MM_DB_Pri and MM_DBLog_Pri. Also, we make them members of the Metro Mirror Consistency Group CG_W2K3_MM. We use the **lsvdisk** command to list all of the volumes in the ITSO_SVC1 system. We then use the **lsrcrelationshipcandidate** command to show the volumes in the ITSO_SVC4 system.

By using this command, we check the possible candidates for MM_DB_Pri. After checking all of these conditions, we use the **mkrcrelationship** command to create the Metro Mirror relationship.

To verify the newly created Metro Mirror relationships, list them with the **lsrcrelationship** command.

Example 9-156 Creating Metro Mirror relationships MMREL1 and MMREL2

```
IBM_2145:ITSO_SVC1:admin>lsvdisk -filtervalue name=MM*
id name        IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name capacity type    FC_id FC_name
RC_id RC_name vdisk_UID                        fc_map_count copy_count fast_write_state se_copy_count
RC_change
0  MM_DB_Pri    0         io_grp0      online 0         Pool_DS3500-1  10.00GB  striped
6005076801AF813F1000000000000031 0          1         empty           0           0
no
1  MM_DBLog_Pri 0         io_grp0      online 0         Pool_DS3500-1  10.00GB  striped
6005076801AF813F1000000000000032 0          1         empty           0           0
no
2  MM_App_Pri   0         io_grp0      online 0         Pool_DS3500-1  10.00GB  striped
6005076801AF813F1000000000000033 0          1         empty           0           0
no
IBM_2145:ITSO_SVC1:admin>lsrcrelationshipcandidate
id vdisk_name
0  MM_DB_Pri
1  MM_DBLog_Pri
2  MM_App_Pri

IBM_2145:ITSO_SVC1:admin>lsrcrelationshipcandidate -aux ITSO_SVC4 -master MM_DB_Pri
id vdisk_name
0  MM_DB_Sec
1  MM_DBLog_Sec
2  MM_App_Sec

IBM_2145:ITSO_SVC1:admin>mkrcrelationship -master MM_DB_Pri -aux MM_DB_Sec -cluster ITSO_SVC4 -consistgrp
CG_W2K3_MM -name MMREL1
RC Relationship, id [0], successfully created
IBM_2145:ITSO_SVC1:admin>mkrcrelationship -master MM_Log_Pri -aux MM_Log_Sec -cluster ITSO_SVC4 -consistgrp
CG_W2K3_MM -name MMREL2
RC Relationship, id [3], successfully created

IBM_2145:ITSO_SVC1:admin>lsrcrelationship
id name   master_cluster_id master_cluster_name master_vdisk_id master_vdisk_name aux_cluster_id
aux_cluster_name aux_vdisk_id aux_vdisk_name primary consistency_group_id consistency_group_name state
bg_copy_priority progress copy_type cycling_mode
0  MMREL1 000002006BE04FC4  ITSO_SVC1          0               MM_DB_Pri         0000020061C06FCA
ITSO_SVC4       0            MM_DB_Sec      master  0                    CG_W2K3_MM
inconsistent_stopped 50              0       metro     none
3  MMREL2 000002006BE04FC4  ITSO_SVC1          3               MM_Log_Pri
0000020061C06FCA ITSO_SVC4       3            MM_Log_Sec     master  0
CG_W2K3_MM             inconsistent_stopped 50               0        metro     none
```

9.14.5 Creating a stand-alone Metro Mirror relationship for MM_App_Pri

In Example 9-157, we create the stand-alone Metro Mirror relationship MMREL3 for MM_App_Pri. After it is created, we check the status of this Metro Mirror relationship.

Notice that the state of MMREL3 is consistent_stopped. MMREL3 is in this state because it was created with the **-sync** option. The **-sync** option indicates that the secondary (auxiliary) volume is already synchronized with the primary (master) volume. Initial background synchronization is skipped when this option is used, even though the volumes are not actually synchronized in this scenario.

We want to illustrate the option of pre-synchronized master and auxiliary volumes before setting up the relationship. We have created the new relationship for MM_App_Sec using the **-sync** option.

> **Tip:** The **-sync** option is only used when the target volume has already mirrored all of the data from the source volume. By using this option, there is no initial background copy between the primary volume and the secondary volume.

MMREL2 and MMREL1 are in the inconsistent_stopped state because they were not created with the **-sync** option. Therefore, their auxiliary volumes need to be synchronized with their primary volumes.

Example 9-157 Creating a stand-alone relationship and verifying it

```
IBM_2145:ITSO_SVC1:admin>mkrcrelationship -master MM_App_Pri -aux MM_App_Sec -sync -cluster
ITSO_SVC4 -name MMREL3
RC Relationship, id [2], successfully created

IBM_2145:ITSO_SVC1:admin>lsrcrelationship 2
id 2
name MMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name MM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name MM_App_Sec
primary master
consistency_group_id
consistency_group_name
state consistent_stopped
bg_copy_priority 50
progress 100
freeze_time
status online
sync in_sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name
```

9.14.6 Starting Metro Mirror

Now that the Metro Mirror Consistency Group and relationships are in place, we are ready to use Metro Mirror relationships in our environment.

When implementing Metro Mirror, the goal is to reach a consistent and synchronized state that can provide redundancy for a data set if a failure occurs that affects the production site.

In the following section, we show how to stop and start stand-alone Metro Mirror relationships and Consistency Groups.

Starting a stand-alone Metro Mirror relationship

In Example 9-158, we start a stand-alone Metro Mirror relationship named `MMREL3`. Because the Metro Mirror relationship was in the Consistent stopped state and no updates have been made to the primary volume, the relationship quickly enters the Consistent synchronized state.

Example 9-158 Starting the stand-alone Metro Mirror relationship

```
IBM_2145:ITSO_SVC1:admin>startrcrelationship MMREL3
IBM_2145:ITSO_SVC1:admin>lsrcrelationship MMREL3
id 2
name MMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name MM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name MM_App_Sec
primary master
consistency_group_id
consistency_group_name
state consistent_synchronized
bg_copy_priority 50
progress
freeze_time
status online
sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name
```

9.14.7 Starting a Metro Mirror Consistency Group

In Example 9-159 on page 265, we start the Metro Mirror Consistency Group `CG_W2K3_MM`. Because the Consistency Group was in the Inconsistent stopped state, it enters the Inconsistent copying state until the background copy has completed for all of the relationships in the Consistency Group.

Upon completion of the background copy, it enters the Consistent synchronized state.

Example 9-159 Starting the Metro Mirror Consistency Group

```
IBM_2145:ITSO_SVC1:admin>startrcconsistgrp CG_W2K3_MM
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp
id name       master_cluster_id master_cluster_name aux_cluster_id   aux_cluster_name
primary state                relationship_count copy_type cycling_mode
0  CG_W2K3_MM 000002006BE04FC4  ITSO_SVC1           0000020061C06FCA ITSO_SVC4
master  inconsistent_copying 2                  metro     none
```

9.14.8 Monitoring the background copy progress

To monitor the background copy progress, we can use the **lsrcrelationship** command. This command shows all of the defined Metro Mirror relationships if it is used without any arguments. In the command output, `progress` indicates the current background copy progress. Our Metro Mirror relationship is shown in Example 9-160.

> **Using SNMP traps:** Setting up SNMP traps for the SAN Volume Controller enables automatic notification when Metro Mirror Consistency Groups or relationships change state.

Example 9-160 Monitoring the background copy progress example

```
IBM_2145:ITSO_SVC1:admin>lsrcrelationship MMREL1
id 0
name MMREL1
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 0
master_vdisk_name MM_DB_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 0
aux_vdisk_name MM_DB_Sec
primary master
consistency_group_id 0
consistency_group_name CG_W2K3_MM
state inconsistent_copying
bg_copy_priority 50
progress 81
freeze_time
status online
sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name

IBM_2145:ITSO_SVC1:admin>lsrcrelationship MMREL2
id 3
name MMREL2
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 3
master_vdisk_name MM_Log_Pri
aux_cluster_id 0000020061C06FCA
```

```
aux_cluster_name ITSO_SVC4
aux_vdisk_id 3
aux_vdisk_name MM_Log_Sec
primary master
consistency_group_id 0
consistency_group_name CG_W2K3_MM
state inconsistent_copying
bg_copy_priority 50
progress 82
freeze_time
status online
sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name
```

When all Metro Mirror relationships have completed the background copy, the Consistency Group enters the Consistent synchronized state, as shown in Example 9-161.

Example 9-161 Listing the Metro Mirror Consistency Group

```
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_MM
id 0
name CG_W2K3_MM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary master
state consistent_synchronized
relationship_count 2
freeze_time
status
sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name MMREL1
RC_rel_id 3
RC_rel_name MMREL2
```

9.14.9 Stopping and restarting Metro Mirror

Now that the Metro Mirror Consistency Group and relationships are running, in this section and in the following sections, we describe how to stop, restart, and change the direction of the stand-alone Metro Mirror relationships and the Consistency Group.

9.14.10 Stopping a stand-alone Metro Mirror relationship

Example 9-162 on page 267 shows how to stop the stand-alone Metro Mirror relationship, while enabling access (write I/O) to both the primary and secondary volumes. It also shows the relationship entering the Idling state.

Example 9-162 Stopping stand-alone Metro Mirror relationship and enabling access to the secondary

```
IBM_2145:ITSO_SVC1:admin>stoprcrelationship -access MMREL3
IBM_2145:ITSO_SVC1:admin>lsrcrelationship MMREL3
id 2
name MMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name MM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name MM_App_Sec
primary
consistency_group_id
consistency_group_name
state idling
bg_copy_priority 50
progress
freeze_time
status
sync in_sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name
```

9.14.11 Stopping a Metro Mirror Consistency Group

Example 9-163 shows how to stop the Metro Mirror Consistency Group without specifying the **-access** flag. The Consistency Group enters the Consistent stopped state.

Example 9-163 Stopping a Metro Mirror Consistency Group

```
IBM_2145:ITSO_SVC1:admin>stoprcconsistgrp CG_W2K3_MM
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_MM
id 0
name CG_W2K3_MM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary master
state consistent_stopped
relationship_count 2
freeze_time
status
sync in_sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name MMREL1
RC_rel_id 3
RC_rel_name MMREL2
```

If, afterward, we want to enable access (write I/O) to the secondary volume, we reissue the **stoprcconsistgrp** command, specifying the **-access** flag. The Consistency Group transits to the Idling state, as shown in Example 9-164.

Example 9-164 Stopping a Metro Mirror Consistency Group and enabling access to the secondary

```
IBM_2145:ITSO_SVC1:admin>stoprcconsistgrp -access CG_W2K3_MM
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_MM
id 0
name CG_W2K3_MM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary
state idling
relationship_count 2
freeze_time
status
sync in_sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name MMREL1
RC_rel_id 3
RC_rel_name MMREL2
```

9.14.12 Restarting a Metro Mirror relationship in the Idling state

When restarting a Metro Mirror relationship in the Idling state, we must specify the copy direction.

If any updates have been performed on either the master or the auxiliary volume, consistency will be compromised. Therefore, we must issue the command with the **-force** flag to restart a relationship, as shown in Example 9-165.

Example 9-165 Restarting a Metro Mirror relationship after updates in the Idling state

```
IBM_2145:ITSO_SVC1:admin>startrcrelationship -primary master -force MMREL3
IBM_2145:ITSO_SVC1:admin>lsrcrelationship MMREL3
id 2
name MMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name MM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name MM_App_Sec
primary master
consistency_group_id
consistency_group_name
state consistent_synchronized
bg_copy_priority 50
progress
freeze_time
status online
sync
```

```
copy_type metro
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name
```

9.14.13 Restarting a Metro Mirror Consistency Group in the Idling state

When restarting a Metro Mirror Consistency Group in the Idling state, we must specify the copy direction.

If any updates have been performed on either the master or the auxiliary volume in any of the Metro Mirror relationships in the Consistency Group, the consistency is compromised. Therefore, we must use the **-force** flag to start a relationship. If the **-force** flag is not used, the command fails.

In Example 9-166, we change the copy direction by specifying the auxiliary volumes to become the primaries.

Example 9-166 Restarting a Metro Mirror relationship while changing the copy direction

```
IBM_2145:ITSO_SVC1:admin>startrcconsistgrp -force -primary aux CG_W2K3_MM
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_MM
id 0
name CG_W2K3_MM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary aux
state consistent_synchronized
relationship_count 2
freeze_time
status
sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name MMREL1
RC_rel_id 3
RC_rel_name MMREL2
```

9.14.14 Changing the copy direction for Metro Mirror

In this section, we show how to change the copy direction of the stand-alone Metro Mirror relationship and the Consistency Group.

9.14.15 Switching the copy direction for a Metro Mirror relationship

When a Metro Mirror relationship is in the Consistent synchronized state, we can change the copy direction for the relationship using the **switchrcrelationship** command, specifying the primary volume. If the specified volume is already a primary when you issue this command, the command has no effect.

In Example 9-167, we change the copy direction for the stand-alone Metro Mirror relationship by specifying the auxiliary volume to become the primary.

> **Important:** When the copy direction is switched, it is crucial that there is no outstanding I/O to the volume that transitions from the primary to the secondary because all of the I/O will be inhibited to that volume when it becomes the secondary. Therefore, careful planning is required before using the **`switchrcrelationship`** command.

Example 9-167 Switching the copy direction for a Metro Mirror Consistency Group

```
IBM_2145:ITSO_SVC1:admin>lsrcrelationship MMREL3
id 2
name MMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name MM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name MM_App_Sec
primary master
consistency_group_id
consistency_group_name
state consistent_synchronized
bg_copy_priority 50
progress
freeze_time
status online
sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name
IBM_2145:ITSO_SVC1:admin>switchrcrelationship -primary aux MMREL3
IBM_2145:ITSO_SVC1:admin>lsrcrelationship MMREL3
id 2
name MMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name MM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name MM_App_Sec
primary aux
consistency_group_id
consistency_group_name
state consistent_synchronized
bg_copy_priority 50
progress
freeze_time
```

```
status online
sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name
```

9.14.16 Switching the copy direction for a Metro Mirror Consistency Group

When a Metro Mirror Consistency Group is in the Consistent synchronized state, we can change the copy direction for the Consistency Group by using the **switchrcconsistgrp** command and specifying the primary volume.

If the specified volume is already a primary when you issue this command, the command has no effect.

In Example 9-168, we change the copy direction for the Metro Mirror Consistency Group by specifying the auxiliary volume to become the primary volume.

> **Important:** When the copy direction is switched, it is crucial that there is no outstanding I/O to the volume that transitions from primary to secondary because all of the I/O will be inhibited when that volume becomes the secondary. Therefore, careful planning is required before using the **switchrcconsistgrp** command.

Example 9-168 Switching the copy direction for a Metro Mirror Consistency Group

```
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_MM
id 0
name CG_W2K3_MM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary master
state consistent_synchronized
relationship_count 2
freeze_time
status
sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name MMREL1
RC_rel_id 3
RC_rel_name MMREL2
IBM_2145:ITSO_SVC1:admin>switchrcconsistgrp -primary aux CG_W2K3_MM
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_MM
id 0
name CG_W2K3_MM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
```

```
primary aux
state consistent_synchronized
relationship_count 2
freeze_time
status
sync
copy_type metro
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name MMREL1
RC_rel_id 3
RC_rel_name MMREL2
```

9.14.17 Creating a SAN Volume Controller partnership among many clustered systems

Starting with SAN Volume Controller 5.1, you can have a clustered system partnership among many SAN Volume Controller systems. This capability allows you to create four configurations using a maximum of four connected systems:

- Star configuration
- Triangle configuration
- Fully connected configuration
- Daisy-chain configuration

In this section, we describe how to configure the SAN Volume Controller system partnership for each configuration.

> **Important:** To have a supported and working configuration, all SAN Volume Controller systems must be at level 5.1 or higher.

In our scenarios, we configure the SAN Volume Controller partnership by referring to the clustered systems as A, B, C, and D:

- `ITSO_SVC1` = A
- `ITSO_SVC2` = B
- `ITSO_SVC3` = C
- `ITSO_SVC4` = D

Example 9-169 shows the available systems for a partnership using the **`lsclustercandidate`** command on each system.

Example 9-169 Available clustered systems

```
IBM_2145:ITSO_SVC1:admin>lspartnershipcandidate
id               configured name
0000020061C06FCA no         ITSO_SVC4
0000020060A06FB8 no         ITSO_SVC3
000002006AC03A42 no         ITSO_SVC2

IBM_2145:ITSO_SVC2:admin>lspartnershipcandidate
id               configured name
0000020061C06FCA no         ITSO_SVC4
000002006BE04FC4 no         ITSO_SVC1
0000020060A06FB8 no         ITSO_SVC3
```

```
IBM_2145:ITSO_SVC3:admin>lspartnershipcandidate
id               configured name
000002006BE04FC4 no         ITSO_SVC1
0000020061C06FCA no         ITSO_SVC4
000002006AC03A42 no         ITSO_SVC2

IBM_2145:ITSO_SVC4:admin>lspartnershipcandidate
id               configured name
000002006BE04FC4 no         ITSO_SVC1
0000020060A06FB8 no         ITSO_SVC3
000002006AC03A42 no         ITSO_SVC2
```

9.14.18 Star configuration partnership

Figure 9-8 shows the star configuration.

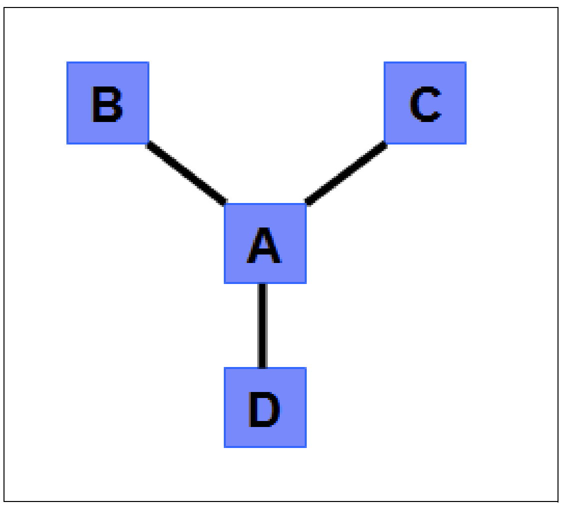


Figure 9-8 Star configuration

Example 9-170 shows the sequence of **mkpartnership** commands to execute to create a star configuration.

Example 9-170 Creating a star configuration using the mkpartnership command

```
From ITSO_SVC1 to multiple systems

IBM_2145:ITSO_SVC1:admin>mkpartnership -bandwidth 50 ITSO_SVC2
IBM_2145:ITSO_SVC1:admin>mkpartnership -bandwidth 50 ITSO_SVC3
IBM_2145:ITSO_SVC1:admin>mkpartnership -bandwidth 50 ITSO_SVC4

From ITSO_SVC2 to ITSO_SVC1

IBM_2145:ITSO_SVC2:admin>mkpartnership -bandwidth 50 ITSO_SVC1

From ITSO_SVC3 to ITSO_SVC1

IBM_2145:ITSO_SVC3:admin>mkpartnership -bandwidth 50 ITSO_SVC1

From ITSO_SVC4 to ITSO_SVC1
```

```
IBM_2145:ITSO_SVC4:admin>mkpartnership -bandwidth 50 ITSO_SVC1
```

From ITSO_SVC1

```
IBM_2145:ITSO_SVC1:admin>lspartnership
id               name      location partnership      bandwidth
000002006BE04FC4 ITSO_SVC1 local
000002006AC03A42 ITSO_SVC2 remote   fully_configured 50
0000020060A06FB8 ITSO_SVC3 remote   fully_configured 50
0000020061C06FCA ITSO_SVC4 remote   fully_configured 50
```

From ITSO_SVC2

```
IBM_2145:ITSO_SVC2:admin>lspartnership
id               name            location partnership              bandwidth
000002006AC03A42 ITSO_SVC2       local
000002006BE04FC4 ITSO_SVC1       remote   fully_configured         50
```

From ITSO_SVC3

```
IBM_2145:ITSO_SVC3:admin>lspartnership
id               name      location partnership      bandwidth
0000020060A06FB8 ITSO_SVC3 local
000002006BE04FC4 ITSO_SVC1 remote   fully_configured 50
```

From ITSO_SVC4

```
IBM_2145:ITSO_SVC4:admin>lspartnership
id               name      location partnership      bandwidth
0000020061C06FCA ITSO_SVC4 local
000002006BE04FC4 ITSO_SVC1 remote   fully_configured 50
```

After the SAN Volume Controller partnership has been configured, you can configure any **rcrelationship** or **rcconsistgrp** that you need. Make sure that a single volume is only in one relationship.

Triangle configuration

Figure 9-9 shows the triangle configuration.

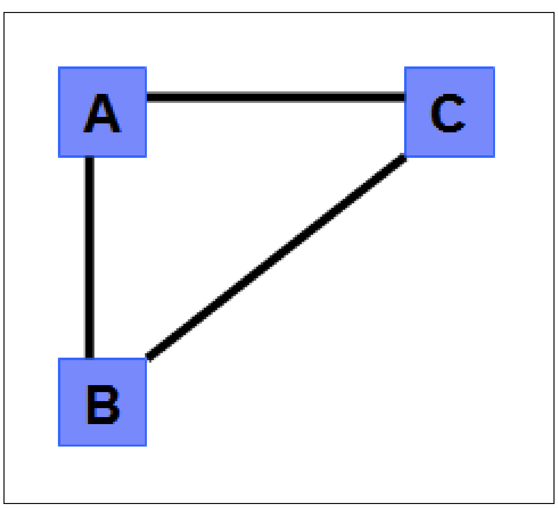


Figure 9-9 Triangle configuration

Example 9-171 shows the sequence of **mkpartnership** commands to execute to create a triangle configuration.

Example 9-171 Creating a triangle configuration

```
From ITSO_SVC1 to ITSO_SVC2 and ITSO_SVC3

IBM_2145:ITSO_SVC1:admin>mkpartnership -bandwidth 50 ITSO_SVC2
IBM_2145:ITSO_SVC1:admin>mkpartnership -bandwidth 50 ITSO_SVC3
IBM_2145:ITSO_SVC1:admin>lspartnership
id               name      location partnership                bandwidth
000002006BE04FC4 ITSO_SVC1 local
000002006AC03A42 ITSO_SVC2 remote   partially_configured_local 50
0000020060A06FB8 ITSO_SVC3 remote   partially_configured_local 50

From ITSO_SVC2 to ITSO_SVC1 and ITSO_SVC3

IBM_2145:ITSO_SVC2:admin>mkpartnership -bandwidth 50 ITSO_SVC1
IBM_2145:ITSO_SVC2:admin>mkpartnership -bandwidth 50 ITSO_SVC3
IBM_2145:ITSO_SVC2:admin>lspartnership
id               name                location partnership                bandwidth
000002006AC03A42 ITSO_SVC2           local
000002006BE04FC4 ITSO_SVC1           remote   fully_configured           50
0000020060A06FB8 ITSO_SVC3           remote   partially_configured_local 50

From ITSO_SVC3 to ITSO_SVC1 and ITSO_SVC2

IBM_2145:ITSO_SVC3:admin>mkpartnership -bandwidth 50 ITSO_SVC1
IBM_2145:ITSO_SVC3:admin>mkpartnership -bandwidth 50 ITSO_SVC2
IBM_2145:ITSO_SVC3:admin>lspartnership
id               name      location partnership      bandwidth
0000020060A06FB8 ITSO_SVC3 local
000002006BE04FC4 ITSO_SVC1 remote   fully_configured 50
000002006AC03A42 ITSO_SVC2 remote   fully_configured 50
```

After the SAN Volume Controller partnership has been configured, you can configure any **rcrelationship** or **rcconsistgrp** that you need. Make sure that a single volume is only in one relationship.

Fully connected configuration

Figure 9-10 on page 276 shows the fully connected configuration.

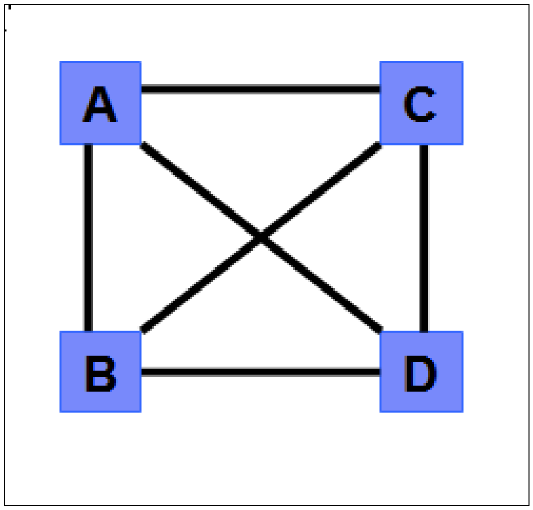


Figure 9-10 Fully connected configuration

Example 9-172 shows the sequence of **mkpartnership** commands to execute to create a fully connected configuration.

Example 9-172 Creating a fully connected configuration

```
From ITSO_SVC1 to ITSO_SVC2, ITSO_SVC3 and ITSO_SVC4

IBM_2145:ITSO_SVC1:admin>mkpartnership -bandwidth 50 ITSO_SVC2
IBM_2145:ITSO_SVC1:admin>mkpartnership -bandwidth 50 ITSO_SVC3
IBM_2145:ITSO_SVC1:admin>mkpartnership -bandwidth 50 ITSO_SVC4
IBM_2145:ITSO_SVC1:admin>lspartnership
id               name      location partnership                bandwidth
000002006BE04FC4 ITSO_SVC1 local
000002006AC03A42 ITSO_SVC2 remote   partially_configured_local 50
0000020060A06FB8 ITSO_SVC3 remote   partially_configured_local 50
0000020061C06FCA ITSO_SVC4 remote   partially_configured_local 50

From ITSO_SVC2 to ITSO_SVC1, ITSO_SVC3 and ITSO-SVC4

IBM_2145:ITSO_SVC2:admin>mkpartnership -bandwidth 50 ITSO_SVC1
IBM_2145:ITSO_SVC2:admin>mkpartnership -bandwidth 50 ITSO_SVC3
IBM_2145:ITSO_SVC2:admin>mkpartnership -bandwidth 50 ITSO_SVC4
IBM_2145:ITSO_SVC2:admin>lspartnership
id               name      location partnership                bandwidth
000002006AC03A42 ITSO_SVC2 local
000002006BE04FC4 ITSO_SVC1 remote   fully_configured           50
0000020060A06FB8 ITSO_SVC3 remote   partially_configured_local 50
0000020061C06FCA ITSO_SVC4 remote   partially_configured_local 50

From ITSO_SVC3 to ITSO_SVC1, ITSO_SVC3 and ITSO-SVC4

IBM_2145:ITSO_SVC3:admin>mkpartnership -bandwidth 50 ITSO_SVC1
IBM_2145:ITSO_SVC3:admin>mkpartnership -bandwidth 50 ITSO_SVC2
IBM_2145:ITSO_SVC3:admin>mkpartnership -bandwidth 50 ITSO_SVC4
IBM_2145:ITSO_SVC3:admin>lspartnership
id               name      location partnership                bandwidth
0000020060A06FB8 ITSO_SVC3 local
000002006BE04FC4 ITSO_SVC1 remote   fully_configured           50
000002006AC03A42 ITSO_SVC2 remote   fully_configured           50
0000020061C06FCA ITSO_SVC4 remote   partially_configured_local 50
```

```
From ITSO-SVC4 to ITSO_SVC1, ITSO_SVC2 and ITSO_SVC3

IBM_2145:ITSO_SVC4:admin>mkpartnership -bandwidth 50 ITSO_SVC1
IBM_2145:ITSO_SVC4:admin>mkpartnership -bandwidth 50 ITSO_SVC2
IBM_2145:ITSO_SVC4:admin>mkpartnership -bandwidth 50 ITSO_SVC3
IBM_2145:ITSO_SVC4:admin>lspartnership
id               name      location partnership      bandwidth
0000020061C06FCA ITSO_SVC4 local
000002006BE04FC4 ITSO_SVC1 remote   fully_configured 50
000002006AC03A42 ITSO_SVC2 remote   fully_configured 50
0000020060A06FB8 ITSO_SVC3 remote   fully_configured 50
```

After the SAN Volume Controller partnership has been configured, you can configure any **rcrelationship** or **rcconsistgrp** that you need. Make sure that a single volume is only in one relationship.

Daisy-chain configuration

Figure 9-11 shows the daisy-chain configuration.

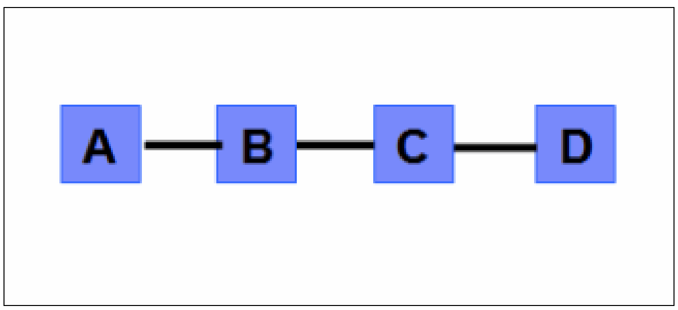


Figure 9-11 Daisy-chain configuration

Example 9-173 shows the sequence of **mkpartnership** commands to execute to create a daisy-chain configuration.

Example 9-173 Creating a daisy-chain configuration

```
From ITSO_SVC1 to ITSO_SVC2

IBM_2145:ITSO_SVC1:admin>mkpartnership -bandwidth 50 ITSO_SVC2
IBM_2145:ITSO_SVC1:admin>lspartnership
id               name      location partnership                bandwidth
000002006BE04FC4 ITSO_SVC1 local
000002006AC03A42 ITSO_SVC2 remote   partially_configured_local 50

From ITSO_SVC2 to ITSO_SVC1 and ITSO_SVC3

IBM_2145:ITSO_SVC2:admin>mkpartnership -bandwidth 50 ITSO_SVC1
IBM_2145:ITSO_SVC2:admin>mkpartnership -bandwidth 50 ITSO_SVC3
IBM_2145:ITSO_SVC2:admin>lspartnership
id               name      location partnership                bandwidth
000002006AC03A42 ITSO_SVC2 local
000002006BE04FC4 ITSO_SVC1 remote   fully_configured           50
0000020060A06FB8 ITSO_SVC3 remote   partially_configured_local 50

From ITSO_SVC3 to ITSO_SVC2 and ITSO_SVC4

IBM_2145:ITSO_SVC3:admin>mkpartnership -bandwidth 50 ITSO_SVC2
```

```
IBM_2145:ITSO_SVC3:admin>mkpartnership -bandwidth 50 ITSO_SVC4
IBM_2145:ITSO_SVC3:admin>lspartnership
id               name      location partnership                bandwidth
0000020060A06FB8 ITSO_SVC3 local
000002006AC03A42 ITSO_SVC2 remote   fully_configured           50
0000020061C06FCA ITSO_SVC4 remote   partially_configured_local 50

From ITSO_SVC4 to ITSO_SVC3

IBM_2145:ITSO_SVC4:admin>mkpartnership -bandwidth 50 ITSO_SVC3
IBM_2145:ITSO_SVC4:admin>lspartnership
id               name      location partnership      bandwidth
0000020061C06FCA ITSO_SVC4 local
0000020060A06FB8 ITSO_SVC3 remote   fully_configured 50
```

After the SAN Volume Controller partnership has been configured, you can configure any **rcrelationship** or **rcconsistgrp** that you need. Make sure that a single volume is only in one relationship.

9.15 Global Mirror operation

In the following scenario, we set up an intercluster Global Mirror relationship between the SAN Volume Controller system ITSO_SVC1 at the primary site and the SAN Volume Controller system ITSO_SVC4 at the secondary site.

Intercluster example: This example is for an *intercluster* Global Mirror operation only. In case you want to set up an *intracluster* operation, we highlight those parts in the following procedure that you do *not* need to perform.

Table 9-4 shows the details of the volumes.

Table 9-4 Details of volumes for Global Mirror relationship scenario

Content of volume	Volumes at primary site	Volumes at secondary site
Database files	GM_DB_Pri	GM_DB_Sec
Database log files	GM_DBLog_Pri	GM_DBLog_Sec
Application files	GM_App_Pri	GM_App_Sec

Because data consistency is needed across GM_DB_Pri and GM_DBLog_Pri, we create a Consistency Group to handle Global Mirror relationships for them. Because, in this scenario, the application files are independent of the database, we create a stand-alone Global Mirror relationship for GM_App_Pri. Figure 9-12 on page 279 illustrates the Global Mirror relationship setup.

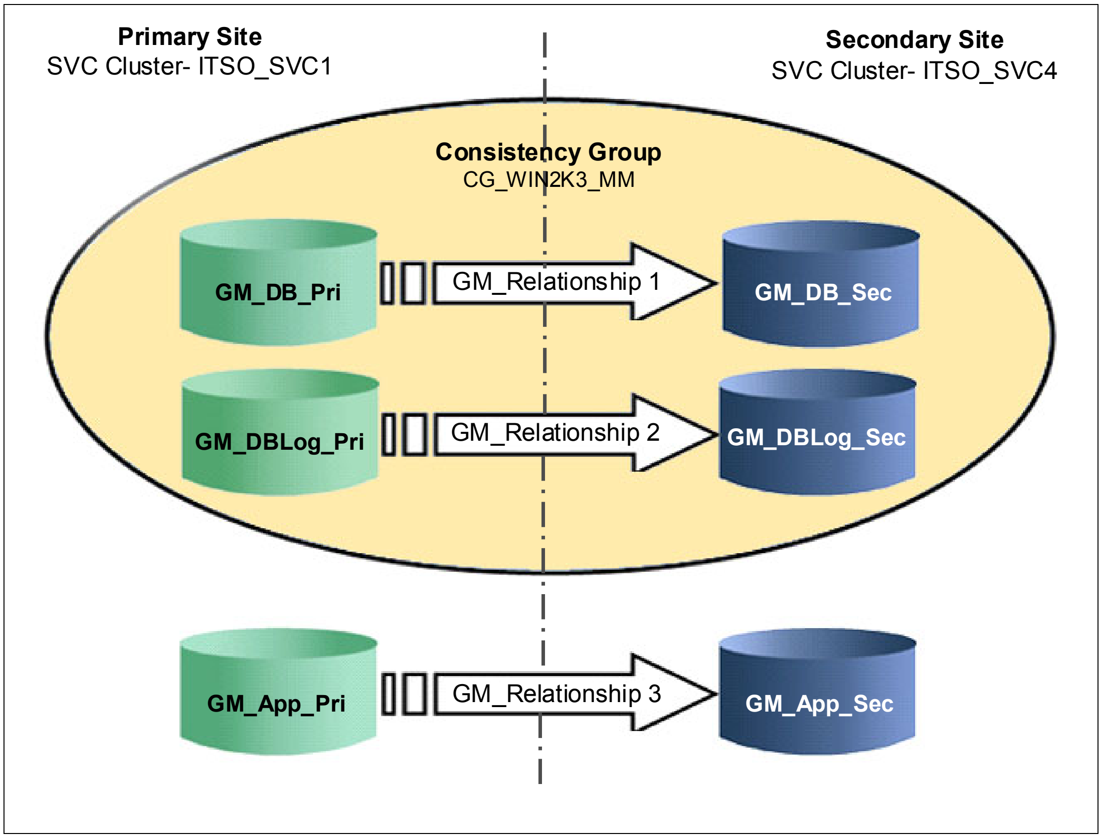


Figure 9-12 Global Mirror scenario

9.15.1 Setting up Global Mirror

In the following section, we assume that the source and target volumes have already been created and that the ISLs and zoning are in place, enabling the SAN Volume Controller systems to communicate.

To set up the Global Mirror, perform the following steps:

1. Create a SAN Volume Controller partnership between `ITSO_SVC1` and `ITSO_SVC4` on both SAN Volume Controller clustered systems.

   Bandwidth: `100` MBps

2. Create a Global Mirror Consistency Group.

   Name: `CG_W2K3_GM`

3. Create the Global Mirror relationship for `GM_DB_Pri`:

   – Master: `GM_DB_Pri`
   – Auxiliary: `GM_DB_Sec`
   – Auxiliary SAN Volume Controller system: `ITSO_SVC4`
   – Name: `GMREL1`
   – Consistency Group: `CG_W2K3_GM`

4. Create the Global Mirror relationship for `GM_DBLog_Pri`:

   – Master: `GM_DBLog_Pri`
   – Auxiliary: `GM_DBLog_Sec`
   – Auxiliary SAN Volume Controller system: `ITSO_SVC4`
   – Name: `GMREL2`

– Consistency Group: CG_W2K3_GM

5. Create the Global Mirror relationship for GM_App_Pri:
   – Master: GM_App_Pri
   – Auxiliary: GM_App_Sec
   – Auxiliary SAN Volume Controller system: ITSO_SVC4
   – Name: GMREL3

In the following sections, we perform each step by using the CLI.

9.15.2 Creating a SAN Volume Controller partnership between ITSO_SVC1 and ITSO_SVC4

We create a SAN Volume Controller partnership between these clustered systems.

Intracluster Global Mirror: If you are creating an intracluster Global Mirror, do *not* perform the next step. Instead, go to 9.15.3, "Changing link tolerance and system delay simulation" on page 281.

Preverification

To verify that both clustered systems can communicate with each other, use the **lspartnership** command. Example 9-174 confirms that our clustered systems are communicating because ITSO_SVC4 is an eligible SAN Volume Controller system candidate at ITSO_SVC1 for the SAN Volume Controller system partnership, and vice versa. Therefore, both systems communicate with each other.

Example 9-174 Listing the available SAN Volume Controller systems for partnership

```
IBM_2145:ITSO_SVC1:admin>lspartnershipcandidate
id               configured name
0000020061C06FCA no         ITSO_SVC4
IBM_2145:ITSO_SVC4:admin>lspartnershipcandidate
id               configured name
000002006BE04FC4 no         ITSO_SVC1
```

In Example 9-175, we show the output of the **lspartnership** command before setting up the SAN Volume Controller systems' partnership for Global Mirror. We show this output for comparison after we have set up the SAN Volume Controller partnership.

Example 9-175 Pre-verification of system configuration

```
IBM_2145:ITSO_SVC1:admin>lspartnership
id               name      location partnership bandwidth
000002006BE04FC4 ITSO_SVC1 local

IBM_2145:ITSO_SVC4:admin>lspartnership
id               name      location partnership bandwidth
0000020061C06FCA ITSO_SVC4 local
```

Partnership between systems

In Example 9-176, we create the partnership from ITSO_SVC1 to ITSO_SVC4, specifying a 100 MBps bandwidth to use for the background copy. To verify the status of the newly created partnership, we issue the **lspartnership** command. Notice that the new partnership is only *partially* configured. It will remain partially configured until we run the **mkpartnership** command on the other clustered system.

Example 9-176 Creating the partnership from ITSO_SVC1 to ITSO_SVC4 and verifying it

```
IBM_2145:ITSO_SVC1:admin>mkpartnership -bandwidth 100 ITSO_SVC4
IBM_2145:ITSO_SVC1:admin>lspartnership
id               name      location partnership                bandwidth
000002006BE04FC4 ITSO_SVC1 local
0000020061C06FCA ITSO_SVC4 remote   partially_configured_local 100
```

In Example 9-177, we create the partnership from ITSO_SVC4 back to ITSO_SVC1, specifying a 100 MBps bandwidth to be used for the background copy. After creating the partnership, verify that the partnership is *fully* configured by reissuing the **lspartnership** command.

Example 9-177 Creating the partnership from ITSO_SVC4 to ITSO_SVC1 and verifying it

```
IBM_2145:ITSO_SVC4:admin>mkpartnership -bandwidth 100 ITSO_SVC1

IBM_2145:ITSO_SVC4:admin>lspartnership
id               name      location partnership      bandwidth
0000020061C06FCA ITSO_SVC4 local
000002006BE04FC4 ITSO_SVC1 remote   fully_configured 100

IBM_2145:ITSO_SVC1:admin>lspartnership
id               name      location partnership      bandwidth
000002006BE04FC4 ITSO_SVC1 local
0000020061C06FCA ITSO_SVC4 remote   fully_configured 100
```

9.15.3 Changing link tolerance and system delay simulation

The **gm_link_tolerance** parameter defines the sensitivity of the SAN Volume Controller to inter-link overload conditions. The value is the number of seconds of continuous link difficulties that will be tolerated before the SAN Volume Controller will stop the remote copy relationships to prevent affecting host I/O at the primary site. To change the value, use the following command:

```
chsystem -gmlinktolerance link_tolerance
```

The *link_tolerance* value is 60 - 86,400 seconds in increments of 10 seconds. The default value for the link tolerance is 300 seconds. A value of 0 disables link tolerance.

> **Important:** We strongly suggest that you use the default value. If the link is overloaded for a period, which affects host I/O at the primary site, the relationships will be stopped to protect those hosts.

Intercluster and intracluster delay simulation

This Global Mirror feature permits a simulation of a delayed write to a remote volume. This feature allows testing to be performed that detects colliding writes, and you can use this feature to test an application before the full deployment of the Global Mirror feature. The delay simulation can be enabled separately for each intracluster or intercluster Global Mirror. To enable this feature, run the following command either for the intracluster or intercluster simulation:

- For intercluster:

  `chsystem -gminterdelaysimulation <inter_cluster_delay_simulation>`

- For intracluster:

  `chsystem -gmintradelaysimulation <intra_cluster_delay_simulation>`

The *inter_cluster_delay_simulation* and *intra_cluster_delay_simulation* values express the amount of time (in milliseconds) that secondary I/Os are delayed respectively for intercluster and intracluster relationships. These values specify the number of milliseconds that I/O activity (that is, copying a primary volume to a secondary volume) is delayed. You can set a value 0 - 100 milliseconds in 1-millisecond increments for the cluster_delay_simulation in the previous commands. A value of zero (0) disables the feature.

To check the current settings for the delay simulation, use the following command:

`lssystem`

In Example 9-178, we show the modification of the delay simulation value and a change of the Global Mirror link tolerance parameters. We also show the changed values of the Global Mirror link tolerance and delay simulation parameters.

Example 9-178 Delay simulation and link tolerance modification

```
IBM_2145:ITSO_SVC1:admin>chsystem -gminterdelaysimulation 20
IBM_2145:ITSO_SVC1:admin>chsystem -gmintradelaysimulation 40
IBM_2145:ITSO_SVC1:admin>chsystem -gmlinktolerance 200
IBM_2145:ITSO_SVC1:admin>lssystem
id 000002006BE04FC4
name ITSO_SVC1
location local
partnership
bandwidth
total_mdisk_capacity 866.5GB
space_in_mdisk_grps 766.5GB
space_allocated_to_vdisks 30.00GB
total_free_space 836.5GB
total_vdiskcopy_capacity 30.00GB
total_used_capacity 30.00GB
total_overallocation 3
total_vdisk_capacity 30.00GB
total_allocated_extent_capacity 31.50GB
statistics_status on
statistics_frequency 15
cluster_locale en_US
time_zone 520 US/Pacific
code_level 6.3.0.0 (build 54.0.1109090000)
console_IP 10.18.228.81:443
id_alias 000002006BE04FC4
gm_link_tolerance 200
gm_inter_cluster_delay_simulation 20
gm_intra_cluster_delay_simulation 40
gm_max_host_delay 5
email_reply
email_contact
email_contact_primary
email_contact_alternate
email_contact_location
```

```
email_contact2
email_contact2_primary
email_contact2_alternate
email_state stopped
inventory_mail_interval 0
cluster_ntp_IP_address
cluster_isns_IP_address
iscsi_auth_method chap
iscsi_chap_secret passw0rd
auth_service_configured no
auth_service_enabled no
auth_service_url
auth_service_user_name
auth_service_pwd_set no
auth_service_cert_set no
auth_service_type tip
relationship_bandwidth_limit 25
tier generic_ssd
tier_capacity 0.00MB
tier_free_capacity 0.00MB
tier generic_hdd
tier_capacity 766.50GB
tier_free_capacity 736.50GB
has_nas_key no
layer appliance
```

9.15.4 Creating a Global Mirror Consistency Group

In Example 9-179, we create the Global Mirror Consistency Group using the **mkrcconsistgrp** command. We will use this Consistency Group for the Global Mirror relationships for the database volumes. The Consistency Group is named CG_W2K3_GM.

Example 9-179 Creating the Global Mirror Consistency Group CG_W2K3_GM

```
IBM_2145:ITSO_SVC1:admin>mkrcconsistgrp -cluster ITSO_SVC4 -name CG_W2K3_GM
RC Consistency Group, id [0], successfully created
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp
id name       master_cluster_id master_cluster_name aux_cluster_id   aux_cluster_name
primary state relationship_count copy_type   cycling_mode
0  CG_W2K3_GM 000002006BE04FC4  ITSO_SVC1          0000020061C06FCA ITSO_SVC4
empty 0                  empty_group none
```

9.15.5 Creating Global Mirror relationships

In Example 9-180 on page 284, we create the GMREL1 and GMREL2 Global Mirror relationships for the GM_DB_Pri and GM_DBLog_Pri volumes. We also make them members of the CG_W2K3_GM Global Mirror Consistency Group.

We use the **lsvdisk** command to list all of the volumes in the ITSO_SVC1 system, and then, use the **lsrcrelationshipcandidate** command to show the possible candidate volumes for GM_DB_Pri in ITSO_SVC4.

After checking all of these conditions, we use the **mkrcrelationship** command to create the Global Mirror relationship. To verify the newly created Global Mirror relationships, we list them with the **lsrcrelationship** command.

Example 9-180 Creating GMREL1 and GMREL2 Global Mirror relationships

```
IBM_2145:ITSO_SVC1:admin>lsvdisk -filtervalue name=GM*
id name         IO_group_id IO_group_name status mdisk_grp_id mdisk_grp_name capacity type    FC_id FC_name
RC_id RC_name vdisk_UID                        fc_map_count copy_count fast_write_state se_copy_count
RC_change
0  GM_DB_Pri    0         io_grp0      online 0          Pool_DS3500-1  10.00GB  striped
6005076801AF813F1000000000000031 0           1          empty          0            0
no
1  GM_DBLog_Pri 0         io_grp0      online 0          Pool_DS3500-1  10.00GB  striped
6005076801AF813F1000000000000032 0           1          empty          0            0
no
2  GM_App_Pri   0         io_grp0      online 0          Pool_DS3500-1  10.00GB  striped
6005076801AF813F1000000000000033 0           1          empty          0            0
no

IBM_2145:ITSO_SVC1:admin>lsrcrelationshipcandidate -aux ITSO_SVC4 -master GM_DB_Pri
id vdisk_name
0  GM_DB_Sec
1  GM_DBLog_Sec
2  GM_App_Sec

IBM_2145:ITSO_SVC1:admin>mkrcrelationship -master GM_DB_Pri -aux GM_DB_Sec -cluster ITSO_SVC4 -consistgrp
CG_W2K3_GM -name GMREL1 -global
RC Relationship, id [0], successfully created

IBM_2145:ITSO_SVC1:admin>mkrcrelationship -master GM_DBLog_Pri -aux GM_DBLog_Sec -cluster ITSO_SVC4
-consistgrp CG_W2K3_GM -name GMREL2 -global
RC Relationship, id [1], successfully created

IBM_2145:ITSO_SVC1:admin>mkrcrelationship -master GM_DB_Pri -aux GM_DB_Sec -cluster ITSO_SVC4 -consistgrp
CG_W2K3_GM -name GMREL1 -global
RC Relationship, id [2], successfully created
IBM_2145:ITSO_SVC1:admin>mkrcrelationship -master GM_DBLog_Pri -aux GM_DBLog_Sec -cluster ITSO_SVC4
-consistgrp CG_W2K3_GM -name GMREL2 -global
RC Relationship, id [3], successfully created
IBM_2145:ITSO_SVC1:admin>lsrcrelationship
id name   master_cluster_id master_cluster_name master_vdisk_id master_vdisk_name aux_cluster_id
aux_cluster_name aux_vdisk_id aux_vdisk_name primary consistency_group_id consistency_group_name state
bg_copy_priority progress copy_type cycling_mode
0  GMREL1 000002006BE04FC4  ITSO_SVC1          0                GM_DB_Pri         0000020061C06FCA
ITSO_SVC4        0             GM_DB_Sec      master  0                    CG_W2K3_GM
inconsistent_stopped 50               0       global    none
1  GMREL2 000002006BE04FC4  ITSO_SVC1          1                GM_DBLog_Pri      0000020061C06FCA
ITSO_SVC4        1             GM_DBLog_Sec   master  0                    CG_W2K3_GM
inconsistent_stopped 50               0       global    none
```

9.15.6 Creating the stand-alone Global Mirror relationship for GM_App_Pri

In Example 9-181 on page 285, we create the stand-alone Global Mirror relationship `GMREL3` for `GM_App_Pri`. After it is created, we will check the status of each of our Global Mirror relationships.

Notice that the status of `GMREL3` is consistent_stopped because it was created with the **`-sync`** option. The **`-sync`** option indicates that the secondary (auxiliary) volume is already synchronized with the primary (master) volume. The initial background synchronization is skipped when this option is used.

GMREL1 and GMREL2 are in the inconsistent_stopped state because they were not created with the **-sync** option, so their auxiliary volumes need to be synchronized with their primary volumes.

Example 9-181 Creating a stand-alone Global Mirror relationship and verifying it

```
IBM_2145:ITSO_SVC1:admin>mkrcrelationship -master GM_App_Pri -aux GM_App_Sec -cluster ITSO_SVC4 -sync -name
GMREL3 -global
RC Relationship, id [2], successfully created

IBM_2145:ITSO_SVC1:admin>lsrcrelationship -delim :
id:name:master_cluster_id:master_cluster_name:master_vdisk_id:master_vdisk_name:aux_cluster_id:aux_cluster_
name:aux_vdisk_id:aux_vdisk_name:primary:consistency_group_id:consistency_group_name:state:bg_copy_priority
:progress:copy_type:cycling_mode
0:GMREL1:000002006BE04FC4:ITSO_SVC1:0:GM_DB_Pri:0000020061C06FCA:ITSO_SVC4:0:GM_DB_Sec:master:0:CG_W2K3_GM:
inconsistent_copying:50:73:global:none
1:GMREL2:000002006BE04FC4:ITSO_SVC1:1:GM_DBLog_Pri:0000020061C06FCA:ITSO_SVC4:1:GM_DBLog_Sec:master:0:CG_W2
K3_GM:inconsistent_copying:50:75:global:none
2:GMREL3:000002006BE04FC4:ITSO_SVC1:2:GM_App_Pri:0000020061C06FCA:ITSO_SVC4:2:GM_App_Sec:master:::consisten
t_stopped:50:100:global:none
```

9.15.7 Starting Global Mirror

Now that we have created the Global Mirror Consistency Group and relationships, we are ready to use the Global Mirror relationships in our environment.

When implementing Global Mirror, the goal is to reach a consistent and synchronized state that can provide redundancy in case a hardware failure occurs that affects the SAN at the production site.

In this section, we show how to start the stand-alone Global Mirror relationships and the Consistency Group.

9.15.8 Starting a stand-alone Global Mirror relationship

In Example 9-182, we start the stand-alone Global Mirror relationship named GMREL3. Because the Global Mirror relationship was in the Consistent stopped state and no updates have been made to the primary volume, the relationship quickly enters the Consistent synchronized state.

Example 9-182 Starting the stand-alone Global Mirror relationship

```
IBM_2145:ITSO_SVC1:admin>startrcrelationship GMREL3
IBM_2145:ITSO_SVC1:admin>lsrcrelationship GMREL3
id 2
name GMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name GM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name GM_App_Sec
primary master
consistency_group_id
consistency_group_name
state consistent_synchronized
```

```
bg_copy_priority 50
progress
freeze_time
status online
sync
copy_type global
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name
```

9.15.9 Starting a Global Mirror Consistency Group

In Example 9-183, we start the `CG_W2K3_GM` Global Mirror Consistency Group. Because the Consistency Group was in the Inconsistent stopped state, it enters the Inconsistent copying state until the background copy has completed for all of the relationships that are in the Consistency Group.

Upon completion of the background copy, the `CG_W2K3_GM` Global Mirror Consistency Group enters the Consistent synchronized state.

Example 9-183 Starting the Global Mirror Consistency Group

```
IBM_2145:ITSO_SVC1:admin>startrcconsistgrp CG_W2K3_GM
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp 0
id 0
name CG_W2K3_GM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary master
state inconsistent_copying
relationship_count 2
freeze_time
status
sync
copy_type global
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name GMREL1
RC_rel_id 1
RC_rel_name GMREL2
```

9.15.10 Monitoring the background copy progress

To monitor the background copy progress, use the **`lsrcrelationship`** command. This command shows us all of the defined Global Mirror relationships if it is used without any parameters. In the command output, progress indicates the current background copy progress. Example 9-184 on page 287 shows our Global Mirror relationships.

Using SNMP traps: Setting up SNMP traps for the SAN Volume Controller enables automatic notification when Global Mirror Consistency Groups or relationships change state.

Example 9-184 Monitoring background copy progress example

```
IBM_2145:ITSO_SVC1:admin>lsrcrelationship GMREL1
id 0
name GMREL1
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 0
master_vdisk_name GM_DB_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 0
aux_vdisk_name GM_DB_Sec
primary master
consistency_group_id 0
consistency_group_name CG_W2K3_GM
state inconsistent_copying
bg_copy_priority 50
progress 38
freeze_time
status online
sync
copy_type global
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name

IBM_2145:ITSO_SVC1:admin>lsrcrelationship GMREL2
id 1
name GMREL2
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 1
master_vdisk_name GM_DBLog_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 1
aux_vdisk_name GM_DBLog_Sec
primary master
consistency_group_id 0
consistency_group_name CG_W2K3_GM
state inconsistent_copying
bg_copy_priority 50
progress 76
freeze_time
status online
sync
copy_type global
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
```

```
aux_change_vdisk_id
aux_change_vdisk_name
```

When all of the Global Mirror relationships complete the background copy, the Consistency Group enters the Consistent synchronized state, as shown in Example 9-185.

Example 9-185 Listing the Global Mirror Consistency Group

```
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_GM
id 0
name CG_W2K3_GM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary master
state consistent_synchronized
relationship_count 2
freeze_time
status
sync
copy_type global
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name GMREL1
RC_rel_id 1
RC_rel_name GMREL2
```

9.15.11 Stopping and restarting Global Mirror

Now that the Global Mirror Consistency Group and relationships are running, we describe how to stop, restart, and change the direction of the stand-alone Global Mirror relationships and the Consistency Group.

First, we show how to stop and restart the stand-alone Global Mirror relationships and the Consistency Group.

9.15.12 Stopping a stand-alone Global Mirror relationship

In Example 9-186, we stop the stand-alone Global Mirror relationship while enabling access (write I/O) to both the primary and the secondary volume. As a result, the relationship enters the Idling state.

Example 9-186 Stopping the stand-alone Global Mirror relationship

```
IBM_2145:ITSO_SVC1:admin>stoprcrelationship -access GMREL3
IBM_2145:ITSO_SVC1:admin>lsrcrelationship GMREL3
id 2
name GMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name GM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
```

```
aux_vdisk_name GM_App_Sec
primary
consistency_group_id
consistency_group_name
state idling
bg_copy_priority 50
progress
freeze_time
status
sync in_sync
copy_type global
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name
```

9.15.13 Stopping a Global Mirror Consistency Group

In Example 9-187, we stop the Global Mirror Consistency Group *without* specifying the **-access** parameter. Therefore, the Consistency Group enters the Consistent stopped state.

Example 9-187 Stopping a Global Mirror Consistency Group without specifying -access

```
IBM_2145:ITSO_SVC1:admin>stoprcconsistgrp CG_W2K3_GM
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp  CG_W2K3_GM
id 0
name CG_W2K3_GM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary master
state consistent_stopped
relationship_count 2
freeze_time
status
sync in_sync
copy_type global
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name GMREL1
RC_rel_id 1
RC_rel_name GMREL2
```

If, afterward, we want to enable access (write I/O) for the secondary volume, we can reissue the **stoprcconsistgrp** command specifying the **-access** parameter. The Consistency Group transits to the Idling state, as shown in Example 9-188.

Example 9-188 Stopping a Global Mirror Consistency Group

```
IBM_2145:ITSO_SVC1:admin>stoprcconsistgrp -access CG_W2K3_GM
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_GM
id 0
name CG_W2K3_GM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
```

```
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary
state idling
relationship_count 2
freeze_time
status
sync in_sync
copy_type global
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name GMREL1
RC_rel_id 1
RC_rel_name GMREL2
```

9.15.14 Restarting a Global Mirror relationship in the Idling state

When restarting a Global Mirror relationship in the Idling state, we must specify the copy direction.

If any updates have been performed on either the master or the auxiliary volume, consistency will be compromised. Therefore, we must issue the **-force** parameter to restart the relationship. If the **-force** parameter is not used, the command will fail, as shown in Example 9-189.

Example 9-189 Restarting a Global Mirror relationship after updates in the Idling state

```
IBM_2145:ITSO_SVC1:admin>startrcrelationship -primary master -force GMREL3
IBM_2145:ITSO_SVC1:admin>lsrcrelationship GMREL3
id 2
name GMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name GM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name GM_App_Sec
primary master
consistency_group_id
consistency_group_name
state consistent_synchronized
bg_copy_priority 50
progress
freeze_time
status online
sync
copy_type global
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name
```

9.15.15 Restarting a Global Mirror Consistency Group in the Idling state

When restarting a Global Mirror Consistency Group in the Idling state, we must specify the copy direction.

If any updates have been performed on either the master or the auxiliary volume in any of the Global Mirror relationships in the Consistency Group, consistency will be compromised. Therefore, we must issue the **-force** parameter to start the relationship. If the **-force** parameter is not used, the command will fail.

In Example 9-190, we restart the Consistency Group and change the copy direction by specifying the auxiliary volumes to become the primaries.

Example 9-190 Restarting a Global Mirror relationship while changing the copy direction

```
IBM_2145:ITSO_SVC1:admin>startrcconsistgrp -primary aux CG_W2K3_GM
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_GM
id 0
name CG_W2K3_GM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary aux
state consistent_synchronized
relationship_count 2
freeze_time
status
sync
copy_type global
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name GMREL1
RC_rel_id 1
RC_rel_name GMREL2
```

9.15.16 Changing the direction for Global Mirror

In this section, we show how to change the copy direction of the stand-alone Global Mirror relationships and the Consistency Group.

9.15.17 Switching the copy direction for a Global Mirror relationship

When a Global Mirror relationship is in the Consistent synchronized state, we can change the copy direction for the relationship by using the **switchrcrelationship** command and specifying the primary volume.

If the volume that is specified as the primary when issuing this command is already a primary, the command has no effect.

In Example 9-191, we change the copy direction for the stand-alone Global Mirror relationship, specifying the auxiliary volume to become the primary.

> **Important:** When the copy direction is switched, it is crucial that there is no outstanding I/O to the volume that transits from primary to secondary because all I/O will be inhibited to that volume when it becomes the secondary. Therefore, careful planning is required before using the **`switchrcrelationship`** command.

Example 9-191 Switching the copy direction for a Global Mirror relationship

```
IBM_2145:ITSO_SVC1:admin>lsrcrelationship GMREL3
id 2
name GMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name GM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name GM_App_Sec
primary master
consistency_group_id
consistency_group_name
state consistent_synchronized
bg_copy_priority 50
progress
freeze_time
status online
sync
copy_type global
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name

IBM_2145:ITSO_SVC1:admin>switchrcrelationship -primary aux GMREL3
IBM_2145:ITSO_SVC1:admin>lsrcrelationship GMREL3
id 2
name GMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name GM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name GM_App_Sec
primary aux
consistency_group_id
consistency_group_name
state consistent_synchronized
bg_copy_priority 50
progress
freeze_time
status online
sync
```

```
copy_type global
cycle_period_seconds 300
cycling_mode none
master_change_vdisk_id
master_change_vdisk_name
aux_change_vdisk_id
aux_change_vdisk_name
```

9.15.18 Switching the copy direction for a Global Mirror Consistency Group

When a Global Mirror Consistency Group is in the Consistent synchronized state, we can change the copy direction for the relationship by using the **switchrcconsistgrp** command and specifying the primary volume. If the volume that is specified as the primary when issuing this command is already a primary, the command has no effect.

In Example 9-192, we change the copy direction for the Global Mirror Consistency Group, specifying the auxiliary to become the primary.

> **Important:** When the copy direction is switched, it is crucial that there is no outstanding I/O to the volume that transits from primary to secondary because all I/O will be inhibited when that volume becomes the secondary. Therefore, careful planning is required before using the **switchrcconsistgrp** command.

Example 9-192 Switching the copy direction for a Global Mirror Consistency Group

```
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_GM
id 0
name CG_W2K3_GM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary master
state consistent_synchronized
relationship_count 2
freeze_time
status
sync
copy_type global
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name GMREL1
RC_rel_id 1
RC_rel_name GMREL2
IBM_2145:ITSO_SVC1:admin>switchrcconsistgrp -primary aux CG_W2K3_GM
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_GM
id 0
name CG_W2K3_GM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary aux
state consistent_synchronized
relationship_count 2
freeze_time
```

```
status
sync
copy_type global
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
RC_rel_name GMREL1
RC_rel_id 1
RC_rel_name GMREL2
```

9.15.19 Changing a Global Mirror relationship to the cycling mode

Starting with SAN Volume Controller 6.3, Global Mirror can operate with or without cycling. When operating without cycling, write operations are applied to the secondary volume as soon as possible after they are applied to the primary volume. The secondary volume is generally less than 1 second behind the primary volume, which minimizes the amount of data that must be recovered in the event of a failover. However, this capability requires that a high-bandwidth link is provisioned between the two sites.

When Global Mirror operates in cycling mode, changes are tracked and, where needed, copied to intermediate Change Volumes. Changes are transmitted to the secondary site periodically. The secondary volumes are much further behind the primary volume, and more data must be recovered in the event of a failover. Because the data transfer can be smoothed over a longer time period, however, lower bandwidth is required to provide an effective solution.

A Global Mirror relationship consists of two volumes: primary and secondary. With SAN Volume Controller 6.3, each of these volumes can be associated to a Change Volume. *Change Volumes* are used to record changes to the remote copy volume. A FlashCopy relationship exists between the remote copy volume and the Change Volume. This relationship cannot be manipulated as a normal FlashCopy relationship. Most commands will fail by design because this relationship is an internal relationship.

Cycling mode transmits a series of FlashCopy images from the primary to the secondary, and it is enabled using **svctask chrcrelationship -cycling=multi**.

The primary Change Volume stores changes to be sent to the secondary volume, and the secondary Change Volume is used to maintain a consistent image at the secondary volume. Every *x* seconds, the primary FlashCopy mapping is started automatically, where *x* is the “cycling period” and is configurable. Data is then copied to the secondary volume from the primary Change Volume. The secondary FlashCopy mapping is started if resynchronization is needed. Therefore, there is always a consistent copy at the secondary volume. The cycling period is configurable, and the default value is 300 seconds.

The recovery point objective (RPO) depends on how long the FlashCopy takes to complete. If the FlashCopy completes within the cycling time, the maximum RPO = 2 x the cycling time; otherwise, the RPO = 2 x the copy completion time.

You can estimate the current RPO using the new **freeze_time** rcrelationship property. It is the time of the last consistent image that is present at the secondary. Figure 9-13 on page 295 shows the cycling mode with Change Volumes.

Change Volume requirements

Follow these rules for the Change Volume:

- The Change Volume can be a thin-provisioned volume.

- It must be the same size as the primary and secondary volumes.
- The Change Volume must be in the same I/O Group as the primary and secondary volumes.
- It cannot be used for the user's remote copy or FlashCopy mappings.
- You must have a Change Volume for both the primary and secondary volumes.
- You cannot manipulate it like a normal FlashCopy mapping.

In this section, we show how to change the cycling mode of the stand-alone Global Mirror relationship (`GMREL3`) and the Consistency Group `CG_W2K3_GM` Global Mirror relationships (`GMREL1` and `GMREL2`).

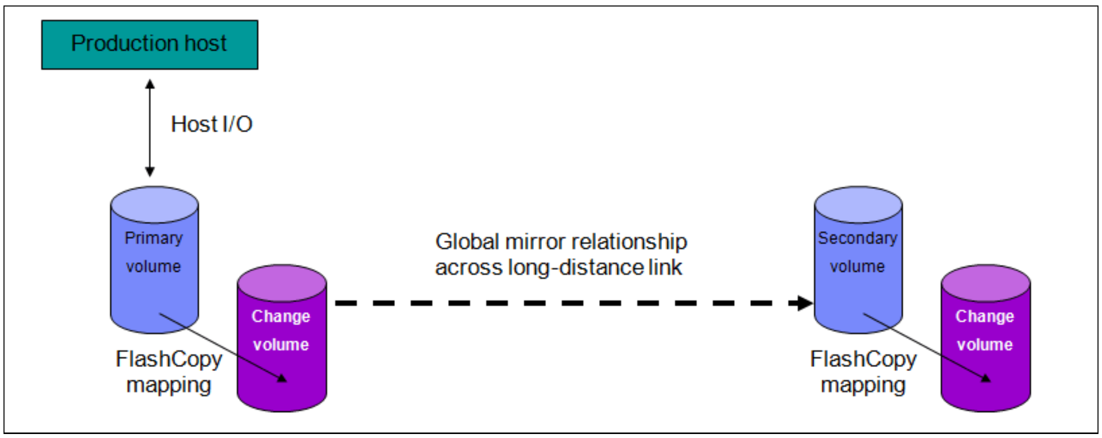


Figure 9-13 Global Mirror with Change Volumes

We assume that the source and target volumes have already been created and that the ISLs and zoning are in place, enabling the SAN Volume Controller systems to communicate. We also assume that the Global Mirror relationship has been already established.

To change the Global Mirror to cycling mode with Change Volumes, perform the following steps:

1. Create thin-provisioned Change Volumes for the primary and secondary volumes at both sites.
2. Stop the stand-alone relationship `GMREL3` to change the cycling mode at the primary site.
3. Set the cycling mode on the stand-alone relationship `GMREL3` at the primary site.
4. Set the Change Volume on the master volume relationship `GMREL3` at the primary site.
5. Set the Change Volume on the auxiliary volume relationship `GMREL3` at the secondary site.
6. Start the stand-alone relationship `GMREL3` in cycling mode at the primary site.
7. Stop the Consistency Group `CG_W2K3_GM` to change the cycling mode at the primary site.
8. Set the cycling mode on the Consistency Group at the primary site.
9. Set the Change Volume on the master volume relationship `GMREL1` of the Consistency Group `CG_W2K3_GM` at the primary site.
10. Set the Change Volume on the auxiliary volume relationship `GMREL1` at the secondary site.
11. Set the Change Volume on the master volume relationship `GMREL2` of the Consistency Group `CG_W2K3_GM` at the primary site.
12. Set the Change Volume on the auxiliary volume relationship `GMREL2` at the secondary site.
13. Start the Consistency Group `CG_W2K3_GM` in the cycling mode at the primary site.

9.15.20 Creating the thin-provisioned Change Volumes

We start the setup by creating thin-provisioned Change Volumes for the primary and secondary volumes at both sites, as shown in Example 9-193.

Example 9-193 Creating the thin-provisioned volumes for Global Mirror cycling mode

```
IBM_2145:ITSO_SVC1:admin>mkvdisk -iogrp 0 -mdiskgrp 0 -size 10 -unit gb -rsize 20%
-autoexpand -grainsize 32 -name GM_DB_Pri_CHANGE_VOL
Virtual Disk, id [3], successfully created
IBM_2145:ITSO_SVC1:admin>mkvdisk -iogrp 0 -mdiskgrp 0 -size 10 -unit gb -rsize 20%
-autoexpand -grainsize 32 -name GM_DBLog_Pri_CHANGE_VOL
Virtual Disk, id [4], successfully created
IBM_2145:ITSO_SVC1:admin>mkvdisk -iogrp 0 -mdiskgrp 0 -size 10 -unit gb -rsize 20%
-autoexpand -grainsize 32 -name GM_App_Pri_CHANGE_VOL
Virtual Disk, id [5], successfully created

IBM_2145:ITSO_SVC4:admin>mkvdisk -iogrp 0 -mdiskgrp 0 -size 10 -unit gb -rsize 20%
-autoexpand -grainsize 32 -name GM_DB_Sec_CHANGE_VOL
Virtual Disk, id [3], successfully created
IBM_2145:ITSO_SVC4:admin>mkvdisk -iogrp 0 -mdiskgrp 0 -size 10 -unit gb -rsize 20%
-autoexpand -grainsize 32 -name GM_DBLog_Sec_CHANGE_VOL
Virtual Disk, id [4], successfully created
IBM_2145:ITSO_SVC4:admin>mkvdisk -iogrp 0 -mdiskgrp 0 -size 10 -unit gb -rsize 20%
-autoexpand -grainsize 32 -name GM_App_Sec_CHANGE_VOL
Virtual Disk, id [5], successfully created
```

9.15.21 Stopping the stand-alone remote copy relationship

We now display the remote copy relationships to ensure that they are in sync, and then we stop the stand-alone relationship GMREL3, as shown in Example 9-194.

Example 9-194 Stopping the remote copy stand-alone relationship

```
IBM_2145:ITSO_SVC1:admin>lsrcrelationship
id name   master_cluster_id master_cluster_name master_vdisk_id master_vdisk_name
aux_cluster_id   aux_cluster_name aux_vdisk_id aux_vdisk_name primary
consistency_group_id consistency_group_name state
bg_copy_priority progress copy_type cycling_mode
0  GMREL1 000002006BE04FC4  ITSO_SVC1          0              GM_DB_Pri
0000020061C06FCA ITSO_SVC4      0          GM_DB_Sec     aux    0
CG_W2K3_GM            consistent_synchronized 50                       global
none
1  GMREL2 000002006BE04FC4  ITSO_SVC1          1              GM_DBLog_Pri
0000020061C06FCA ITSO_SVC4      1          GM_DBLog_Sec   aux    0
CG_W2K3_GM            consistent_synchronized 50                       global
none
2  GMREL3 000002006BE04FC4  ITSO_SVC1          2              GM_App_Pri
0000020061C06FCA ITSO_SVC4     2       GM_App_Sec   aux
consistent_synchronized 50                        global    none

IBM_2145:ITSO_SVC1:admin>stoprcrelationship GMREL3
```

9.15.22 Setting the cycling mode on the stand-alone remote copy relationship

In Example 9-195, we set the cycling mode on the relationship using the **chrcrelationship** command. Note that the **cyclingmode** and **masterchange** parameters cannot be entered in the same command.

Example 9-195 Setting the cycling mode

```
IBM_2145:ITSO_SVC1:admin>chrcrelationship -cyclingmode multi  GMREL3
```

9.15.23 Setting the Change Volume on the master volume

In Example 9-196, we set the Change Volume for the primary volume. A display shows the name of the master Change Volume.

Example 9-196 Setting the Change Volume

```
IBM_2145:ITSO_SVC1:admin>chrcrelationship -masterchange GM_App_Pri_CHANGE_VOL

IBM_2145:ITSO_SVC1:admin>lsrcrelationship GMREL3
id 2
name GMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name GM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name GM_App_Sec
primary aux
consistency_group_id
consistency_group_name
state consistent_stopped
bg_copy_priority 50
progress 100
freeze_time
status online
sync in_sync
copy_type global
cycle_period_seconds 300
cycling_mode multi
master_change_vdisk_id 5
master_change_vdisk_name GM_App_Pri_CHANGE_VOL
aux_change_vdisk_id
aux_change_vdisk_name
```

9.15.24 Setting the Change Volume on the auxiliary volume

In Example 9-197, we set the Change Volume on the auxiliary volume in the secondary site. From the display, we can see the name of the volume.

Example 9-197 Setting the Change Volume on the auxiliary volume

```
IBM_2145:ITSO_SVC4:admin>chrcrelationship -auxchange GM_App_Sec_CHANGE_VOL 2
IBM_2145:ITSO_SVC4:admin>
IBM_2145:ITSO_SVC4:admin>lsrcrelationship GMREL3
id 2
```

```
name GMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name GM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name GM_App_Sec
primary aux
consistency_group_id
consistency_group_name
state consistent_stopped
bg_copy_priority 50
progress 100
freeze_time
status online
sync in_sync
copy_type global
cycle_period_seconds 300
cycling_mode multi
master_change_vdisk_id 5
master_change_vdisk_name GM_App_Pri_CHANGE_VOL
aux_change_vdisk_id 5
aux_change_vdisk_name GM_App_Sec_CHANGE_VOL
```

9.15.25 Starting the stand-alone relationship in the cycling mode

In Example 9-198, we start the stand-alone relationship GMREL3. After a few minutes, we check the **freeze_time** parameter to see how it changes.

Example 9-198 Starting the stand-alone relationship in the cycling mode

```
IBM_2145:ITSO_SVC1:admin>startrcrelationship GMREL3

IBM_2145:ITSO_SVC1:admin>lsrcrelationship GMREL3
id 2
name GMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name GM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name GM_App_Sec
primary aux
consistency_group_id
consistency_group_name
state consistent_copying
bg_copy_priority 50
progress 100
freeze_time 2011/10/04/20/37/20
status online
sync
copy_type global
cycle_period_seconds 300
cycling_mode multi
master_change_vdisk_id 5
```

```
master_change_vdisk_name GM_App_Pri_CHANGE_VOL
aux_change_vdisk_id 5
aux_change_vdisk_name GM_App_Sec_CHANGE_VOL

IBM_2145:ITSO_SVC1:admin>lsrcrelationship GMREL3
id 2
name GMREL3
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 2
master_vdisk_name GM_App_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 2
aux_vdisk_name GM_App_Sec
primary aux
consistency_group_id
consistency_group_name
state consistent_copying
bg_copy_priority 50
progress 100
freeze_time 2011/10/04/20/42/25
status online
sync
copy_type global
cycle_period_seconds 300
cycling_mode multi
master_change_vdisk_id 5
master_change_vdisk_name GM_App_Pri_CHANGE_VOL
aux_change_vdisk_id 5
aux_change_vdisk_name GM_App_Sec_CHANGE_VOL
```

9.15.26 Stopping the Consistency Group to change the cycling mode

In Example 9-199, we stop the Consistency Group with two relationships, and you must stop it to change Global Mirror to cycling mode. A display shows that the state of the Consistency Group changes to consistent_stopped.

Example 9-199 Stopping the Consistency Group to change the cycling mode

```
IBM_2145:ITSO_SVC1:admin>stoprcconsistgrp CG_W2K3_GM

IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_GM
id 0
name CG_W2K3_GM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary aux
state consistent_stopped
relationship_count 2
freeze_time
status
sync in_sync
copy_type global
cycle_period_seconds 300
cycling_mode none
RC_rel_id 0
```

```
RC_rel_name GMREL1
RC_rel_id 1
RC_rel_name GMREL2
```

9.15.27 Setting the cycling mode on the Consistency Group

In Example 9-200, we change the cycling mode of the Consistency Group `CG_W2K3_GM`. To change it, we need to stop the Consistency Group; otherwise, the command fails.

Example 9-200 Setting the Global Mirror cycling mode on the Consistency Group

```
IBM_2145:ITSO_SVC1:admin>chrcconsistgrp -cyclingmode multi CG_W2K3_GM

IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_GM
id 0
name CG_W2K3_GM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary aux
state consistent_stopped
relationship_count 2
freeze_time
status
sync in_sync
copy_type global
cycle_period_seconds 300
cycling_mode multi
RC_rel_id 0
RC_rel_name GMREL1
RC_rel_id 1
RC_rel_name GMREL2
```

9.15.28 Setting the Change Volume on the master volume relationships of the Consistency Group

In Example 9-201, we change both of the relationships of the Consistency Group to add the Change Volumes on the primary volumes. A display shows the name of the master Change Volumes.

Example 9-201 Setting the Change Volume on the master volume

```
IBM_2145:ITSO_SVC1:admin>chrcrelationship -masterchange GM_DB_Pri_CHANGE_VOL GMREL1

IBM_2145:ITSO_SVC1:admin>lsrcrelationship GMREL1
id 0
name GMREL1
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 0
master_vdisk_name GM_DB_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 0
aux_vdisk_name GM_DB_Sec
primary aux
```

```
consistency_group_id 0
consistency_group_name CG_W2K3_GM
state consistent_stopped
bg_copy_priority 50
progress 100
freeze_time
status online
sync in_sync
copy_type global
cycle_period_seconds 300
cycling_mode multi
master_change_vdisk_id 3
master_change_vdisk_name GM_DB_Pri_CHANGE_VOL
aux_change_vdisk_id
aux_change_vdisk_name
IBM_2145:ITSO_SVC1:admin>

IBM_2145:ITSO_SVC1:admin>chrcrelationship -masterchange GM_DBLog_Pri_CHANGE_VOL GMREL2

IBM_2145:ITSO_SVC1:admin>lsrcrelationship GMREL2
id 1
name GMREL2
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 1
master_vdisk_name GM_DBLog_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 1
aux_vdisk_name GM_DBLog_Sec
primary aux
consistency_group_id 0
consistency_group_name CG_W2K3_GM
state consistent_stopped
bg_copy_priority 50
progress 100
freeze_time
status online
sync in_sync
copy_type global
cycle_period_seconds 300
cycling_mode multi
master_change_vdisk_id 4
master_change_vdisk_name GM_DBLog_Pri_CHANGE_VOL
aux_change_vdisk_id
aux_change_vdisk_name
```

9.15.29 Setting the Change Volumes on the auxiliary volumes

In Example 9-202, we change both of the relationships of the Consistency Group to add the Change Volumes to the secondary volumes. The display shows the names of the auxiliary Change Volumes.

Example 9-202 Setting the Change Volumes on the auxiliary volumes

```
IBM_2145:ITSO_SVC4:admin>chrcrelationship -auxchange GM_DB_Sec_CHANGE_VOL GMREL1
IBM_2145:ITSO_SVC4:admin>lsrcrelationship GMREL1
id 0
name GMREL1
```

```
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 0
master_vdisk_name GM_DB_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 0
aux_vdisk_name GM_DB_Sec
primary aux
consistency_group_id 0
consistency_group_name CG_W2K3_GM
state consistent_stopped
bg_copy_priority 50
progress 100
freeze_time
status online
sync in_sync
copy_type global
cycle_period_seconds 300
cycling_mode multi
master_change_vdisk_id 3
master_change_vdisk_name GM_DB_Pri_CHANGE_VOL
aux_change_vdisk_id 3
aux_change_vdisk_name GM_DB_Sec_CHANGE_VOL

IBM_2145:ITSO_SVC4:admin>chrcrelationship -auxchange GM_DBLog_Sec_CHANGE_VOL GMREL2
IBM_2145:ITSO_SVC4:admin>lsrcrelationship GMREL2
id 1
name GMREL2
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
master_vdisk_id 1
master_vdisk_name GM_DBLog_Pri
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
aux_vdisk_id 1
aux_vdisk_name GM_DBLog_Sec
primary aux
consistency_group_id 0
consistency_group_name CG_W2K3_GM
state consistent_stopped
bg_copy_priority 50
progress 100
freeze_time
status online
sync in_sync
copy_type global
cycle_period_seconds 300
cycling_mode multi
master_change_vdisk_id 4
master_change_vdisk_name GM_DBLog_Pri_CHANGE_VOL
aux_change_vdisk_id 4
aux_change_vdisk_name GM_DBLog_Sec_CHANGE_VOL
```

9.15.30 Starting the Consistency Group CG_W2K3_GM in the cycling mode

In Example 9-203, we start the Consistency Group in the cycling mode. Looking at the field **freeze_time**, you can see that the Consistency Group has been started in the cycling mode, and it is taking consistency images.

Example 9-203 Starting the Consistency Group with cycling mode

```
IBM_2145:ITSO_SVC1:admin>startrcconsistgrp CG_W2K3_GM
IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_GM
id 0
name CG_W2K3_GM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary aux
state consistent_copying
relationship_count 2
freeze_time 2011/10/04/21/02/33
status
sync
copy_type global
cycle_period_seconds 300
cycling_mode multi
RC_rel_id 0
RC_rel_name GMREL1
RC_rel_id 1
RC_rel_name GMREL2

IBM_2145:ITSO_SVC1:admin>lsrcconsistgrp CG_W2K3_GM
id 0
name CG_W2K3_GM
master_cluster_id 000002006BE04FC4
master_cluster_name ITSO_SVC1
aux_cluster_id 0000020061C06FCA
aux_cluster_name ITSO_SVC4
primary aux
state consistent_copying
relationship_count 2
freeze_time 2011/10/04/21/07/42
status
sync
copy_type global
cycle_period_seconds 300
cycling_mode multi
RC_rel_id 0
RC_rel_name GMREL1
RC_rel_id 1
RC_rel_name GMREL2
```

9.16 Service and maintenance

In this section, we describe the various service and maintenance tasks that you can execute within the SAN Volume Controller environment.

9.16.1 Upgrading software

In this section, we explain how to upgrade the SAN Volume Controller software.

Package numbering and version

The format for software upgrade packages is four positive integers that are separated by periods. For example, a software upgrade package is similar to 7.3.0.0, and each software package is given a unique number. Check the recommended software levels at this website:

http://www.ibm.com/systems/storage/software/virtualization/svc/index.html

SAN Volume Controller software upgrade test utility

The SAN Volume Controller Software Upgrade Test Utility, which resides on the Master Console, checks the software levels in the system against the recommended levels, which are documented on the support website. You will be informed if the software levels are current or if you need to download and install newer levels. You can download the utility and installation instructions from this link:

http://www.ibm.com/support/docview.wss?rs=591&uid=ssg1S4000585

After the software file has been uploaded to the system (to the `/home/admin/upgrade` directory), you can select the software and apply it to the system. Use the web script and the **`applysoftware`** command. When a new code level is applied, it is automatically installed on all of the nodes within the system.

The underlying command-line tool runs the **`sw_preinstall`** script. This script checks the validity of the upgrade file and whether it can be applied over the current level. If the upgrade file is unsuitable, the preinstall script deletes the files, which prevents the buildup of invalid files on the system.

Precaution before you perform the upgrade

Software installation is normally considered to be a client's task. The SAN Volume Controller supports concurrent software upgrade. You can perform the software upgrade concurrently with I/O user operations and certain management activities, but only limited CLI commands will be operational from the time that the install command starts until the upgrade operation has either terminated successfully or been backed out. Certain commands will fail with a message indicating that a software upgrade is in progress.

Before you upgrade the SAN Volume Controller software, ensure that all I/O paths between all hosts and SANs are working. Otherwise, the applications might have I/O failures during the software upgrade. Ensure that all I/O paths between all hosts and SANs are working by using the Subsystem Device Driver (SDD) command. Example 9-204 shows the output.

Example 9-204 Query adapter

```
#datapath query adapter
Active Adapters :2

Adpt#     Name   State     Mode             Select     Errors  Paths  Active
    0   fscsi0  NORMAL   ACTIVE              1445          0      4       4
    1   fscsi1  NORMAL   ACTIVE              1888          0      4       4

#datapath query device
Total Devices : 2

DEV#:   0  DEVICE NAME: vpath0  TYPE: 2145           POLICY:    Optimized
SERIAL: 60050768018201BF2800000000000000
```

```
==========================================================================
Path#      Adapter/Hard Disk          State     Mode     Select     Errors
    0          fscsi0/hdisk3           OPEN   NORMAL          0          0
    1          fscsi1/hdisk7           OPEN   NORMAL        972          0

DEV#:   1  DEVICE NAME: vpath1  TYPE: 2145            POLICY:    Optimized
SERIAL: 60050768018201BF2800000000000002
==========================================================================
Path#      Adapter/Hard Disk          State     Mode     Select     Errors
    0          fscsi0/hdisk4           OPEN   NORMAL        784          0
    1          fscsi1/hdisk8           OPEN   NORMAL          0          0
```

Write-through mode: During a software upgrade, there are periods when not all of the nodes in the system are operational, and as a result, the cache operates in write-through mode. Note that write-through mode has an effect on the throughput, latency, and bandwidth aspects of performance.

Verify that your uninterruptible power supply unit configuration is also set up correctly (even if your system is running without problems). Specifically, make sure that the following conditions are true:

- Your uninterruptible power supply units are all getting their power from an external source, and they are not daisy chained. Make sure that each uninterruptible power supply unit is *not* supplying power to another node's uninterruptible power supply unit.
- The power cable and the serial cable, which come from each node, go back to the *same* uninterruptible power supply unit. If the cables are crossed and go back to separate uninterruptible power supply units, during the upgrade, while one node is shut down, another node might also be shut down mistakenly.

Important: Do not share the SAN Volume Controller uninterruptible power supply unit with any other devices.

You must also ensure that all I/O paths are working for each host that runs I/O operations to the SAN during the software upgrade. You can check the I/O paths by using the **`datapath query`** commands.

You do not need to check for hosts that have no active I/O operations to the SAN during the software upgrade.

Upgrade procedure

To upgrade the SAN Volume Controller system software, perform the following steps:

1. Before starting the upgrade, you must back up the configuration (see 9.17, "Backing up the SAN Volume Controller system configuration" on page 319) and save the backup config file in a safe place.
2. Before starting to transfer the software code to the clustered system, clear the previously uploaded upgrade files in the `/home/admin/upgrade` SAN Volume Controller system directory, as shown in Example 9-205.

Example 9-205 cleardumps -prefix /home/admin/upgrade command

```
IBM_2145:ITSO_SVC1:admin>cleardumps -prefix /home/admin/upgrade
IBM_2145:ITSO_SVC1:admin>
```

3. Save the data collection for support diagnosis in case of problems, as shown in Example 9-206.

Example 9-206 svc_snap -c command

```
IBM_2145:ITSO_SVC1:admin>svc_snap -c
Collecting system information...
Creating Config Backup
Dumping error log...
Creating
Snap data collected in /dumps/snap.110711.111003.111031.tgz
```

4. List the dump that was generated by the previous command, as shown in Example 9-207.

Example 9-207 lsdumps command

```
IBM_2145:ITSO_SVC1:admin>lsdumps
id filename
0  svc.config.cron.bak_108283
1  sel.110711.trc
2  rtc.race_mq_log.txt.110711.trc
3  ethernet.110711.trc
4  svc.config.cron.bak_110711
5  svc.config.cron.xml_110711
6  svc.config.cron.log_110711
7  svc.config.cron.sh_110711
8  svc.config.backup.bak_110711
9  svc.config.backup.tmp.xml
10 110711.trc
11 svc.config.backup.xml_110711
12 svc.config.backup.now.xml
13 snap.110711.111003.111031.tgz
```

5. Save the generated dump in a safe place using the **pscp** command, as shown in Example 9-208 on page 307.

> **Note:** The **pscp** command will not work if you have not uploaded your PuTTY SSH private key or if you are not using the user ID and password with the PuTTY **pageant** agent, as shown in Figure 9-14 on page 307.

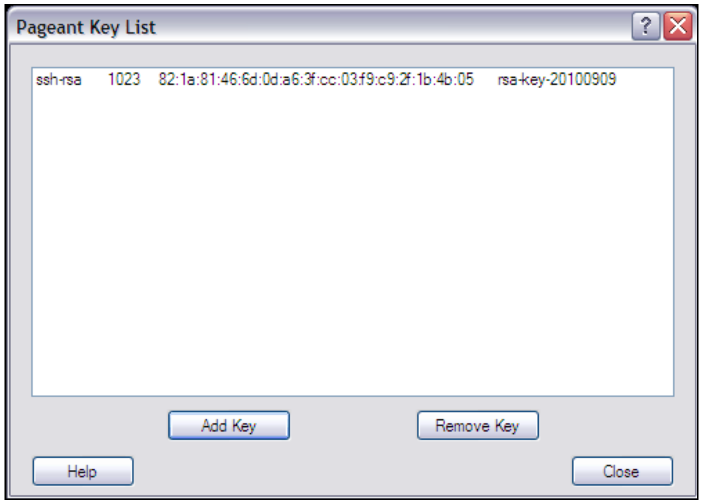


Figure 9-14 Pageant example

Example 9-208 pscp -load command

```
C:\Program Files (x86)\PuTTY>pscp -load ITSO_SVC1
admin@10.18.228.173:/dumps/snap
.110711.111003.111031.tgz c:snap.110711.111003.111031.tgz
snap.110711.111003.111031 | 4999 kB | 4999.8 kB/s | ETA: 00:00:00 |
100%
```

6. Upload the new software package using PuTTY Secure Copy. Enter the command, as shown in Example 9-209.

Example 9-209 pscp -load command

```
C:\Program Files (x86)\PuTTY>pscp -load ITSO_SVC1 c:\IBM2145_INSTALL_7.3.0.0.
admin@10.18.228.81:/home/admin/upgrade
110926.tgz.gpg            | 353712 kB | 11053.5 kB/s | ETA: 00:00:00 | 100%
```

7. Upload the SAN Volume Controller Software Upgrade Test Utility by using PuTTY Secure Copy. Enter the command, as shown in Example 9-210.

Example 9-210 Upload utility

```
C:\>pscp -load ITSO_SVC1 IBM2145_INSTALL_svcupgradetest_7.3.0.0
admin@10.18.229.81:/home/admin/upgrade
IBM2145_INSTALL_svcupgrad | 11 kB |  12.0 kB/s | ETA: 00:00:00 | 100%
```

8. Verify that the packages were successfully delivered through the PuTTY command-line application by entering the **lsdumps** command, as shown in Example 9-211.

Example 9-211 lsdumps command

```
IBM_2145:ITSO_SVC1:admin>lsdumps -prefix /home/admin/upgrade
id filename
0  IBM2145_INSTALL_7.3.0.0.
1  IBM2145_INSTALL_svcupgradetest_7.3
```

9. Now that the packages are uploaded, install the SAN Volume Controller Software Upgrade Test Utility, as shown in Example 9-212.

Example 9-212 applysoftware command

```
IBM_2145:ITSO_SVC1:admin>applysoftware -file IBM2145_INSTALL_svcupgradetest_7.3
CMMVC6227I The package installed successfully.
```

10. Using the following command, test the upgrade for known issues that might prevent a software upgrade from completing successfully, as shown in Example 9-213.

Example 9-213 svcupgradetest command

```
IBM_2145:ITSO_SVC1:admin>svcupgradetest -v 7.3.0.0
svcupgradetest version 7.2 Please wait while the tool tests
for issues that may prevent a software upgrade from completing
successfully. The test will take approximately one minute to complete.
The test has not found any problems with the 2145 cluster.
Please proceed with the software upgrade.
```

Important: If the **svcupgradetest** command produces any errors, troubleshoot the errors using the maintenance procedures before continuing.

11. Use the **applysoftware** command to apply the software upgrade, as shown in Example 9-214.

Example 9-214 Applysoftware upgrade command example

```
IBM_2145:ITSO_SVC1:admin>applysoftware -file IBM2145_INSTALL_7.3.0.0
```

While the upgrade runs, you can check the status as shown in Example 9-215.

Example 9-215 Checking the update status

```
IBM_2145:ITSO_SVC1:admin>lssoftwareupgradestatus
status
upgrading
```

12. The new code is distributed and applied to each node in the SAN Volume Controller system. After installation, each node is automatically restarted one at a time. If a node does not restart automatically during the upgrade, you must repair it manually.
13. Eventually, both nodes display `Cluster:` on line one on the SAN Volume Controller front panel and the name of your system on line two of the panel. Be prepared for a wait (in our case, we waited approximately 40 minutes).

Performance: During this process, both your CLI and GUI vary from sluggish (slow) to unresponsive. The important thing is that I/O to the hosts can continue throughout this process.

14. To verify that the upgrade was successful, you can perform either of the following options:
    – You can run the **lssystem** and **lsnodevpd** commands, as shown in Example 9-216 on page 309. (We truncated the **lssystem** and **lsnodevpd** information for this example.)

Example 9-216 lssystem and lsnodevpd commands

```
IBM_2145:ITSO_SVC1:admin>lssystem
id 000002006BE04FC4
name ITSO_SVC1
location local
partnership
bandwidth
.
.
cluster_locale en_US
time_zone 520 US/Pacific
code_level 7.3.0.0 (96.3.1406022)
console_IP 10.18.228.81:443
id_alias 000002006BE04FC4
gm_link_tolerance 200
gm_inter_cluster_delay_simulation 20
gm_intra_cluster_delay_simulation 40
gm_max_host_delay 5
.
.
tier_capacity 766.50GB
tier_free_capacity 736.50GB
has_nas_key no
layer appliance

IBM_2145:ITSO_SVC1:admin>lsnodevpd 1
id 1
system board: 23 fields
part_number 31P1090
.
.
software: 4 fields
id 1
node_name SVC1N1
WWNN 0x50050768010027e2
code_level 7.3.0.0 (build 86.6.1310161200)
```

– Or you can check whether the code installation has completed without error by copying the log to your management workstation, as explained in the next section. Open the event log in WordPad and search for the "`Software Install completed`." message.

At this point, you have completed the required tasks to upgrade the SAN Volume Controller software.

9.16.2 Running the maintenance procedures

Use the **`finderr`** command to generate a list of any unfixed errors in the system. This command analyzes the last generated log that resides in the `/dumps/elogs/` directory on the system.

To generate a new log before analyzing unfixed errors, run the **`dumperrlog`** command (Example 9-217).

Example 9-217 dumperrlog command

```
IBM_2145:ITSO_SVC1:admin>dumperrlog
```

This command generates an `errlog_timestamp` file, such as errlog_KD8P1CG_140508_123608, where:

- `errlog` is part of the default prefix for all event log files.
- `KD8P1CG` is the panel name of the current configuration node.
- `111003` is the date (*YYMMDD*).
- `090500` is the time (*HHMMSS*).

You can add the **`-prefix`** parameter to your command to change the default prefix of `errlog` to something else (Example 9-218).

Example 9-218 dumperrlog -prefix command

```
IBM_2145:ITSO_SVC1:admin>dumperrlog -prefix ITSO_SVC1_errlog
```

This command creates a file called `ITSO_SVC1_errlog_110711_111003_141111`.

To see the file name, enter the following command (Example 9-219).

Example 9-219 lsdumps command

```
IBM_2145:ITSO_SVC1:admin>IBM_2145:ITSO_SVC2:superuser>lsdumps -prefix /dumps/elogs
id filename
0  errlog_KD8P1BP_140520_202738
1  errlog_KD8P1BP_140704_030209
2  errlog_KD8P1CG_140508_123608
```

Maximum number of event log dump files: A maximum of ten event log dump files per node will be kept on the system. When the eleventh dump is made, the oldest existing dump file for that node will be overwritten. Note that the directory might also hold log files that are retrieved from other nodes. These files are *not* counted.

The SAN Volume Controller deletes the oldest file (when necessary) for this node to maintain the maximum number of files. The SAN Volume Controller will *not* delete files from other nodes unless you issue the **`cleardumps`** command.

After you generate your event log, you can issue the **`finderr`** command to scan the event log for any unfixed events, as shown in Example 9-220.

Example 9-220 finderr command

```
IBM_2145:ITSO_SVC1:admin>finderr
Highest priority unfixed error code is [1550]
```

As you can see, we have one unfixed event on our system. To analyze this event, we download it onto our personal computer. To know more about this unfixed event, we look at the event log in more detail. We use the PuTTY Secure Copy process to copy the file from the system to our local management workstation, as shown in Example 9-221.

Example 9-221 pscp command: Copy event logs off the SVC

In W2K3 → **Start** → **Run** → cmd

```
C:\Program Files (x86)\PuTTY>pscp -load ITSO_SVC1 admin@10.18.228.81:/dumps/elog
s/ITSO_SVC1_errlog_110711_111003_141111 c:\ITSO_SVC1_errlog_110711_111003_141111

ITSO_SVC1_errlog_110711_1 | 6 kB |   6.8 kB/s | ETA: 00:00:00 | 100%
```

```
C:\Program Files (x86)\PuTTY>
```

To use the **Run** option, you must know where your `pscp.exe` file is located. In our case, it is in the `C:\Program Files (x86)\PuTTY>` folder.

This command copies the file called `ITSO_SVC1_errlog_110711_111003_141111` to the `C:\` directory on our local workstation and calls the file `ITSO_SVC1_errlog_110711_111003_141111`. Open the file in WordPad. (Notepad does not format the window as well.) You will see information that is similar to the information that is shown in Example 9-222. (We truncated this list for the purposes of this example.)

Example 9-222 errlog in WordPad

```
//-------------------
// Error Log Entries
//-------------------


Error Log Entry 0
  Node Identifier       : SVC1N2
  Object Type           : node
  Object ID             : 2
  Copy ID               :
  Sequence Number       : 101
  Root Sequence Number  : 101
  First Error Timestamp : Mon Oct  3 10:50:13 2011
                        : Epoch + 1317664213
  Last Error Timestamp  : Mon Oct  3 10:50:13 2011
                        : Epoch + 1317664213
  Error Count           : 1
  Error ID              : 980221 : Error log cleared
  Error Code            :
  Status Flag           : SNMP trap raised
  Type Flag             : INFORMATION

00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
```

By scrolling through the list, or searching for the term *unfixed*, you can find more detail about the problem. You might see more entries in the errorlog that have the status of `unfixed`.

After rectifying the problem, you can mark the event as fixed in the log by issuing the **`cherrstate`** command against its sequence number; see Example 9-223.

Example 9-223 cherrstate command

```
IBM_2145:ITSO_SVC1:admin>cherrstate -sequencenumber 106
```

If you accidentally mark the wrong event as `fixed`, you can mark it as `unfixed` again by entering the same command and appending the **`-unfix`** flag to the end, as shown in Example 9-224.

Example 9-224 -unfix flag

```
IBM_2145:ITSO_SVC1:admin>cherrstate -sequencenumber 106 -unfix
```

9.16.3 Setting up SNMP notification

To set up event notification, use the **`mksnmpserver`** command. Example 9-225 shows an example of the **`mksnmpserver`** command.

Example 9-225 mksnmpserver command

```
IBM_2145:ITSO_SVC1:admin>mksnmpserver -error on -warning on -info on  -ip
9.43.86.160 -community SVC
SNMP Server id [0] successfully created
```

This command sends *all* events and warnings to the SAN Volume Controller community on the SNMP manager with the IP address `9.43.86.160`.

9.16.4 Setting the syslog event notification

You can save a syslog to a defined syslog server because the SAN Volume Controller provides support for syslog in addition to email and SNMP traps.

The syslog protocol is a client/server standard for forwarding log messages from a sender to a receiver on an IP network. You can use syslog to integrate log messages from various types of systems into a central repository. You can configure SAN Volume Controller to send information to six syslog servers.

You use the **`mksyslogserver`** command to configure the SAN Volume Controller using the CLI, as shown in Example 9-226.

Using this command with the **`-h`** parameter gives you information about all of the available options. In our example, we only configure the SAN Volume Controller to use the default values for our syslog server.

Example 9-226 Configuring the syslog

```
IBM_2145:ITSO_SVC1:admin>mksyslogserver -ip 10.64.210.231 -name Syslogserv1
Syslog Server id [0] successfully created
```

When we have configured our syslog server, we can display the current syslog server configurations in our system, as shown in Example 9-227.

Example 9-227 lssyslogserver command

```
IBM_2145:ITSO_SVC1:admin>lssyslogserver
id name        IP_address    facility error warning info
0  Syslogserv1 10.64.210.231 0        on    on      on
1               Syslogserv1 10.64.210.231                         0
on             on             on
```

9.16.5 Configuring error notification using an email server

The SAN Volume Controller can use an email server to send event notification and inventory emails to email users. It can transmit any combination of events, warning, and informational notification types. The SAN Volume Controller supports up to six email servers to provide redundant access to the external email network. The SAN Volume Controller uses the email servers in sequence until the email is successfully sent from the SAN Volume Controller.

> **Important:** Before the SAN Volume Controller can start sending emails, we must run the **`startemail`** command, which enables this service.

The attempt is successful when the SAN Volume Controller gets a positive acknowledgment from an email server that the email has been received by the server.

If no port is specified, port `25` is the default port, as shown in Example 9-228.

Example 9-228 The mkemailserver command syntax

```
IBM_2145:ITSO_SVC1:admin>mkemailserver -ip 192.168.1.1
Email Server id [0] successfully created
IBM_2145:ITSO_SVC1:admin>lsemailserver 0
id 0
name emailserver0
IP_address 192.168.1.1
port 25
```

We can configure an email user that will receive email notifications from the SAN Volume Controller system. We can define 12 users to receive emails from our SAN Volume Controller.

Using the **`lsemailuser`** command, we can verify which user is already registered and what type of information is sent to that user, as shown in Example 9-229.

Example 9-229 lsemailuser command

```
IBM_2145:ITSO_SVC1:admin>lsemailuser
id              name              address
user_type      error          warning        info           inventory
0              IBM_Support_Center callhome0@de.ibm.com
support        on             off            off            on
```

We can also create a new user, as shown in Example 9-230, for a SAN administrator.

Example 9-230 mkemailuser command

```
IBM_2145:ITSO_SVC1:admin>mkemailuser -address SANadmin@ibm.com -error on -warning
on -info on -inventory on
User, id [0], successfully created
```

9.16.6 Analyzing the event log

The following types of events are logged in the event log. An *event* is an occurrence of significance to a task or system. Events can include the completion or failure of an operation, a user action, or a change in the state of a process.

Node event codes now have two classifications:

- Critical events: Critical events put the node into the service state and prevent the node from joining the system. The critical events are numbered 500 - 699.

  > **Deleting a node:** Deleting a node from a system will cause the node to enter the service state, as well.

- Non-critical events: Non-critical events are partial hardware faults, for example, one power-supply unit (PSU) failed in the 2145-CF8. The non-critical events are numbered 800 - 899.

To display the event log, use the **lseventlog** command, as shown in Example 9-231.

Example 9-231 lseventlog command used to display the event log

```
IBM_2145:ITSO_SVC1:admin>lseventlog -count 2
sequence_number last_timestamp object_type object_id object_name copy_id status  fixed
event_id error_code description
102             111003105018   cluster               ITSO_SVC1          message no
981004              FC discovery occurred, no configuration changes were detected
103             111003111036   cluster               ITSO_SVC1          message no
981004              FC discovery occurred, no configuration changes were detected

IBM_2145:ITSO_SVC1:admin>lseventlog 103
sequence_number 103
first_timestamp 111003111036
first_timestamp_epoch 1317665436
last_timestamp 111003111036
last_timestamp_epoch 1317665436
object_type cluster
object_id
object_name ITSO_SVC1
copy_id
reporting_node_id 1
reporting_node_name SVC1N1
root_sequence_number
event_count 1
status message
fixed no
auto_fixed no
notification_type informational
event_id 981004
event_id_text FC discovery occurred, no configuration changes were detected
error_code
error_code_text
sense1 01 01 00 00 7E 0B 00 00 04 02 00 00 01 00 01 00
sense2 00 00 00 00 10 00 00 00 08 00 08 00 00 00 00 00
sense3 00 00 00 00 00 00 00 00 F2 FF 01 00 00 00 00 00
sense4 0E 00 00 00 FC FF FF FF 03 00 00 00 07 00 00 00
sense5 00 00 06 00 00 00 00 00 00 00 00 00 00 00 00 00
sense6 00 00 00 00 03 00 00 00 00 00 00 00 00 00 00 00
sense7 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
```

```
sense8 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
```

These commands allow you to view the last events (you can specify the ***-count*** parameter to define how many events you need to display) that were generated. Use the method that is described in 9.16.2, "Running the maintenance procedures" on page 309 to upload and analyze the event log in more detail.

To clear the event log, you can issue the **clearerrlog** command, as shown in Example 9-232.

Example 9-232 clearerrlog command

```
IBM_2145:ITSO_SVC1:admin>clearerrlog
Do you really want to clear the log?  y
```

Using the **-force** flag will stop any confirmation requests from appearing. When executed, this command clears all of the entries from the event log. This process proceeds even if there are unfixed errors in the log. It also clears any status events that are in the log.

Important: This command is a destructive command for the event log. Only use this command when you have either rebuilt the system, or when you have fixed a major problem that has caused many entries in the event log that you do not want to fix manually.

9.16.7 License settings

To change the licensing feature settings, use the **chlicense** command.

Before you change the licensing, you can display the licenses that you already have by issuing the **lslicense** command, as shown in Example 9-233.

Example 9-233 lslicense command

```
IBM_2145:ITSO_SVC1:admin>lslicense
used_flash 0.00
used_remote 0.03
used_virtualization 0.75
license_flash 500
license_remote 500
license_virtualization 500
license_physical_disks 0
license_physical_flash off
license_physical_remote off
```

The current license settings for the system are displayed in the viewing license settings log window. These settings show whether you are licensed to use the FlashCopy, Metro Mirror, Global Mirror, or Virtualization features. The license settings log window also shows the storage capacity that is licensed for virtualization. Typically, the license settings log contains entries because feature options must be set as part of the web-based system creation process.

Consider, for example, that you have purchased an additional 5 TB of licensing for the Metro Mirror and Global Mirror feature from your actual 20 TB license. Example 9-234 shows the command that you enter.

Example 9-234 chlicense command

```
IBM_2145:ITSO_SVC1:admin>chlicense -remote 25
```

To turn a feature off, add 0 TB as the capacity for the feature that you want to disable.

To verify that the changes that you have made are reflected in your SAN Volume Controller configuration, you can issue the **lslicense** command (see Example 9-235).

Example 9-235 lslicense command: Verifying changes

```
IBM_2145:ITSO_SVC1:admin>lslicense
used_flash 0.00
used_remote 0.03
used_virtualization 0.75
license_flash 500
license_remote 25
license_virtualization 500
license_physical_disks 0
license_physical_flash off
license_physical_remote off
```

9.16.8 Listing dumps

Starting with SAN Volume Controller 6.3, a new command is available to list the dumps that were generated over a period of time. You can use **lsdumps** with the **-prefix** parameter to return a list of dumps in the appropriate directory. The command produces a list of the files in the specified directory on the specified node. If no node is specified, the config node is used. If no **-prefix** is set, the files in the /dumps directory are listed.

Error or event dump

The dumps that are contained in the /dumps/elogs directory are dumps of the contents of the event log at the time that the dump was taken. You create an error or event log dump by using the **dumperrlog** command. This command dumps the contents of the error or event log to the /dumps/elogs directory.

If you do not supply a file name prefix, the system uses the default errlog_ file name prefix. The full, default file name is errlog_*NNNNNN_YYMMDD_HHMMSS*. In this file name, *NNNNNN* is the node front panel name. If the command is used with the **-prefix** option, the value that is entered for the **-prefix** is used instead of errlog.

The **lsdumps -prefix** command lists all of the dumps in the /dumps/elogs directory (Example 9-236).

Example 9-236 lsdumps -prefix /dumps/elogs

```
IBM_2145:ITSO_SVC1:admin>lsdumps -prefix /dumps/elogs
id filename
0  errlog_110711_111003_111056
1  testerrorlog_110711_111003_135358
2  ITSO_SVC1_errlog_110711_111003_141111
3  ITSO_SVC1_errlog_110711_111003_141620
```

```
4  errlog_110711_111003_154759
```

Featurization log dump

The dumps that are contained in the `/dumps/feature` directory are dumps of the featurization log. A featurization log dump is created by using the **`dumpinternallog`** command. This command dumps the contents of the featurization log to the `/dumps/feature` directory to a file called `feature.txt`. Only one of these files exists, so every time that the **`dumpinternallog`** command is run, this file is overwritten.

The **`lsdumps -prefix /dumps/feature`** command lists all of the dumps in the `/dumps/feature` directory (Example 9-237).

Example 9-237 lsdumps with -prefix /dumps/feature command

```
IBM_2145:ITSO_SVC1:admin>lsdumps -prefix /dumps/feature
id filename
0  feature.txt
```

I/O trace dump

Dumps that are contained in the `/dumps/iotrace` directory are dumps of I/O trace data. The type of data that is traced depends on the options that are specified by the **`settrace`** command. The collection of the I/O trace data is started by using the **`starttrace`** command. The I/O trace data collection is stopped when the **`stoptrace`** command is used. When the trace is stopped, the data is written to the file.

The file name is *`prefix_NNNNNN_YYMMDD_HHMMSS`*, where *`NNNNNN`* is the node front panel name, and *`prefix`* is the value that is entered by the user for the **`-filename`** parameter in the **`settrace`** command.

The command to list all of the dumps in the `/dumps/iotrace` directory is **`lsdumps -prefix /dumps/iotrace`** (Example 9-238).

Example 9-238 lsdumps with -prefix /dumps/iotrace command

```
IBM_2145:ITSO_SVC1:admin>lsdumps -prefix /dumps/iotrace
id               iotrace_filename
0                tracedump_104643_080624_172208
1                iotrace_104643_080624_172451
```

I/O statistics dump

The dumps that are contained in the `/dumps/iostats` directory are the dumps of the I/O statistics for the disks on the cluster. You create an I/O statistics dump by using the **`startstats`** command. As part of this command, you can specify a time interval at which you want the statistics to be written to the file (the default is `15` minutes). Every time that the time interval is encountered, the I/O statistics that are collected up to that point are written to a file in the `/dumps/iostats` directory.

The file names that are used for storing I/O statistics dumps are `m_stats_`*`NNNNNN_YYMMDD_HHMMSS`* or `v_stats_`*`NNNNNN_YYMMDD_HHMMSS`*, depending on whether the statistics are for MDisks or volumes. In these file names, *`NNNNNN`* is the node front panel name.

The command to list all of the dumps that are in the /dumps/iostats directory is **lsdumps -prefix** (Example 9-239).

Example 9-239 lsdumps -prefix /dumps/iostats command

```
IBM_2145:ITSO_SVC1:admin>lsdumps -prefix /dumps/iostats
id filename
0  Nm_stats_110711_111003_125706
1  Nn_stats_110711_111003_125706
2  Nv_stats_110711_111003_125706
3  Nd_stats_110711_111003_125706
4  Nv_stats_110711_111003_131204
5  Nd_stats_110711_111003_131204
6  Nn_stats_110711_111003_131204
........
```

Software dump

The **lsdumps** command lists the contents of the /dumps directory. The general debug information, software, application dumps, and live dumps are copied into this directory. Example 9-240 shows the command.

Example 9-240 lsdumps command without prefix

```
IBM_2145:ITSO_SVC1:admin>lsdumps
id filename
0  svc.config.cron.bak_108283
1  sel.110711.trc
2  rtc.race_mq_log.txt.110711.trc
3  ethernet.110711.trc
4  svc.config.cron.bak_110711
5  svc.config.cron.xml_110711
6  svc.config.cron.log_110711
7  svc.config.cron.sh_110711
8  svc.config.backup.bak_110711
9  svc.config.backup.tmp.xml
10 110711.trc
11 svc.config.backup.xml_110711
12 svc.config.backup.now.xml
13 snap.110711.111003.111031.tgz
```

Other node dumps

The **lsdumps** commands can accept a node identifier as input (for example, append the node name to the end of any of the node dump commands). If this identifier is not specified, the list of files on the current configuration node is displayed. If the node identifier is specified, the list of files on that node is displayed.

However, files can only be copied from the current configuration node (using PuTTY Secure Copy). Therefore, you must issue the **cpdumps** command to copy the files from a non-configuration node to the current configuration node. Subsequently, you can copy them to the management workstation using PuTTY Secure Copy.

For example, suppose that you discover a dump file and want to copy it to your management workstation for further analysis. In this case, you must first copy the file to your current configuration node.

To copy dumps from other nodes to the configuration node, use the **cpdumps** command.

In addition to the directory, you can specify a file filter. For example, if you specified `/dumps/elogs/*.txt`, all of the files, in the `/dumps/elogs` directory, that end in `.txt` are copied.

> **Wildcards:** The following rules apply to the use of wildcards with the SAN Volume Controller CLI:
>
> - The wildcard character is an asterisk (*).
> - The command can contain a maximum of one wildcard.
> - When you use a wildcard, you must surround the filter entry with double quotation marks (""), for example:
>
>   `>cleardumps -prefix "/dumps/elogs/*.txt"`

Example 9-241 shows an example of the **cpdumps** command.

Example 9-241 cpdumps command

```
IBM_2145:ITSO_SVC1:admin>cpdumps -prefix /dumps/configs n4
```

Now that you have copied the configuration dump file from Node `n4` to your configuration node, you can use PuTTY Secure Copy to copy the file to your management workstation for further analysis.

To clear the dumps, you can run the **cleardumps** command. Again, you can append the node name if you want to clear dumps off a node other than the current configuration node (the default for the **cleardumps** command).

The commands in Example 9-242 clear all logs or dumps from the SAN Volume Controller Node `SVC1N2`.

Example 9-242 cleardumps command

```
IBM_2145:ITSO_SVC1:admin>cleardumps -prefix /dumps SVC1N2
IBM_2145:ITSO_SVC1:admin>cleardumps -prefix /dumps/iostats SVC1N2
IBM_2145:ITSO_SVC1:admin>cleardumps -prefix /dumps/iotrace SVC1N2
IBM_2145:ITSO_SVC1:admin>cleardumps -prefix /dumps/feature SVC1N2
IBM_2145:ITSO_SVC1:admin>cleardumps -prefix /dumps/config SVC1N2
IBM_2145:ITSO_SVC1:admin>cleardumps -prefix /dumps/elog SVC1N2
IBM_2145:ITSO_SVC1:admin>cleardumps -prefix /home/admin/upgrade SVC1N2
```

9.17 Backing up the SAN Volume Controller system configuration

You can back up your system configuration by using the Backing Up a Cluster Configuration window or the CLI **svcconfig** command. In this section, we describe the overall procedure for backing up your system configuration and the conditions that must be satisfied to perform a successful backup.

The **backup** command extracts configuration data from the system and saves it to the `svc.config.backup.xml` file in the `/tmp` directory. This process also produces an `svc.config.backup.sh` file. You can study this file to see what other commands were issued to extract information.

An `svc.config.backup.log` log is also produced. You can study this log for the details of what was done and when it was done. This log also includes information about the other commands that were issued.

Any pre-existing `svc.config.backup.xml` file is archived as the `svc.config.backup.bak` file. The system only keeps one archive. We strongly suggest that you immediately move the `.XML` file and related KEY files (see the following limitations) *off* the system for archiving. Then, erase the files from the `/tmp` directory using the **`svcconfig clear -all`** command.

We further advise that you change all of the objects having default names to *non-default* names. Otherwise, a warning is produced for objects with default names.

Also, the object with the default name is restored with its original name with an `_r` appended. The prefix _(underscore) is reserved for backup and restore command usage. Do not use this prefix in any object names.

Important: The tool backs up logical configuration data only, *not* client data. It does not replace a traditional data backup and restore tool. Instead, this tool supplements a traditional data backup and restore tool with a way to back up and restore the client's configuration.

To provide a complete backup and disaster recovery solution, you must back up both user (non-configuration) data and configuration (non-user) data. After the restoration of the SAN Volume Controller configuration, you must fully restore user (non-configuration) data to the system's disks.

9.17.1 Prerequisites

You must have the following prerequisites in place:

- All nodes must be online.
- No object name can begin with an underscore.
- All objects must have non-default names, that is, names that are not assigned by the SAN Volume Controller.

Although we advise that objects have non-default names at the time that the backup is taken, this prerequisite is not mandatory. Objects with default names are renamed when they are restored.

Example 9-243 shows an example of the **`svcconfig backup`** command.

Example 9-243 svcconfig backup command

```
IBM_2145:ITSO_SVC1:admin>svcconfig backup
..................
CMMVC6130W Cluster ITSO_SVC4 with inter-cluster partnership fully_configured will
not be restored
..................................................................................
.......
CMMVC6155I SVCCONFIG processing completed successfully
```

As you can see in Example 9-243, we received a `CMMVC6130W Cluster ITSO_SVC4 with inter-cluster partnership fully_configured will not be restored` message. This message indicates that individual systems in a multisystem environment will need to be backed up individually.

If recovery is required, recovery will only be performed on the system where the recovery commands are executed.

Example 9-244 shows the **pscp** command.

Example 9-244 pscp command

```
C:\Program Files\PuTTY>pscp -load ITSO_SVC1
admin@10.18.229.81:/tmp/svc.config.backup.xml c:\temp\clibackup.xml
clibackup.xml             | 97 kB |  97.2 kB/s | ETA: 00:00:00 | 100%
```

The following scenario illustrates the value of the configuration backup:

1. Use the **svcconfig** command to create a backup file on the clustered system that contains details about the current system configuration.
2. Store the backup configuration on a form of tertiary storage. You must copy the backup file from the clustered system or it becomes lost if the system crashes.
3. If a sufficiently severe failure occurs, the system might be lost. Both the configuration data (for example, the system definitions of hosts, I/O Groups, MDGs, and MDisks) and the application data on the virtualized disks are lost.

   In this scenario, it is assumed that the application data can be restored from normal client backup procedures. However, before you can perform this restoration, you must reinstate the system as it was configured at the time of the failure. Therefore, you restore the same MDGs, I/O Groups, host definitions, and volumes that existed before the failure. Then, you can copy the application data back onto these volumes and resume operations.
4. Recover the hardware: hosts, SAN Volume Controllers, disk controller systems, disks, and SAN fabric. The hardware and SAN fabric must physically be the same as the hardware and SAN fabric that were used before the failure.
5. Reinitialize the clustered system with the configuration node; the other nodes will be recovered when restoring the configuration.
6. Restore your clustered system configuration using the backup configuration file that was generated before the failure.
7. Restore the data on your volumes using your preferred restoration solution or with help from IBM Support.
8. Resume normal operations.

9.18 Restoring the SAN Volume Controller clustered system configuration

> **Important:** It is extremely important that you always consult IBM Support before you restore the SAN Volume Controller clustered system configuration from the backup. IBM Support can assist you in analyzing the root cause of why the system configuration was lost.

After the **svcconfig restore -execute** command is started, consider any prior user data on the volumes *destroyed.* The user data must be recovered through your usual application data backup and restore process.

See the *IBM TotalStorage Open Software Family SAN Volume Controller: Command-Line Interface User's Guide*, GC27-2287, for more information about this topic.

For a detailed description of the SAN Volume Controller configuration backup and restore functions, see the *IBM TotalStorage Open Software Family SAN Volume Controller: Configuration Guide*, GC27-2286.

9.18.1 Deleting the configuration backup

We describe in detail the tasks that you can perform to delete the configuration backup that is stored in the configuration file directory on the system. Never clear this configuration without having a backup of your configuration that is stored in a separate, secure place.

When using the **`clear`** command, you erase the files in the `/tmp` directory. This command does not clear the running configuration and prevent the system from working, but the command clears all of the configuration backup that is stored in the `/tmp` directory; see Example 9-245.

Example 9-245 svcconfig clear command

```
IBM_2145:ITSO_SVC1:admin>svcconfig clear -all
.
CMMVC6155I SVCCONFIG processing completed successfully
```

9.19 Working with the SAN Volume Controller Quorum MDisks

In this section, we show how to list and change the SAN Volume Controller system Quorum Managed Disks (MDisks).

9.19.1 Listing the SAN Volume Controller Quorum MDisks

To list SAN Volume Controller system Quorum MDisks and view their numbers and status, issue the **`lsquorum`** command, as shown in Example 9-246.

For more information about SAN Volume Controller Quorum Disk planning and configuration, see Chapter 4, "Planning and configuration" on page 57.

Example 9-246 lsquorum command and detail

```
IBM_2145:ITSO_SVC1:admin>lsquorum
quorum_index status id name   controller_id controller_name active object_type override
0            online 1  mdisk1 2             ITSO-DS3500     no     mdisk       no
1            online 0  mdisk0 2             ITSO-DS3500     yes    mdisk       no
2            online 3  mdisk3 2             ITSO-DS3500     no     mdisk       no
IBM_2145:ITSO_SVC1:admin>lsquorum 1
quorum_index 1
status online
id 0
name mdisk0
controller_id 2
controller_name ITSO-DS3500
active yes
object_type mdisk
override no
```

9.19.2 Changing the SAN Volume Controller Quorum Disks

To move one of your SAN Volume Controller Quorum MDisks from one MDisk to another, or from one storage subsystem to another, use the **chquorum** command, as shown in Example 9-247.

Example 9-247 chquorum command

```
IBM_2145:ITSO_SVC1:admin>lsquorum
quorum_index status id name   controller_id controller_name active object_type override
0            online 1  mdisk1 2             ITSO-DS3500     no     mdisk       no
1            online 0  mdisk0 2             ITSO-DS3500     yes    mdisk       no
2            online 3  mdisk3 2             ITSO-DS3500     no     mdisk       no



chquorum -mdisk 9 2



IBM_2145:ITSO_SVC1:admin>lsquorum
quorum_index status id name   controller_id controller_name active object_type override
0            online 1  mdisk1 2             ITSO-DS3500     no     mdisk       no
1            online 0  mdisk0 2             ITSO-DS3500     yes    mdisk       no
2            online 9  mdisk9 3             ITSO-DS5000     no     mdisk       no
```

As you can see in Example 9-247, the quorum index `2` has been moved from `MDisk3` on the ITSO-DS3500 controller to `MDisk9` on the ITSO-DS5000 controller.

9.20 Working with the Service Assistant menu

SAN Volume Controller V6.1 introduced a new method for performing service tasks on the system. In addition to being able to perform service tasks from the front panel, you can now also service a node through an Ethernet connection using either a web browser or the CLI. The web browser runs a new service application called *Service Assistant*. Service Assistant offers almost all of the function that was previously available through the front panel. Now, the function is available from the Ethernet connection with an interface that is easier to use and that you can use remotely from the system.

9.20.1 SAN Volume Controller CLI Service Assistant menu

A set of commands relating to the new method for performing service tasks on the system has been introduced.

Two major command sets are available:

- The **sainfo** command set allows you to query the various components within the SAN Volume Controller environment.
- The **satask** command set allows you to make changes to the various components within the SAN Volume Controller.

When the command syntax is shown, you will see certain parameters in square brackets, for example `[parameter]`, indicating that the parameter is optional in most if not all instances.

Any information that is not in square brackets is required information. You can view the syntax of a command by entering one of the following commands:

- **sainfo -?**: Shows a complete list of information commands
- **satask -?**: Shows a complete list of task commands
- **sainfo commandname -?**: Shows the syntax of information commands
- **satask commandname -?**: Shows the syntax of task commands

Example 9-248 shows the two new sets of commands that are introduced with Service Assistant.

Example 9-248 sainfo and satask commands

```
IBM_2145:ITSO_SVC1:admin>sainfo -h
The following actions are available with this command :
        lscmdstatus
        lsfiles
        lsservicenodes
        lsservicerecommendation
        lsservicestatus
IBM_2145:ITSO_SVC1:admin>satask -h
The following actions are available with this command :
        chenclosurevpd
        chnodeled
        chserviceip
        chwwnn
        cpfiles
        installsoftware
        leavecluster
        mkcluster
        rescuenode
        setlocale
        setpacedccu
        settempsshkey
        snap
        startservice
        stopnode
        stopservice
        t3recovery
```

Important: You must use the **sainfo** and **satask** command sets under the direction of IBM Support. The incorrect use of these commands can lead to unexpected results.

9.21 SAN troubleshooting and data collection

When we encounter a SAN issue, the SAN Volume Controller is often extremely helpful in troubleshooting the SAN because the SAN Volume Controller is at the center of the environment through which the communication travels.

SAN Volume Controller Best Practices and Performance Guidelines, SG24-7521, contains a detailed description of how to troubleshoot and collect data from the SAN Volume Controller:

http://www.redbooks.ibm.com/abstracts/sg247521.html?Open

Use the **lsfabric** command regularly to obtain a complete picture of the devices that are connected and visible from the SAN Volume Controller cluster through the SAN. The

lsfabric command generates a report that displays the FC connectivity between nodes, controllers, and hosts.

Example 9-249 shows the output of an **lsfabric** command.

Example 9-249 lsfabric command

```
IBM_2145:ITSO_SVC1:admin>lsfabric
remote_wwpn      remote_nportid id node_name local_wwpn       local_port local_nportid
state    name        cluster_name    type
5005076801405034 030A00         1  SVC1N1    50050768014027E2 1          030800
active   SVC1N2      ITSO_SVC1       node
5005076801405034 030A00         1  SVC1N1    50050768011027E2 3          030900
active   SVC1N2      ITSO_SVC1       node
5005076801305034 040A00         1  SVC1N1    50050768013027E2 2          040800
active   SVC1N2      ITSO_SVC1       node
5005076801305034 040A00         1  SVC1N1    50050768012027E2 4          040900
active   SVC1N2      ITSO_SVC1       node
50050768012027E2 040900         2  SVC1N2    5005076801305034 2          040A00
active   SVC1N1      ITSO_SVC1       node
50050768012027E2 040900         2  SVC1N2    5005076801205034 4          040B00
active   SVC1N1      ITSO_SVC1       node
500507680120505C 040F00         1  SVC1N1    50050768013027E2 2          040800
active   SVC4N2      ITSO_SVC4       node
500507680120505C 040F00         1  SVC1N1    50050768012027E2 4          040900
active   SVC4N2      ITSO_SVC4       node
500507680120505C 040F00         2  SVC1N2    5005076801305034 2          040A00
active   SVC4N2      ITSO_SVC4       node
500507680120505C 040F00         2  SVC1N2    5005076801205034 4          040B00
active   SVC4N2      ITSO_SVC4       node
50050768013027E2 040800         2  SVC1N2    5005076801305034 2          040A00
active   SVC1N1      ITSO_SVC1       node
....
Above and below rows has been removed for brevity
....
20690080E51B09E8 041900         1  SVC1N1    50050768013027E2 2          040800
inactive ITSO-DS3500                 controller
20690080E51B09E8 041900         1  SVC1N1    50050768012027E2 4          040900
inactive ITSO-DS3500                 controller
20690080E51B09E8 041900         2  SVC1N2    5005076801305034 2          040A00
inactive ITSO-DS3500                 controller
20690080E51B09E8 041900         2  SVC1N2    5005076801205034 4          040B00
inactive ITSO-DS3500                 controller
50050768013037DC 041400         1  SVC1N1    50050768013027E2 2          040800
active   ITSOSVC3N1  ITSO_SVC3       node
50050768013037DC 041400         1  SVC1N1    50050768012027E2 4          040900
active   ITSOSVC3N1  ITSO_SVC3       node
50050768013037DC 041400         2  SVC1N2    5005076801305034 2          040A00
active   ITSOSVC3N1  ITSO_SVC3       node
50050768013037DC 041400         2  SVC1N2    5005076801205034 4          040B00
active   ITSOSVC3N1  ITSO_SVC3       node
5005076801101D1C 031500         1  SVC1N1    50050768014027E2 1          030800
active   ITSOSVC3N2  ITSO_SVC3       node
5005076801101D1C 031500         1  SVC1N1    50050768011027E2 3          030900
active   ITSOSVC3N2  ITSO_SVC3       node
5005076801101D1C 031500         2  SVC1N2    5005076801405034 1          030A00
active   ITSOSVC3N2  ITSO_SVC3       node
.....
Above and below rows has been removed for brevity
.....
```

```
5005076801201D22 021300        1  SVC1N1    50050768013027E2 2          040800
active   SVC2N2      ITSO_SVC2       node
5005076801201D22 021300        1  SVC1N1    50050768012027E2 4          040900
active   SVC2N2      ITSO_SVC2       node
5005076801201D22 021300        2  SVC1N2    5005076801305034 2          040A00
active   SVC2N2      ITSO_SVC2       node
5005076801201D22 021300        2  SVC1N2    5005076801205034 4          040B00
active   SVC2N2      ITSO_SVC2       node
50050768011037DC 011513        1  SVC1N1    50050768014027E2 1          030800
active   ITSOSVC3N1  ITSO_SVC3       node
50050768011037DC 011513        1  SVC1N1    50050768011027E2 3          030900
active   ITSOSVC3N1  ITSO_SVC3       node
50050768011037DC 011513        2  SVC1N2    5005076801405034 1          030A00
active   ITSOSVC3N1  ITSO_SVC3       node
50050768011037DC 011513        2  SVC1N2    5005076801105034 3          030B00
active   ITSOSVC3N1  ITSO_SVC3       node
5005076801301D22 021200        1  SVC1N1    50050768013027E2 2          040800
active   SVC2N2      ITSO_SVC2       node
5005076801301D22 021200        1  SVC1N1    50050768012027E2 4          040900
active   SVC2N2      ITSO_SVC2       node
....
Above and below rows have been removed for brevity
....
```

For more information about the **lsfabric** command, see the *IBM System Storage SAN Volume Controller and Storwize V7000 Command-Line Interface User's Guide Version 6.3.0*, GC27-2287.

9.22 T3 recovery process

A procedure, which is known as *T3 recovery*, has been tested and used in select cases where a system has been completely destroyed. (One example is simultaneously pulling power cords from all nodes to their uninterruptible power supply units. In this case, all nodes boot up to node error 578 when the power is restored.)

This procedure, in certain circumstances, is able to recover most user data. However, it is *not* to be used by the client or IBM service representative without the direct involvement from IBM Level 3 technical support. This procedure is not published, but we refer to it here only to indicate that the loss of a system can be recoverable without total data loss. However, it requires a restoration of application data from the backup. T3 recovery is an extremely sensitive procedure that is only to be used as a last resort, and it cannot recover any data that was destaged from cache at the time of the total system failure.

10

SAN Volume Controller operations using the GUI

In this chapter, we illustrate IBM SAN Volume Controller operational management and system administration using the IBM SAN Volume Controller graphical user interface (GUI). The IBM SAN Volume Controller management GUI is an easy-to-use tool that helps you to monitor, manage, and configure your system.

The information is presented at a high level as this book is based on the new hardware, and is not intended to be an in-depth deep dive of the software. For more detailed information about using the GUI, refer to *Implementing the IBM System Storage SAN Volume Controller V7.2*, SG24-7933.

Although the Storage Tier Advisory Tool is not part of the GUI, it is a strong and useful tool to determine the use of your tiered storage, as the 7.3 code level now supports 3 tiered storage using the Easy Tier functionality.

Important: It is possible for more than one user to be logged in to the GUI at any given time. However, no locking mechanism exists, so be aware that if two users change the same object at the same time, the last action entered from the GUI is the one that will take effect.

Data entries made through the GUI are case-sensitive.

10.1 Introduction to IBM SAN Volume Controller operations using the GUI

The IBM SAN Volume Controller Overview panel is the first panel that you will see after you have logged in to the system. It is an important user interface and throughout this chapter we refer to it as IBM SAN Volume Controller Overview panel or just Overview panel.

Initially you will start at the login window, as shown in Figure 10-1.

Figure 10-1 Login window

Dynamic menu

From any page inside the IBM SAN Volume Controller GUI, you always have access to the dynamic menu. The IBM SAN Volume Controller GUI dynamic menu is located on the left side of the IBM SAN Volume Controller GUI panel. To navigate using this menu, move the mouse cursor over the various icons and choose a page that you want to display.

The IBM SAN Volume Controller dynamic menu consists of multiple panels. These panels group together common configuration and administration objects, and present individual administrative objects to the IBM SAN Volume Controller GUI users.

10.1.1 Overview

After you have successfully logged in to the system, you will start at the overview panel as shown in Figure 10-2. The overview will provide the user with a quick summary of the system, as well as a link to the SAN Volume Controller IBM Knowledge Center:

http://www-01.ibm.com/support/knowledgecenter/STPVGU/landing/SVC_welcome.html?lang=en

Figure 10-2 Overview

10.1.2 Monitoring

Figure 10-3 shows the Monitoring menu where you can work with the following details:

- Information about the code level
- Hardware configuration: See installed hardware, change memory allocation also known as bitmap allocation.
- Events: See warnings, alerts, and run the maintenance procedure
- Real-time performance graphs: See CPU usage IOPS for volumes, MDisks, and so on.

Figure 10-3 Monitoring

10.1.3 Pools

In the Pools menu, shown in Figure 10-4, you can administer everything that is related to storage pools, from creating new storage pools and working with internal or external storage.

Storage Pool Balancing is introduced in code level 7.3, which means that if you add new or additional MDisks to an existing pool, it balances the extents across all of the MDisks in a pool. Before release 7.3, you had to do this manually, or use a script to balance extents after adding new MDisks to an existing pool. Note this is an automated process and it is not configurable.

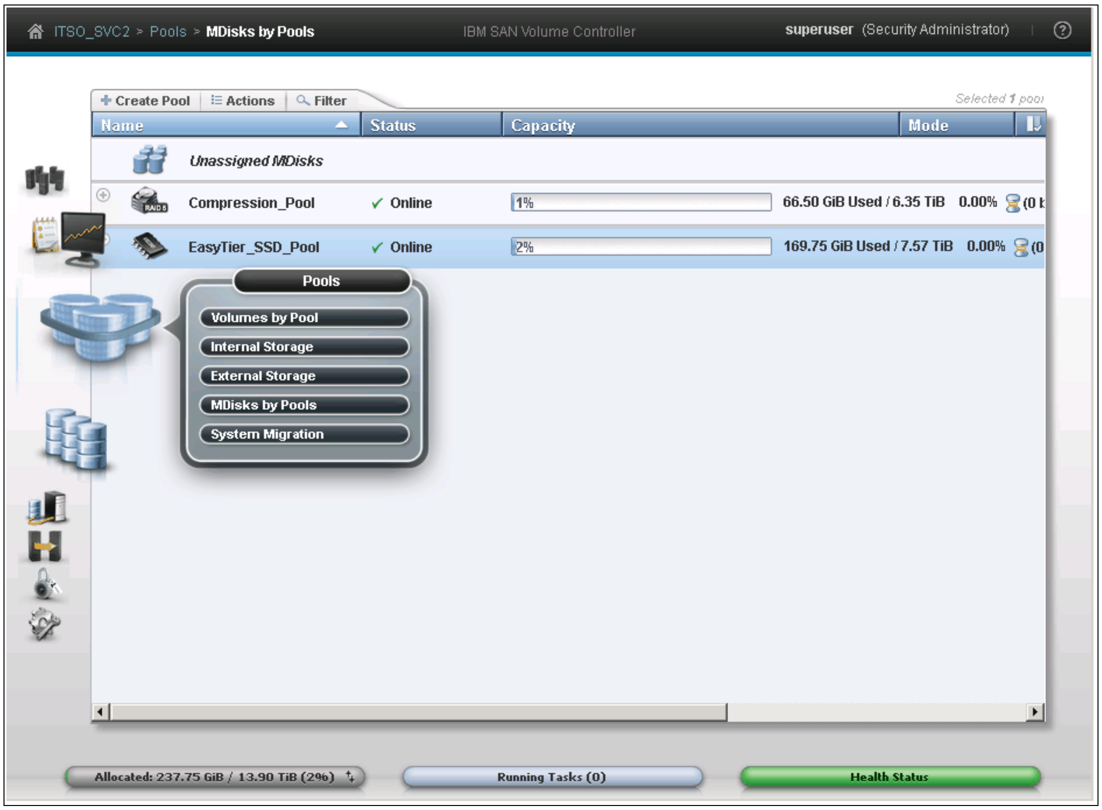


Figure 10-4 Pools

10.1.4 Volumes

The Volumes menu contains the following administrative options:

- View volumes, create volumes, delete volumes.
- See details about volumes, if they are mapped or unmapped to a host.
- See details about volumes mapped to host.

Figure 10-5 shows the Volume menu expanded.

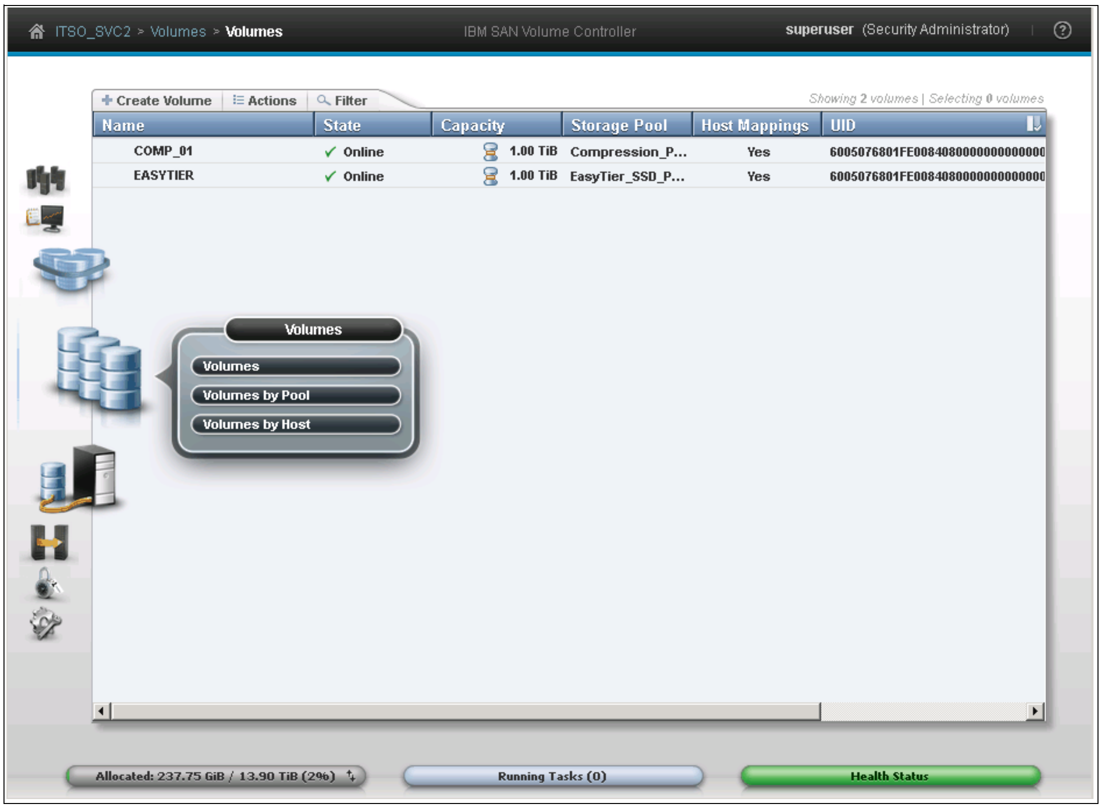


Figure 10-5 Volumes

10.1.5 Hosts

All host-related administration can be executed from this menu, such as:

- Creating and deleting hosts
- View Details about each host, such as ports, worldwide port names.
- Volume mappings for each host.

Figure 10-6 shows the expanded view of the Hosts menu.

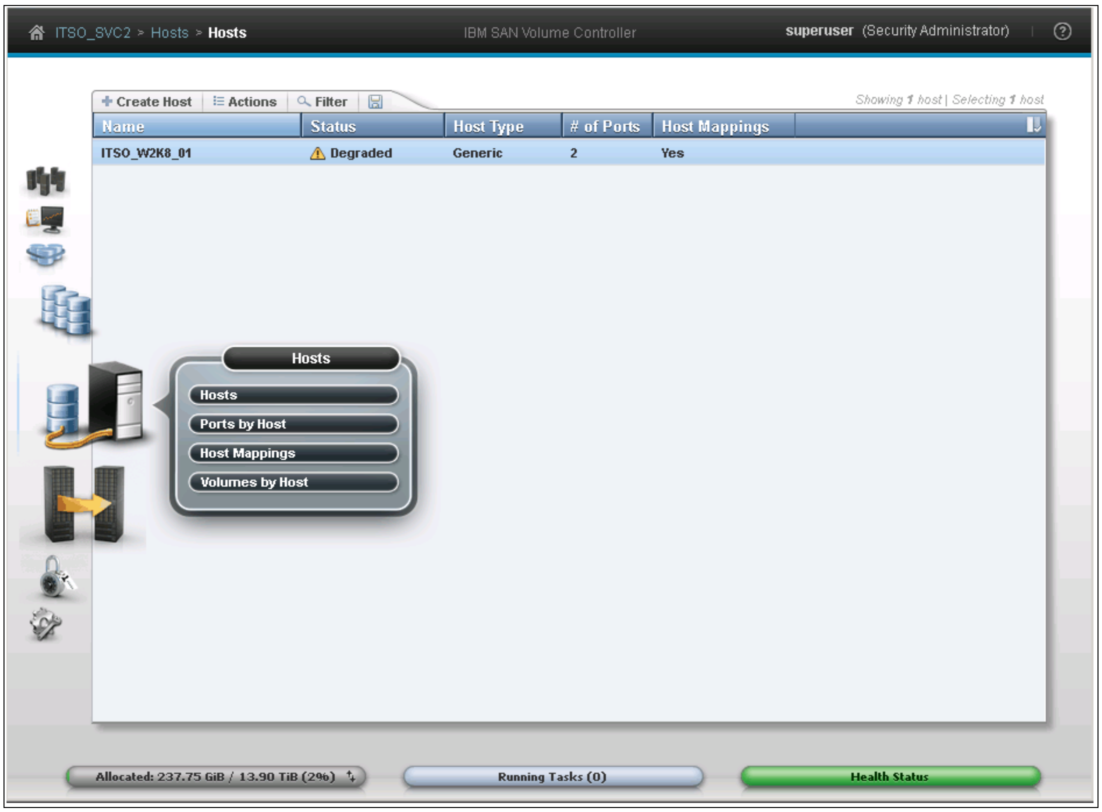


Figure 10-6 Hosts

10.1.6 Copy Services

In the Copy Services menu, you can perform administration of all Copy Services-related activities, such as:

- Create partnerships with other SVC and Storwize systems
- Create and delete Metro Mirrored volumes
- Create and delete Global Mirrored volumes
- Create and delete FlashCopy volumes
- View details about the copy services configured

Figure 10-7 shows the expanded Copy Services menu.

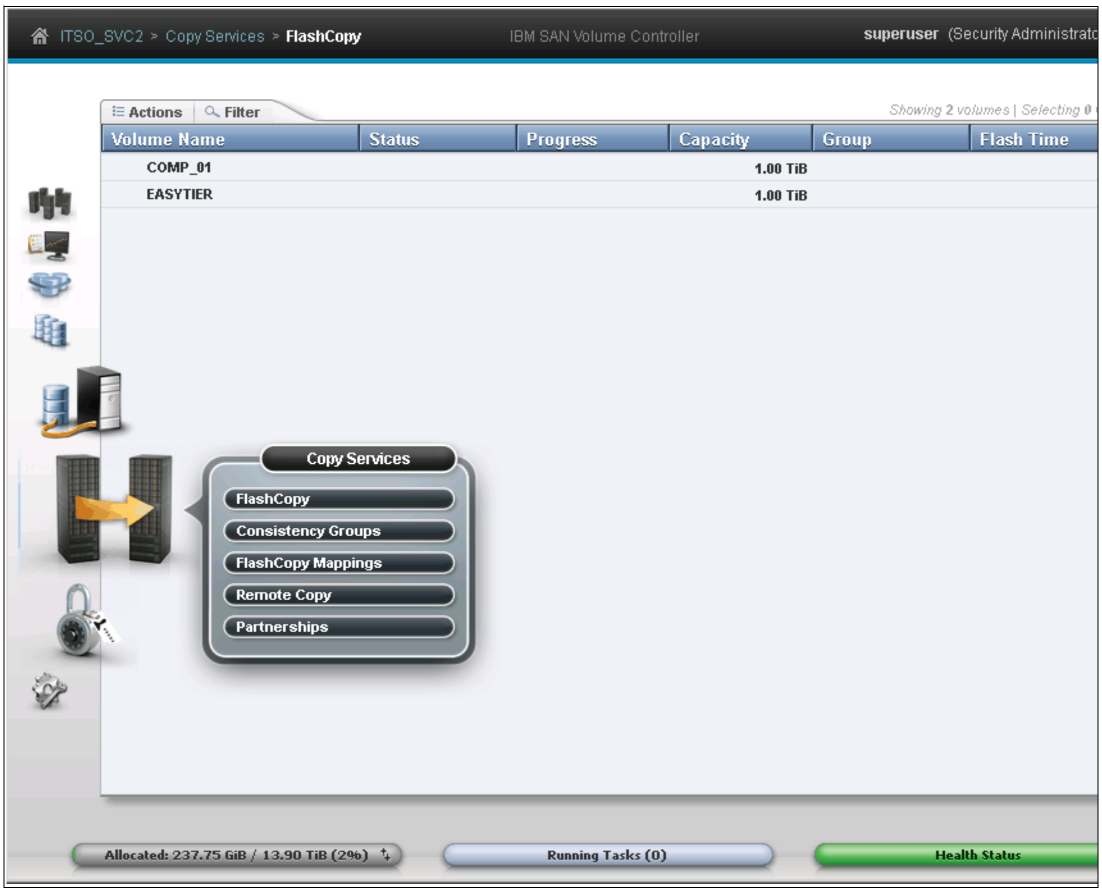


Figure 10-7 Copy Services

10.1.7 Access

In the Access menu, you can work with user administration, such as create and delete users, and user group administration.

Furthermore, you can see the auditlog, which shows all of the activities that have been executed on the system. Figure 10-8 shows the expanded view of the Access menu.

Figure 10-8 Access

10.1.8 Settings

In the Settings menu, you have access to the following settings:

- Event Notifications, such as call home (via email), SNMP, SMTP, and syslog.
- Directory Services, for enabling of remote authentication of users.
- Network, Both Fibre Channel settings, as well as IP settings.
- Support, where you can manage dumps, snaps, heatmap files, and so on.
- General, where you can upgrade the system, time and date settings, and so on.

Figure 10-9 shows an expanded view of the Settings menu.

Figure 10-9 Settings

10.2 New functions in 7.3

In this section, we show you the new functions in the GUI added in 7.3.

If you are configuring a Stretched System, you now have the option to select or change the system from a standard system to a Stretched System. You do have the option to select Stretched System during the initial configuration of the system and we covered this in Chapter 6, "SAN Volume Controller initial configuration" on page 113.

The option that we explain here is if you decide to change the system from a standard one-site system to a Stretched System after the initial configuration. With code level 7.2, you can only change that using the command-line interface (CLI).

So, log in to the system and go to **Monitoring** → **System details**. Click the system name, then click the Actions drop-down menu, as shown in Figure 10-10.

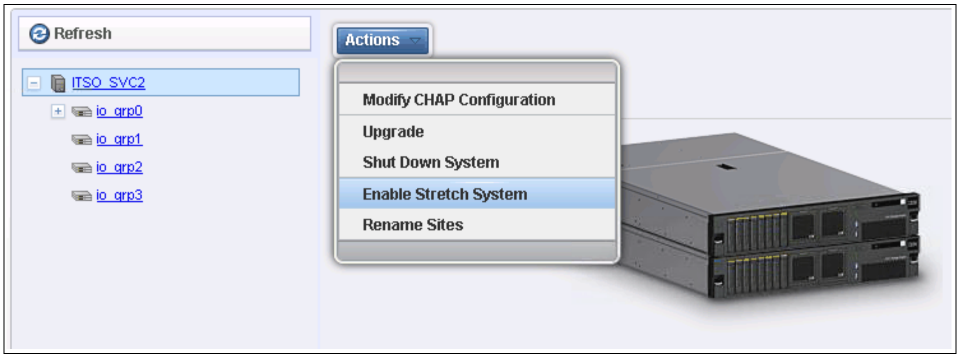


Figure 10-10 Enable Stretched System

When you select **Enable Stretch System**, you see a warning; see Figure 10-11.

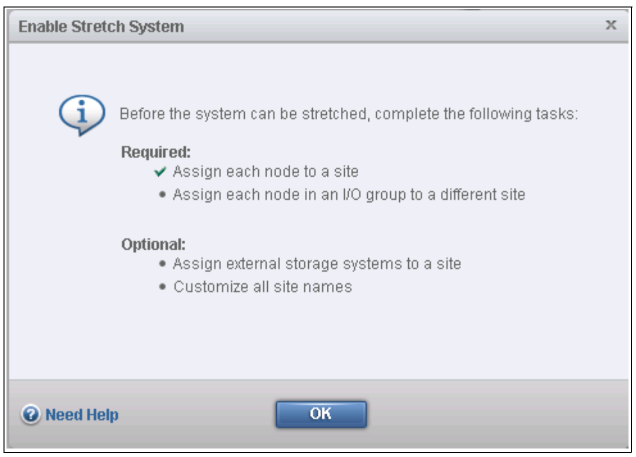


Figure 10-11 Enable Stretched System warning

You need to select each individual node, and assign a site to it, as seen in Figure 10-12.

Note: Ensure that you assign each node in an I/O group to a different site.

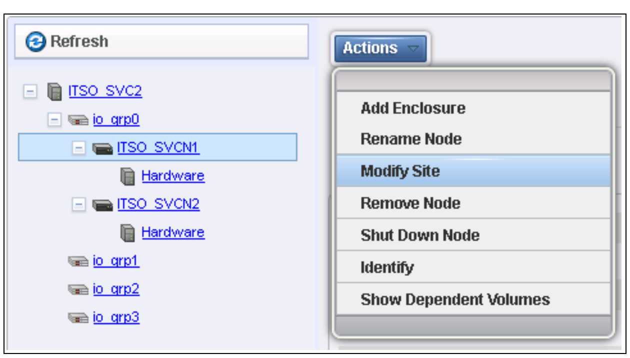


Figure 10-12 Modify Site

Figure 10-13 shows the option to select the site.

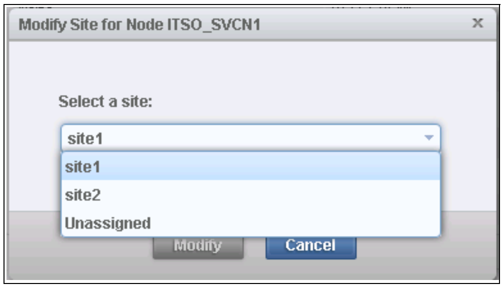


Figure 10-13 Site selection

You have to perform this task for all the nodes in the system. When completed, go back to the System Name, click the Actions drop-down menu, and select **Enable Stretch System**. When this is selected, you see a warning, as shown in Figure 10-14.

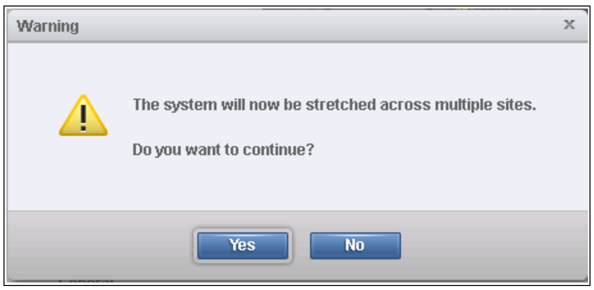


Figure 10-14 Warning

Click **Yes**, and the topology is changed from Standard to Stretched, as shown in Figure 10-15.

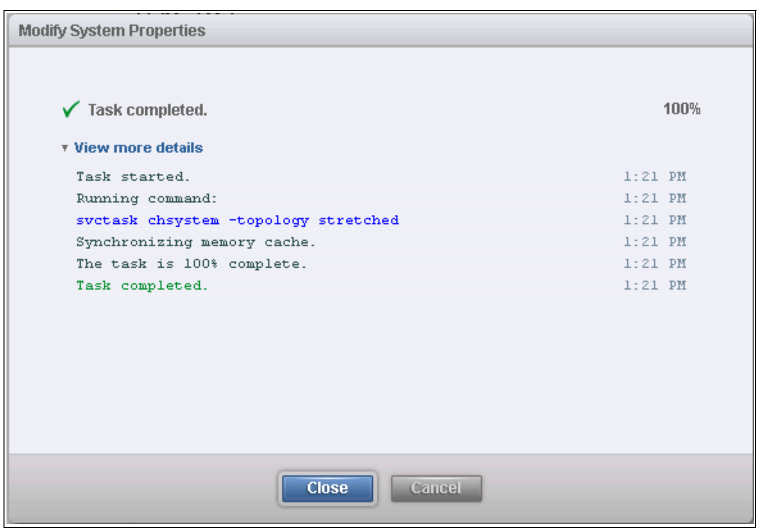


Figure 10-15 Change topology

Now you have successfully changed the nodes from standard topology to Stretched System topology.

You can rename the sites by going to **Monitoring** → **System details**. Click the system name, then click the Actions drop-down menu and select **Rename Sites**. A pop-up window is displayed, as shown in Figure 10-16.

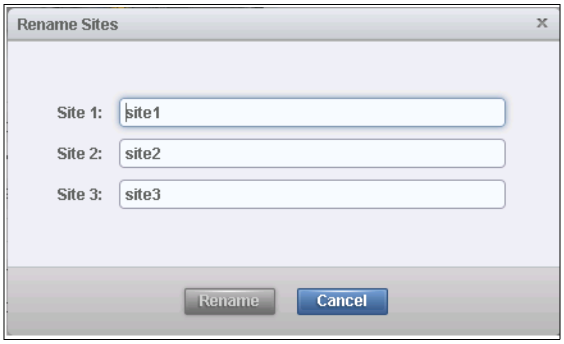


Figure 10-16 Change Site names

You can of course change the topology from Stretched System to Standard System, by reversing the above process intuitively.

10.2.1 Extent size selection during pool creation

Another new GUI function in 7.3 is the option to select and deselect the option to allow extent size selection during pool creation, and from an accessibility point, the option to use low graphic mode.

Go to menu **General** → **GUI preferences**, as shown in Figure 10-17 and tick/untick the wanted options.

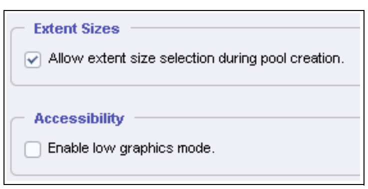


Figure 10-17 Extent size selection and low graphics mode selection

10.3 Storage Tier Advisor Tool

This section shows how to use the IBM Storage Tier Advisor Tool, also known as the STAT tool.

> **Note:** The STAT tool is not a part of the graphical user interface but can be downloaded from the IBM support pages:
>
> `http://www.ibm.com/support/docview.wss?uid=ssg1S4000935`

The Storage Tier Advisor Tool uses limited storage performance measurement data from a user's operational environment to model potential unbalanced workload (also known as *skew*) on disk and array resources. It is intended to supplement and support, but not replace, detailed pre-installation sizing and planning analysis. It is most useful to obtain a "rule of thumb" system-wide performance projection of cumulative latency reduction on arrays and disks when a solid-state drive configuration and the IBM Easy Tier function are used in combination to handle workload growth or skew management.

The "hot data" identification methodology in the tool is an engineering estimation based on expected cumulative latency reduction if the suggested solid-state device configuration is used with the measured workload and storage configuration. Care has been taken in the development of this tool, but the accuracy of any prediction of performance improvement is subject to various storage system configurations, conditions, and other variables beyond the scope of this tool. Accordingly, actual results may vary.

The **STAT.exe** command creates an HTML report of the IO distribution. IBM SAN Volume Controller input files are found under /dumps on the configuration node and are named "dpa_heat.node_name.time_stamp.data". The file must be offloaded manually using the CLI or GUI.

You can install the STAT tool on any Windows based PC or notebook, and you do not need to have direct access to the IBM SAN Volume Controller.

When the STAT tool is installed, it is time to offload/download heat files from your IBM SAN Volume Controller system.

The next few screen captures show how you can download the heat files from the GUI. The heat files can also be downloaded using the CLI/PuTTY Secure Copy (PSCP); however, we show how to offload/download these files using the GUI.

Log in to the GUI and select the **Settings menu** → **Support** and press the "**Show full log listing**" link, as shown in Figure 10-18.

Figure 10-18 Show full log listing

Now you can select the heat files that you want to use for the STAT tool. Select the files and right-click, and select **Download**; see Figure 10-19.

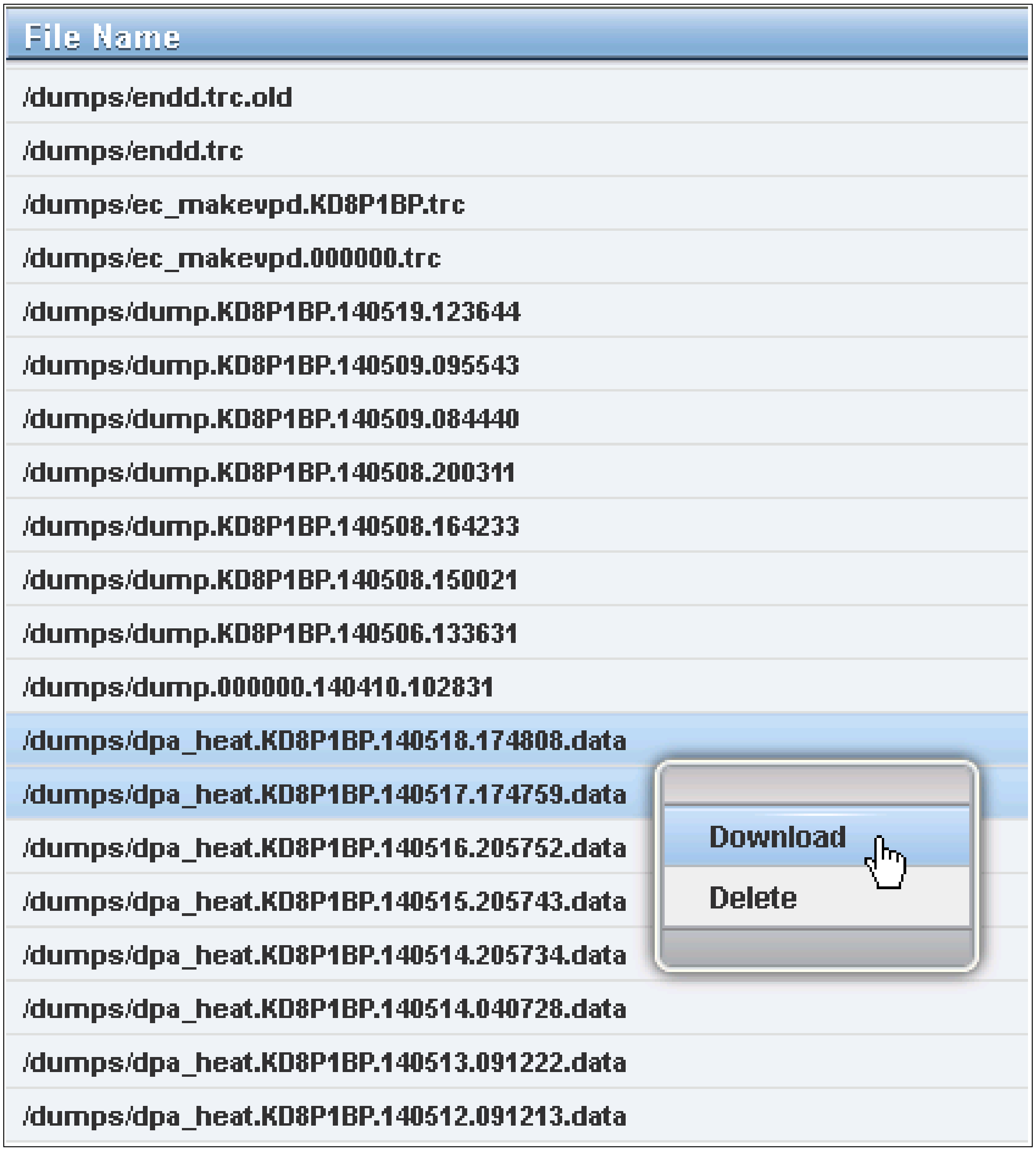


Figure 10-19 Download heat file

When the files have been offloaded/downloaded, open a command prompt and go to the directory where you installed the STAT tool (default path on a 64-bit Windows OS are `c:\Program Files (x86)\IBM\STAT`).

> **Note:** If the config node of the system reboots/asserts, and so on, note that the new config node will start the Easy Tier heatmap cycle count from 0, which means that it takes 24 hours until you see a new heatmap file in the `/dumps` directory.

You might want to copy and move the offloaded/downloaded files to the directory where you installed the STAT tool for ease of the usage. Otherwise, you must define the entire input file path every time you create a report.

To generate the report (in this case we already copied the input file to the STAT directory), execute the following command:

```
stat.exe -o "c:\Program Files (x86)\IBM\STAT\ITSO_SVC2"dpa_heat.KD8P1BP.140518.174808.data
```

Replace the heat file with the correct file names of the ones you have offloaded/downloaded. For IBM SAN Volume Controller systems, you can only execute one file concurrently.

In this scenario we used the syntax -o, which specifies an output path (a folder) which is very useful if you are generating STAT files from more than one system.

> **Note:** For detailed information about the usage of the STAT tool, refer to the `readme` file for the tool that is contained within the same directory where you installed it.

The returning message will show this:

CMUA00019I The STAT.exe command has completed.

Press any key to continue . . .

Go to the directory where the report was generated. In this case:

```
C:\Program Files (x86)\IBM\STAT\ITSO_SVC2
```

Open the `index.html` file. Your default browser will open the report as shown in Figure 10-20. Or you can simply open your browser and navigate to the `index.html` file.

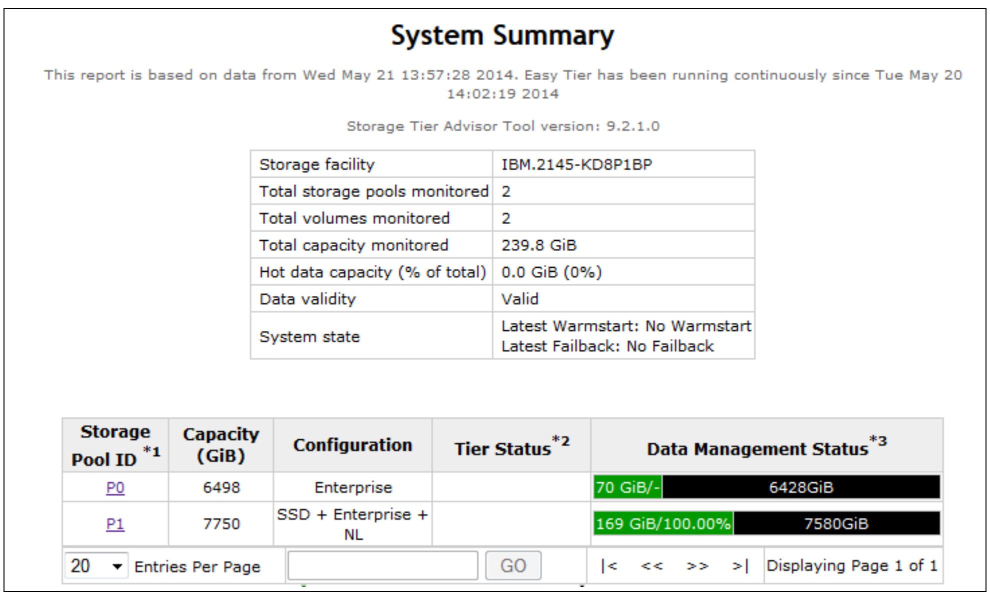
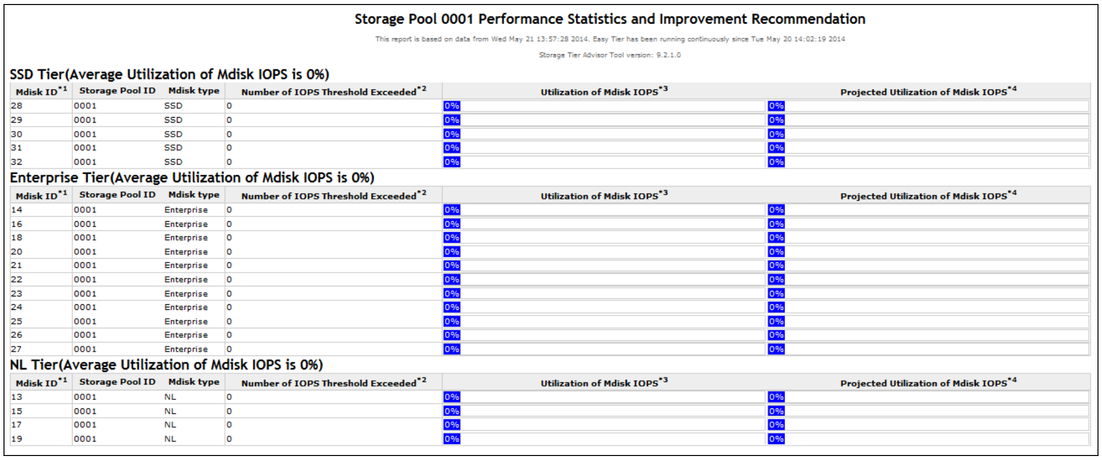


Figure 10-20 STAT System Summary

Note that not the entire report panels are shown in Figure 10-20 and this is an abbreviated example of the report.

You can click any of the Storage Pool IDs, which will open the Performance Statistics and Improvement Recommendation for the selected Pool ID, which we have done in Figure 10-21, with Pool ID named P1. The Pool ID **P1** is based on a 3 tiered pool, consisting of flash drives “SSD”, Enterprise drives, and Nearline drives.



Storage Pool 0001 Performance Statistics and Improvement Recommendation

This report is based on data from Wed May 21 13:57:28 2014. Easy Tier has been running continuously since Tue May 20 14:02:19 2014

Storage Tier Advisor Tool version: 9.2.1.0

SSD Tier(Average Utilization of Mdisk IOPS is 0%)

Mdisk ID*1	Storage Pool ID	Mdisk type	Number of IOPS Threshold Exceeded*2	Utilization of Mdisk IOPS*3	Projected Utilization of Mdisk IOPS*4
28	0001	SSD	0	0%	0%
29	0001	SSD	0	0%	0%
30	0001	SSD	0	0%	0%
31	0001	SSD	0	0%	0%
32	0001	SSD	0	0%	0%

Enterprise Tier(Average Utilization of Mdisk IOPS is 0%)

Mdisk ID*1	Storage Pool ID	Mdisk type	Number of IOPS Threshold Exceeded*2	Utilization of Mdisk IOPS*3	Projected Utilization of Mdisk IOPS*4
14	0001	Enterprise	0	0%	0%
16	0001	Enterprise	0	0%	0%
18	0001	Enterprise	0	0%	0%
20	0001	Enterprise	0	0%	0%
21	0001	Enterprise	0	0%	0%
22	0001	Enterprise	0	0%	0%
23	0001	Enterprise	0	0%	0%
24	0001	Enterprise	0	0%	0%
25	0001	Enterprise	0	0%	0%
26	0001	Enterprise	0	0%	0%
27	0001	Enterprise	0	0%	0%

NL Tier(Average Utilization of Mdisk IOPS is 0%)

Mdisk ID*1	Storage Pool ID	Mdisk type	Number of IOPS Threshold Exceeded*2	Utilization of Mdisk IOPS*3	Projected Utilization of Mdisk IOPS*4
13	0001	NL	0	0%	0%
15	0001	NL	0	0%	0%
17	0001	NL	0	0%	0%
19	0001	NL	0	0%	0%

Figure 10-21 Performance Statistics and Improvement Recommendation

Further details can be seen in the lower section of the Performance Statistics and Improvement Recommendation page, where you can expand hyperlinks for:

- Workload Distribution Across Tiers
- Recommended NL Configuration
- Volume Heat Distribution

More details for planning and configuration are available in the following IBM Redbooks publications:

- *Implementing IBM Easy Tier with IBM Real-time Compression*, TIPS1072
- *IBM System Storage SAN Volume Controller Best Practices and Performance Guidelines*, SG24-7521
- *IBM DS8000 Easy Tier*, REDP-4667 (this concept is similar to IBM SAN Volume Controller Easy Tier)

11

SAN Volume Controller Stretched System

In this chapter, we briefly describe the IBM System Storage SAN Volume Controller (SVC) Stretched System (formerly known as Split I/O Group, or Stretched Cluster). We also explain the term *Enhanced Stretched System*.

We do not provide deep technical details or implementation guidelines because those are fully covered in *IBM SAN and SVC Enhanced Stretched Cluster and VMware Solution Implementation*, SG24-8211.

For more information about Enhanced Stretched System prerequisites, visit the IBM Knowledge Center:

http://pic.dhe.ibm.com/infocenter/svc/ic/index.jsp

11.1 Stretched System overview

In a standard SVC configuration, all nodes are physically located within the same physical location, usually a single rack within a data center. Version 5.1 and later provide support for Stretched System configurations, where nodes within an I/O group can be physically separated from one another by up to 10 km (6.2 miles). This capability allows nodes to be placed in separate failure domains, or site, which provides protection against failures that affect a single failure domain or site. That initial support included the restriction that all communication between SVC node ports cannot traverse inter-switch links (ISLs). This limited the maximum supported distance between failure domains or sites. Starting with SVC 6.3, the ISL restriction was removed, which allowed the distance between failure domains to be extended to 300 km (186.4 miles). Additionally, in SVC 6.3, the maximum supported distance for non-ISL configurations was extended to 40 km (24.9 miles).

The SVC Stretched System configuration provides a continuous availability platform, whereby host access is maintained in the event of the loss of any single failure domain. This availability is accomplished through the inherent active/active architecture of SVC along with the use of volume mirroring. During a failure, the SVC nodes and associated mirror copy of the data remain online and available to service all host I/O.

The existing SVC Stretched System configuration has two locations, each with one node from each I/O group pair. The quorum disks are usually held in a third location. This solution is very successful in the field for many IBM clients.

It allows SVC hardware to be geographically distributed, resembling how IBM Metro Mirror is deployed. Stretched System deployments can span as short a distance as two racks in a data center, two buildings in a campus, across a city, or as far as 100 km (62.1 miles) or potentially more.

The key benefit of a Stretched System, compared to Metro Mirror, is that it allows for fast non-disruptive failover in the event of small scale outages. For example, if there is impact to just a single storage device, SVC fails over internally with minimal delay. If there is a failure in a fabric element, or SVC node, a host can fail over to another SVC node and continue performing I/O.

One of the attractions of SVC Stretched System, and a key advantage over some of the alternatives, is that the SVC failover uses the same multi-pathing driver as is used with conventional “unstretched” SVC deployments. This gives a wide interoperability matrix, matching the hosts and controllers that clients already deployed in their data centers. Many of the multi-pathing drivers are also the default/standard multi-pathing drivers for the OS in question.

Another aspect of this attractiveness is that SVC always has an automatic quorum to act as a tie-break. This means that no external management software or human intervention is ever required to perform a failover.

Enhanced Stretched System

The Enhanced Stretched System (ESS) was released in Version 7.2, which kept all of the previous SVC Stretched System functions and benefits and added high availability (HA), failure and recovery capabilities, and bandwidth savings.

Remember these key points about ESS:

- Use of ESS is optional. Existing Stretched System configurations are still supported. However, we encourage clients to use the new feature for its benefits.
- For the cluster topology that was introduced, a cluster is either *standard* or *stretched*. Configure it by using the **chsystem** command, or view it by using the **lssystem** command.
- Topology set to **stretched** enables site awareness features and disaster recovery (DR) capability.
- It is possible to convert an existing Stretched System to an ESS to change the topology non-disruptively any time after upgrade to Version 7.2.
- The ESS must be configured in advance to use the disaster recovery features.
- Use of internal flash drives (SSD) with SVC nodes is not supported in an ESS deployment.

Version 7.2 also introduced a *site awareness* concept.

- Site awareness can be used only when Topology is set to **stretched**.
- Topology set to **stretched** also means that the DR feature is enabled.
- Site object is now added, and the valid sites are 1, 2, and 3. You can set a name for each site if you prefer.
- The default names for the sites are site1, site2, and site3. Sites 1 and 2 are where the two halves of the ESS are located. Site 3 is the optional third site for a quorum tiebreaker disk.
- A Site field is added to nodes and controllers. You can set it only by using these SVC command-line interface commands, or via the graphical user interface (a new feature in the 7.3 release): **addnode**, **chnode**, and **chcontroller**. The nodes and controller must have sites set in advance, before you set Topology to **stretched**, and must have a site assigned.
- You can view site fields by using **lsnode**, **lscontroller**, **lsmdisk**, or **lsmdiskgrp** commands.
- Nodes can be assigned only to sites 1 or 2. Nodes cannot be assigned to site 3.
- You have the option to specify the site for each controller. The default for a controller is for its site to be undefined. This is the default for pre-existing controllers for upgrades to Version 7.2 or 7.3. Controllers can be assigned to sites 1, 2, or 3, or can be set to **undefined** again.
- An MDisk derives its site value from the controller that it is associated with it at that time. Some back-end storage devices are presented by SVC as multiple controller objects, and an MDisk might be associated with any of them from time to time. Make sure that all such controller objects have the same site specified to ensure that any MDisks associated with that controller are associated with a well-defined single site.
- The site for a controller can be changed when the DR feature is disabled. It can also be changed if the controller has no managed (or image mode) MDisks. The site for a controller cannot be changed when the DR feature is enabled if the controller uses managed (or image mode) MDisks.

The site property for a controller adjusts the I/O routing and error reporting for connectivity between nodes and the associated MDisks. These changes are effective for any MDisk controller that has a site defined, even if the DR feature is disabled.

SVC supports two Stretched System configurations:

- No ISL configuration:
  - Passive wave division multiplexing (WDM) devices can be used between both sites.
  - No ISLs can be located between the SVC nodes (similar to SVC 5.1-supported configurations).
  - The supported distance is up to 40 km (24.9 miles).

  Figure 11-1 shows an example of a Stretched System configuration with no ISL configuration.

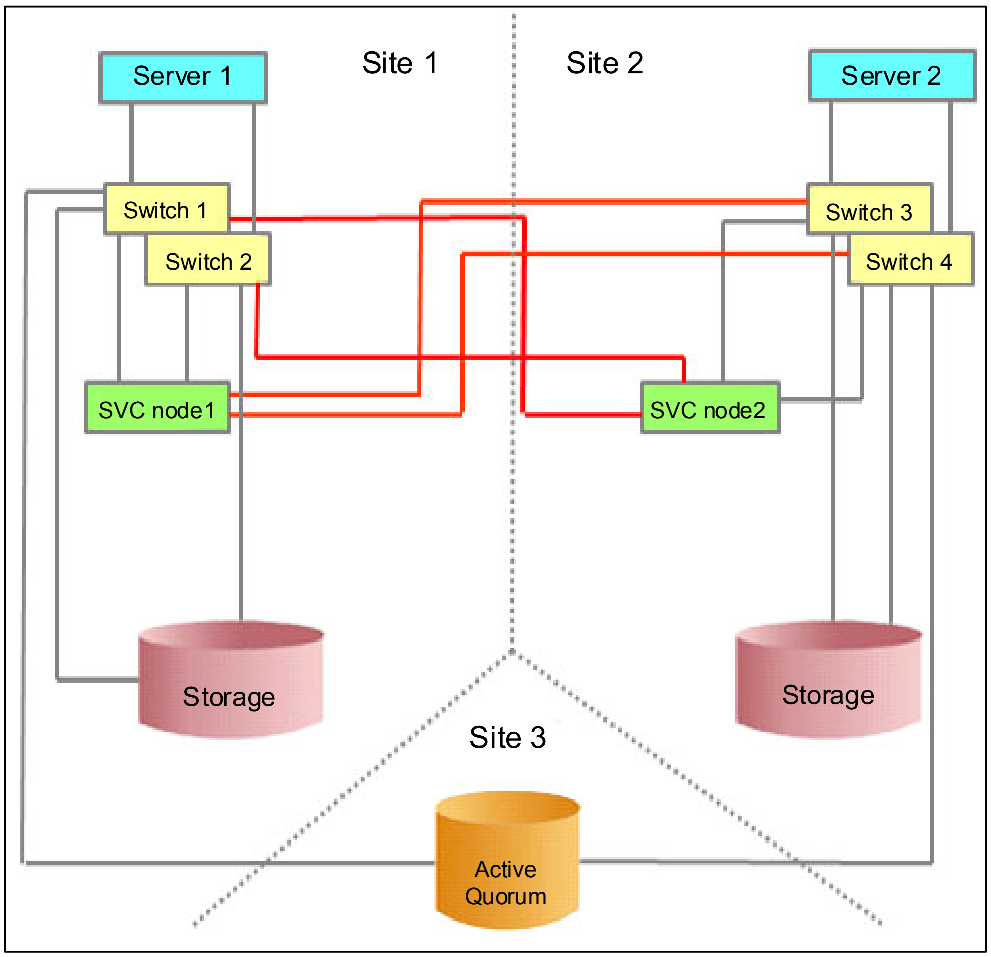


Figure 11-1 Stretched System with no ISL configuration

- ISL configuration:
  - ISLs located between the SVC nodes
  - Maximum distance similar to Metro Mirror distances
  - Physical requirements similar to Metro Mirror requirements
  - ISL distance extension with active and passive WDM devices

Figure 11-2 shows an example of a Stretched System with ISL configuration.

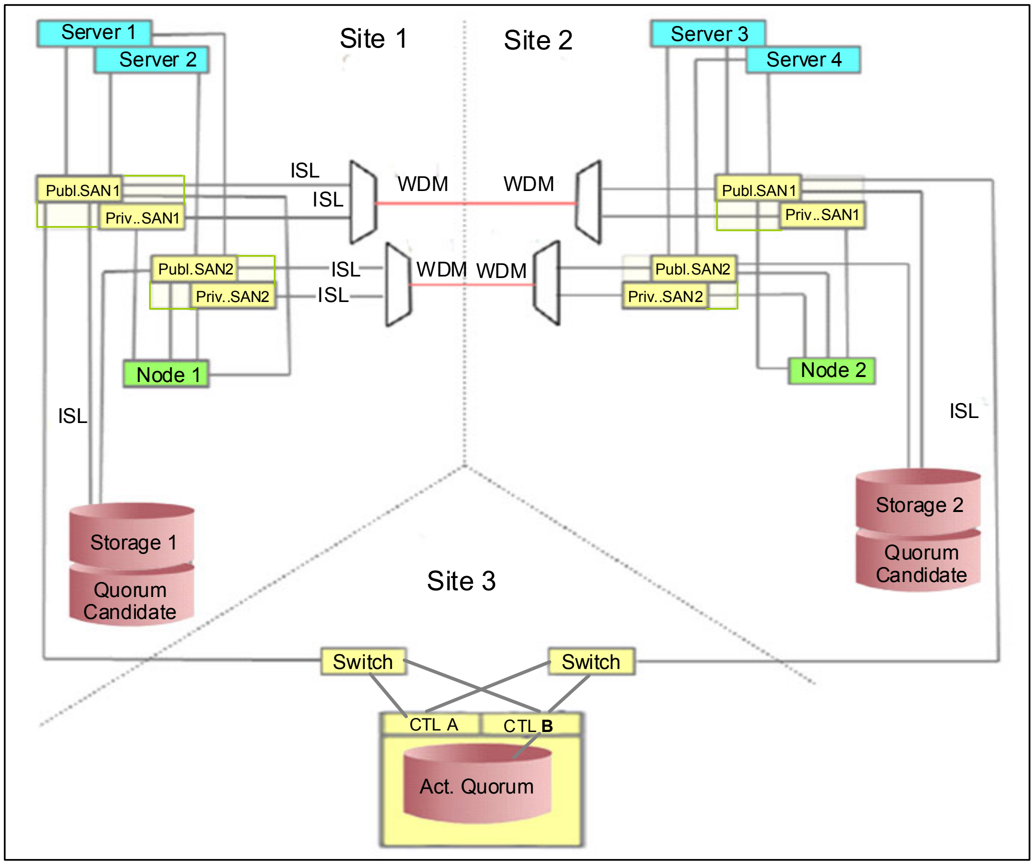


Figure 11-2 Stretched System with ISL Configuration

Use the Stretched System configuration with the volume mirroring option to realize an availability benefit.

After volume mirroring is configured, use the **lscontrollerdependentvdisks** command to validate that the volume mirrors reside on separate storage controllers. Having the volume mirrors reside on separate storage controllers ensures that access to the volumes is maintained in the event of the loss of a storage controller.

When implementing a Stretched System configuration, two of the three quorum disks can be co-located in the same room where the SVC nodes are located. However, the active quorum disk *must* reside in a separate room. This configuration ensures that a quorum disk is always available, even after a single-site failure.

For Stretched System configuration, configure the SVC in the following manner:

- Site 1: Half of the SVC clustered system nodes plus one quorum disk candidate
- Site 2: Half of the SVC clustered system nodes plus one quorum disk candidate
- Site 3: Active quorum disk

When a Stretched System configuration is used with volume mirroring, this configuration provides a high-availability solution that is tolerant of a failure at a single site. If either the primary or secondary site fails, the remaining sites can continue performing I/O operations.

11.1.1 More information

More information about SVC Stretched System and Enhanced Stretched System including planning, implementation, configuration steps, and troubleshooting is available in the following materials:

- *IBM SAN and SVC Enhanced Stretched Cluster and VMware Solution Implementation*, SG24-8211
- *IBM SAN Volume Controller Stretched Cluster with PowerVM and PowerHA*, SG24-8142
- *IBM System Storage SAN Volume Controller Best Practices and Performance Guidelines*, SG24-7521
- http://pic.dhe.ibm.com/infocenter/svc/ic/index.jsp

Related publications

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

IBM Redbooks

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

- *Implementing the IBM System Storage SAN Volume Controller V7.2*, SG24-7933
- *Implementing the IBM Storwize V7000 V7.2*, SG24-7938
- *IBM b-type Gen 5 16 Gbps Switches and Network Advisor*, SG24-8186
- *Introduction to Storage Area Networks and System Networking*, SG24-5470
- *IBM SAN Volume Controller and IBM FlashSystem 820: Best Practices and Performance Capabilities*, REDP-5027
- *Implementing the IBM SAN Volume Controller and FlashSystem 820*, SG24-8172
- *Implementing IBM FlashSystem 840*, SG24-8189
- *IBM FlashSystem in IBM PureFlex System Environments*, TIPS1042
- *IBM FlashSystem 840 Product Guide*, TIPS1079
- *IBM FlashSystem 820 Running in an IBM StorwizeV7000 Environment*, TIPS1101
- *Implementing FlashSystem 840 with SAN Volume Controller*, TIPS1137
- *IBM FlashSystem V840 Enterprise Performance Solution*, TIPS1158
- *IBM Midrange System Storage Implementation and Best Practices Guide*, SG24-6363
- *IBM System Storage b-type Multiprotocol Routing: An Introduction and Implementation*, SG24-7544
- *IBM Tivoli Storage Area Network Manager: A Practical Introduction*, SG24-6848
- *Tivoli Storage Productivity Center for Replication for Open Systems*, SG24-8149
- *Tivoli Storage Productivity Center V5.2 Release Guide*, SG24-8204
- *Implementing an IBM b-type SAN with 8 Gbps Directors and Switches*, SG24-6116

You can search for, view, download, or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Other resources

These publications are also relevant as further information sources:

- *IBM System Storage Master Console: Installation and User's Guide*, GC30-4090

- *IBM System Storage Open Software Family SAN Volume Controller: CIM Agent Developers Reference*, SC26-7545
- *IBM System Storage Open Software Family SAN Volume Controller: Command-Line Interface User's Guide*, SC26-7544
- *IBM System Storage Open Software Family SAN Volume Controller: Configuration Guide*, SC26-7543
- *IBM System Storage Open Software Family SAN Volume Controller: Host Attachment Guide*, SC26-7563
- *IBM System Storage Open Software Family SAN Volume Controller: Installation Guide*, SC26-7541
- *IBM System Storage Open Software Family SAN Volume Controller: Planning Guide*, GA22-1052
- *IBM System Storage Open Software Family SAN Volume Controller: Service Guide*, SC26-7542
- *IBM System Storage SAN Volume Controller - Software Installation and Configuration Guide,* SC23-6628
- *IBM System Storage SAN Volume Controller V6.2.0 - Software Installation and Configuration Guide,* GC27-2286

  http://www-01.ibm.com/support/knowledgecenter/STPVGU_7.3.0/com.ibm.storage.svc.console.730.doc/svc_ichome_730.html
- *IBM System Storage SAN Volume Controller 6.2.0 Configuration Limits and Restrictions,* S1003799
- *IBM TotalStorage Multipath Subsystem Device Driver User's Guide*, SC30-4096
- *IBM XIV and SVC Best Practices Implementation Guide*

  https://ibm.biz/BdFNLY
- *Considerations and Comparisons between IBM SDD for Linux and DM-MPIO*

  http://www.ibm.com/support/docview.wss?rs=540&context=ST52G7&q1=linux&uid=ssg1S7001664&loc=en_US&cs=utf-8&lang=en

Referenced websites

These websites are also relevant as further information sources:

- IBM Storage home page

  http://www.storage.ibm.com
- IBM site to download SSH for AIX

  http://www.ibm.com/developerworks/systems/articles/openssh_updated.html
- IBM TotalStorage Virtualization home page

  http://www-1.ibm.com/servers/storage/software/virtualization/index.html
- SAN Volume Controller supported platform

  http://www-1.ibm.com/servers/storage/support/software/sanvc/index.html
- SAN Volume Controller IBM Knowledge Center

  http://pic.dhe.ibm.com/infocenter/svc/ic/index.jsp

- Cygwin Linux-like environment for Windows

  http://www.cygwin.com

- Open source site for SSH for Windows and Mac

  http://www.openssh.com/windows.html

- Sysinternals home page

  http://www.sysinternals.com

- Download site for Windows SSH freeware

  http://www.chiark.greenend.org.uk/~sgtatham/putty

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

IBM

Redbooks

IBM SAN Volume Controller 2145-DH8 Introduction and Implementation

(0.5" spine)
0.475"<->0.873"
250 <-> 459 pages

IBM SAN Volume Controller 2145-DH8 Introduction and Implementation

Learn about the latest addition to the SVC and Storwize family

Understand the new functions and features

Enjoy a painless implementation

Data is the new currency of business, the most critical asset of the modern organization. In fact, enterprises that can gain business insights from their data are twice as likely to outperform their competitors; yet, 72 percent of them have not started or are only planning big data activities. In addition, organizations often spend too much money and time managing where their data is stored. The average firm purchases 24% more storage every year, but uses less than half of the capacity it already has.

A member of the IBM Storwize family, IBM SAN Volume Controller (SVC) Data Platform is a storage virtualization system that enables a single point of control for storage resources to help support improved business application availability and greater resource utilization. The objective is to manage storage resources in your IT infrastructure and to make sure they are used to the advantage of your business, and do it quickly, efficiently, and in real time, while avoiding increases in administrative costs.

Virtualizing storage with SVC Data Platform helps make new and existing storage more effective. SVC Data Platform includes many functions traditionally deployed separately in disk systems. By including these in a virtualization system, SVC Data Platform standardizes functions across virtualized storage for greater flexibility and potentially lower costs.

SVC Data Platform functions benefit all virtualized storage. For example, IBM Easy Tier optimizes use of flash storage. And IBM Real-time Compression enhances efficiency even further by enabling the storage of up to five times as much active primary data in the same physical disk space. Finally, high-performance thin provisioning helps automate provisioning. These benefits can help extend the useful life of existing storage assets, reducing costs.

Integrating these functions into SVC Data Platform also means that they are designed to operate smoothly together, reducing management effort.

In this IBM Redbooks publication, we discuss the latest features and functions of the SVC 2145-DH8 and software version 7.3, implementation, architectural improvements, and Easy Tier.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-8229-00 ISBN 0738440078