IBM System Storage SAN Volume Controller

Planning Guide

Version 4.1.0

GA32-0551-00

IBM System Storage SAN Volume Controller

Planning Guide

Version 4.1.0

GA32-0551-00

Seventh Edition (June 2006)

Before using this information and the product it supports, read the information in ″Notices.″

Contents

Figures

Tables

About this guide

This publication introduces the IBM System Storage SAN Volume Controller, its components, and its features.

It also provides planning guidelines for installing and configuring the SAN Volume Controller.

Who should use this guide?

This publication is intended for anyone who is planning to install and configure an IBM System Storage SAN Volume Controller.

Summary of changes

This document contains terminology, maintenance, and editorial changes.

Technical changes or additions to the text and illustrations are indicated by a vertical line to the left of the change. This summary of changes describes new functions that have been added to this release.

Summary of changes for GA32-0551-00 SAN Volume Controller Planning Guide

The Summary of changes provides a list of new, modified, and changed information since the last version of the guide.

New information

This topic describes the changes to this guide since the previous edition, GA22-1052-05. The following sections summarize the changes that have since been implemented from the previous version.

This version includes the following new information:

- There is a new SAN Volume Controller node model, the SAN Volume Controller 2145-8F4.

The following new topics have been added:

- Metro & Global Mirror
- Global Mirror
- Standard and persistent reserves

Changed information

This section lists the updates that were made in this document.

The following topics have been updated:

- SAN Volume Controller overview
- SAN Volume Controller operating environment
- UPS
- UPS configuration

- UPS operation
- Preparing your SAN Volume Controller environment
- Hardware location guidelines
- Cable connection table
- Switch zoning for the SAN Volume Controller
- Cluster configuration using SAN fabrics with long-distance fibre links
- Nodes
- UPS and power domains
- HBAs
- Power requirements
- Fibre-channel switches and interswitch links

Summary of changes for GA22-1052-05 SAN Volume Controller Planning Guide

The Summary of Changes provides a list of new, modified, and changed information since the last version of the guide.

Changed information

This topic describes the changes to this guide since the previous edition, GA22-1052-04. The following sections summarize the changes that have since been implemented from the previous version.

- The previous release referred to the uninterruptible power supply (UPS) as UPS 5115 and UPS 5125, by model number. For this release, the UPS is referred to by machine type. For example, this publication states 2145 uninterruptible power supply-1U (2145 UPS-1U) and uninterruptible power supply (2145 UPS). 2145 UPS-1U refers to UPS 5115 and 2145 UPS refers to UPS 5125.

  Note: If text is referring to the UPS or to the uninterruptible power supply, then it is referring to a generic UPS and can be referring to either UPS. When the UPS is referred to as the 2145 UPS-1U or the 2145 UPS, then the specific UPS is designated.
- There is a new SAN Volume Controller supported model. The SAN Volume Controller is now documented by model number. For example, this publication states two SAN Volume Controller models types: SAN Volume Controller 2145-4F2 and the new SAN Volume Controller 2145-8F2.

  Note: If text is referring to the SAN Volume Controller then it is referring to a generic SAN Volume Controller and can be referring to either SAN Volume Controller model. When the SAN Volume Controller is referred to as the SAN Volume Controller 2145-4F2 or the SAN Volume Controller 2145-8F2, then the specific SAN Volume Controller is designated.
- The IBM® TotalStorage® FAStT series is now called the IBM TotalStorage DS4000 series.
- Cache modes
- VDisk to host mappings
- Hardware location guidelines
- Cable connection table
- Switch zoning for the SAN Volume Controller
- Zoning considerations for Metro Mirror

- UPS and power domains
- Storage subsystems
- Mixing manufacturer switches in a single SAN fabric
- UPS
- UPS configuration
- UPS operation
- Maximum configuration

Deleted information

This section lists the updates that were made in this document.

- The SAN Volume Controller Console no longer arrives with a CD set. All publication and product upgrades are available from the following Web site:

  http://www-1.ibm.com/servers/storage/support/virtual/2145.html

Emphasis

Different typefaces are used in this guide to show emphasis.

The following typefaces are used to show emphasis:

Boldface	Text in **boldface** represents menu items and command names.
Italics	Text in *italics* is used to emphasize a word. In command syntax, it is used for variables for which you supply actual values, such as a default directory or the name of a cluster.
`Monospace`	Text in `monospace` identifies the data or commands that you type, samples of command output, examples of program code or messages from the system, or names of command flags, parameters, arguments, and name-value pairs.

SAN Volume Controller library and related publications

A list of other publications that are related to this product are provided to you for your reference.

The tables in this section list and describe the following publications:

- The publications that make up the library for the IBM System Storage SAN Volume Controller
- Other IBM publications that relate to the SAN Volume Controller

SAN Volume Controller library

The following table lists and describes the publications that make up the SAN Volume Controller library. Unless otherwise noted, these publications are available in Adobe portable document format (PDF) from the following Web site:

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

Title	Description	Order number
IBM System Storage SAN Volume Controller: CIM agent Developer's Reference	This reference guide describes the objects and classes in a Common Information Model (CIM) environment.	GA32-0552
IBM System Storage SAN Volume Controller: Command-Line Interface User's Guide	This guide describes the commands that you can use from the SAN Volume Controller command-line interface (CLI).	SC26-7903
IBM System Storage SAN Volume Controller: Configuration Guide	This guide provides guidelines for configuring your SAN Volume Controller.	SC26-7902
IBM System Storage SAN Volume Controller: Host Attachment Guide	This guide provides guidelines for attaching the SAN Volume Controller to your host system.	SC26-7905
IBM System Storage SAN Volume Controller: Installation Guide	This guide includes the instructions the service representative uses to install the SAN Volume Controller.	GC26-7900
IBM System Storage SAN Volume Controller: Planning Guide	This guide introduces the SAN Volume Controller and lists the features you can order. It also provides guidelines for planning the installation and configuration of the SAN Volume Controller.	GA32-0551
IBM System Storage SAN Volume Controller: Service Guide	This guide includes the instructions the service representative uses to service the SAN Volume Controller.	GC26-7901
IBM System Safety Notices	This guide contains the danger and caution notices for the SAN Volume Controller. The notices are shown in English and in numerous other languages.	G229-9054
IBM System Storage Master Console for SAN File System and SAN Volume Controller: Installation and User's Guide	This guide includes the instructions on how to install and use the SAN Volume Controller Console	GC30-4090

Other IBM publications

The following table lists and describes other IBM publications that contain additional information related to the SAN Volume Controller.

Title	Description	Order number
IBM System Storage Multipath Subsystem Device Driver: User's Guide	This guide describes the IBM System Storage Multipath Subsystem Device Driver Version 1.5 for TotalStorage Products and how to use it with the SAN Volume Controller. This publication is referred to as the *IBM System Storage Multipath Subsystem Device Driver: User's Guide*.	SC30-4131

Related Web sites

The following Web sites provide information about the SAN Volume Controller or related products or technologies.

Type of information	Web site
SAN Volume Controller support	http://www.ibm.com/storage/support/2145
Technical support for IBM storage products	http://www.ibm.com/storage/support/

How to order IBM publications

The publications center is a worldwide central repository for IBM product publications and marketing material.

The IBM publications center

The IBM publications center offers customized search functions to help you find the publications that you need. Some publications are available for you to view or download free of charge. You can also order publications. The publications center displays prices in your local currency. You can access the IBM publications center through the following Web site:

http://www.ibm.com/shop/publications/order/

Publications notification system

The IBM publications center Web site offers you a notification system for IBM publications. Register and you can create your own profile of publications that interest you. The publications notification system sends you a daily e-mail that contains information about new or revised publications that are based on your profile.

If you want to subscribe, you can access the publications notification system from the IBM publications center at the following Web site:

http://www.ibm.com/shop/publications/order/

How to send your comments

Your feedback is important to help us provide the highest quality information. If you have any comments about this book or any other documentation, you can submit them in one of the following ways:

- e-mail

  Submit your comments electronically to the following e-mail address:

  starpubs@us.ibm.com

  Be sure to include the name and order number of the book and, if applicable, the specific location of the text you are commenting on, such as a page number or table number.

- Mail

  Fill out the Readers' Comments form (RCF) at the back of this book. If the RCF has been removed, you can address your comments to:

  International Business Machines Corporation
  RCF Processing Department
  Department 61C
  9032 South Rita Road
  Tucson, Arizona 85775-4401
  U.S.A.

Chapter 1. Virtualization

Virtualization is a concept that applies to many areas of the information technology industry.

For data storage, virtualization includes the creation of a pool of storage that contains several disk subsystems. These subsystems can be supplied from various vendors. The pool can be split into virtual disks (VDisks) that are visible to the host systems that use them. Therefore, VDisks can use mixed back-end storage and provide a common way to manage a storage area network (SAN).

Historically, the term *virtual storage* has described the virtual memory techniques that have been used in operating systems. The term *storage virtualization*, however, describes the shift from managing physical volumes of data to logical volumes of data. This shift can be made on several levels of the components of storage networks. Virtualization separates the representation of storage between the operating system and its users from the actual physical storage components. This technique has been used in mainframe computers for many years through methods such as system-managed storage and products like the IBM Data Facility Storage Management Subsystem (DFSMS). Virtualization can be applied at the following four main levels:

At the server level
: Manages volumes on the operating systems servers. An increase in the amount of logical storage over physical storage is suitable for environments that do not have storage networks.

At the storage device level
: Uses striping, mirroring and RAIDs to create disk subsystems. This type of virtualization can range from simple RAID controllers to advanced volume management such as that provided by the IBM TotalStorage Enterprise Storage Server® (ESS) or by Log Structured Arrays (LSA). The Virtual Tape Server (VTS) is another example of virtualization at the device level.

At the fabric level
: Enables storage pools to be independent of the servers and the physical components that make up the storage pools. One management interface can be used to manage different storage systems without affecting the servers. The SAN Volume Controller performs virtualization at the fabric level.

At the file system level
: Provides the highest benefit because data is shared, allocated, and protected at the data level rather than the volume level.

Virtualization is a radical departure from traditional storage management. In traditional storage management, storage is attached directly to a host system, which controls storage management. SANs introduced the principle of networks of storage, but storage is still primarily created and maintained at the RAID subsystem level. Multiple RAID controllers of different types require knowledge of, and software that is specific to, the given hardware. Virtualization provides a central point of control for disk creation and maintenance.

One problem area that virtualization addresses is unused capacity. Before virtualization, individual host systems each had their own storage, which wasted unused storage capacity. Using virtualization, storage is pooled so that jobs from any attached system that need large amounts of storage capacity can use it as needed. Virtualization makes it easier to regulate the amount of available storage without having to use host system resources or to turn storage devices off and on to add or remove capacity. Virtualization also provides the capability to move storage between storage subsystems transparently to host systems.

Types of virtualization

Virtualization can be performed either asymmetrically or symmetrically. Figure 1 provides a diagram of the levels of virtualization.

Asymmetric
A virtualization engine is outside the data path and performs a metadata style service.

Symmetric
A virtualization engine sits in the data path and presents disks to the hosts, but hides the physical storage from the hosts. Advanced functions, such as cache and Copy Services, can therefore be implemented in the engine itself.

Virtualization at any level provides benefits. When several levels are combined, the benefits of those levels can also be combined. For example, you can gain the most benefits if you attach a low-cost RAID controller to a virtualization engine that provides virtual volumes for use by a virtual file system.

Note: The SAN Volume Controller implements fabric-level *virtualization*. Within the context of the SAN Volume Controller and throughout this document, *virtualization* refers to symmetric fabric-level virtualization.

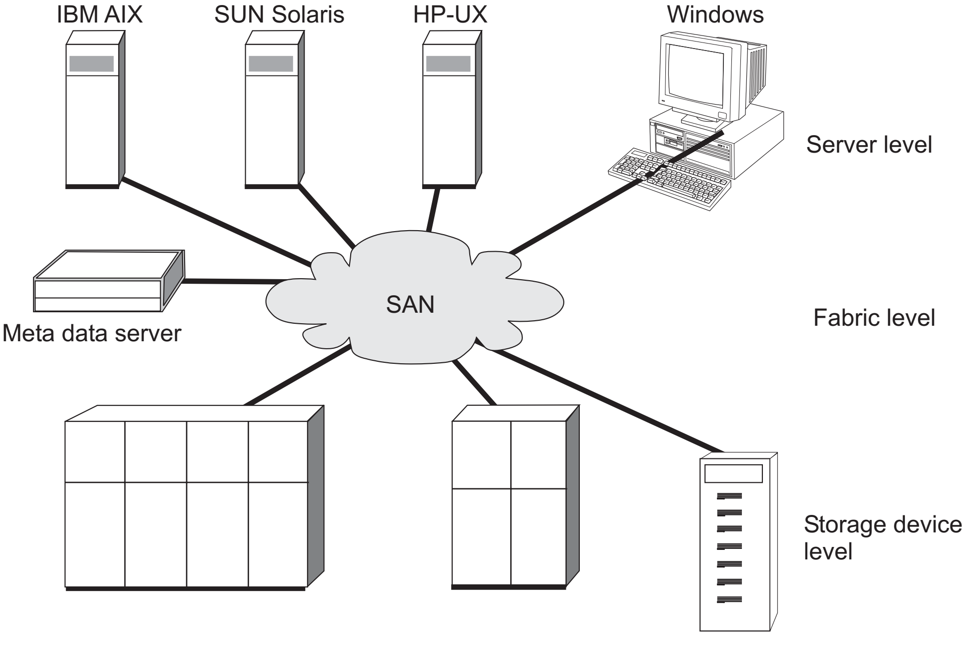


Figure 1. Levels of virtualization

Related concepts

"VDisks" on page 23
A *virtual disk (VDisk)* is a logical disk that the cluster presents to the storage area network (SAN).

The need for virtualization

Storage is a facility that computer users want to access at any time, from any location, with a minimum amount of management.

Users expect the storage devices to provide enough capacity and to be reliable. The amount of storage that users require, however, is increasing quickly. Internet users use large amounts of storage daily. Many users are mobile, access patterns cannot be predicted, and the content of the data becomes more and more interactive. Because the amount of data that is processed is large, it can no longer be managed manually. Automatic management is required, as are new levels of bandwidth and load balancing. Also, it is important that all this data can be shared between different types of operating systems, because the communication networks cannot process the large replication, download, and copying operations that would otherwise be required.

Storage area networks (SANs) are high-speed switched networks that let multiple computers share access to many storage devices. SANs allow for the use of advanced software that automatically manages the storage of data. With such advanced software, the computers that are connected to a particular network can, therefore, access storage wherever that storage is available in the network. The user is no longer aware of, and no longer needs to know, which physical devices contain which data. The storage has become virtualized. In a similar way to how virtual memory has solved the problems of the management of a limited resource in application programs, the virtualization of storage has provided a more intuitive use of storage, while software quietly manages the storage network in the background.

Fabric-level virtualization models

In a SAN without virtualization, storage devices are connected directly to host systems and are maintained locally by those host systems.

Although storage area networks (SANs) have introduced the principle of networks, storage devices are still mainly assigned to individual host systems and storage is still mainly created and maintained at the RAID subsystem level. Therefore, RAID controllers of different types require access to both the hardware and the software that is used.

Virtualization provides a complete change from the traditional storage management. It provides a central point of control for disk creation and management, and therefore requires changes to the way in which storage management is done.

Fabric level virtualization is the principle in which a pool of storage is created from more than one disk subsystem. This pool is then used to set up virtual disks (VDisks) that are made visible to the host systems. These VDisks use whatever storage is available and permit a common way to manage SAN storage.

Fabric level virtualization can be done in either of two ways: asymmetric or symmetric.

With asymmetric virtualization, the virtualization engine is outside the data path. It provides a metadata server that contains all the mapping and the locking tables. The storage devices contain only data.

Because the flow of control is separated from the flow of data, input/output (I/O) operations can use the full bandwidth of the SAN. A separate network or SAN link is used for control.

However, there are disadvantages to asymmetric virtualization:

- Data is at risk to increased security exposures, and the control network must be protected with a firewall.
- Metadata can become very complicated when files are distributed across several devices.
- Each host that accesses the SAN must know how to access and interpret the metadata. Specific device driver or agent software must therefore be running on each of these hosts.
- The metadata server cannot run advanced functions, such as caching or Copy Services, because it only has access to the metadata, not the data itself.

Symmetric virtualization

The SAN Volume Controller provides symmetric virtualization.

Virtualization splits the storage that is presented by the storage subsystems into smaller chunks that are known as extents. These extents are then concatenated, using various policies, to make virtual disks (VDisks). With symmetric virtualization, host systems can be isolated from the physical storage. Advanced functions, such as data migration, can run without the need to reconfigure the host. With symmetric virtualization, the virtualization engine is the central configuration point for the SAN.

Figure 2 shows that the storage is pooled under the control of the virtualization engine, because the separation of the control from the data occurs in the data path. The virtualization engine performs the logical-to-physical mapping.

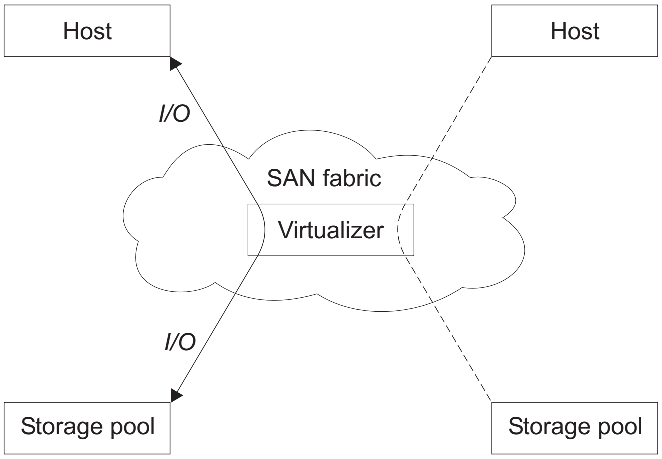


Figure 2. Symmetrical virtualization

The virtualization engine directly controls access to the storage and to the data that is written to the storage. As a result, locking functions that provide data integrity and advanced functions, such as cache and Copy Services, can be run in the virtualization engine itself. Therefore, the virtualization engine is a central point of control for device and advanced function management. Symmetric virtualization allows you to build a firewall in the storage network. Only the virtualization engine can grant access through the firewall.

Symmetric virtualization can cause some problems. The main problem that is associated with symmetric virtualization is scalability. Scalability can cause poor performance because all input/output (I/O) must flow through the virtualization engine. To solve this problem, you can use an *n-way* cluster of virtualization engines that has failover capacity. You can scale the additional processor power, cache memory, and adapter bandwidth to achieve the level of performance that you want. Additional memory and processing power are needed to run advanced services such as Copy Services and caching.

The SAN Volume Controller uses symmetric virtualization. Single virtualization engines, which are known as *nodes*, are combined to create *clusters*. Each cluster can contain between two and eight nodes.

SAN Volume Controller overview

The *SAN Volume Controller* is a SAN (storage area network) appliance that attaches open-systems storage devices to supported open-systems hosts.

The SAN Volume Controller is a rack-mounted unit that you can install in a standard Electrical Industries Association (EIA) 19-inch rack. It provides symmetric virtualization by creating a pool of managed disks (MDisks) from the attached storage subsystems. Those storage systems are then mapped to a set of virtual disks (VDisks) for use by attached host systems. System administrators can view and access a common pool of storage on the SAN. This lets the administrators use storage resources more efficiently and provides a common base for advanced functions.

A *SAN* is a high-speed fibre-channel network that connects host systems and storage devices. It allows a host system to be connected to a storage device across the network. The connections are made through units such as routers, gateways, hubs, and switches. The area of the network that contains these units is known as the *fabric* of the network.

The SAN Volume Controller is analogous to a logical volume manager (LVM) on a SAN. The SAN Volume Controller performs the following functions for the SAN storage that it controls:

- Creates a single pool of storage
- Provides logical unit virtualization
- Manages logical volumes
- Provides the following advanced functions for the SAN:
  - Large scalable cache
  - Copy Services
    - FlashCopy® (point-in-time copy)
    - Metro Mirror (synchronous copy)
    - Global Mirror (asynchronous copy)
    - Data migration
  - Space management
    - Mapping that is based on desired performance characteristics
    - Metering of service quality

Each SAN Volume Controller is a *node*. The nodes are always installed in pairs, with one-to-four pairs of nodes constituting a *cluster*. Each node in a pair is configured to back up the other. Each pair of nodes is known as an *I/O group*.

There are three models of SAN Volume Controller nodes: the SAN Volume Controller 2145-4F2, the SAN Volume Controller 2145-8F2 and the SAN Volume Controller 2145-8F4. Figure 3 and Figure 4 provide illustrations of the three types of SAN Volume Controller nodes.

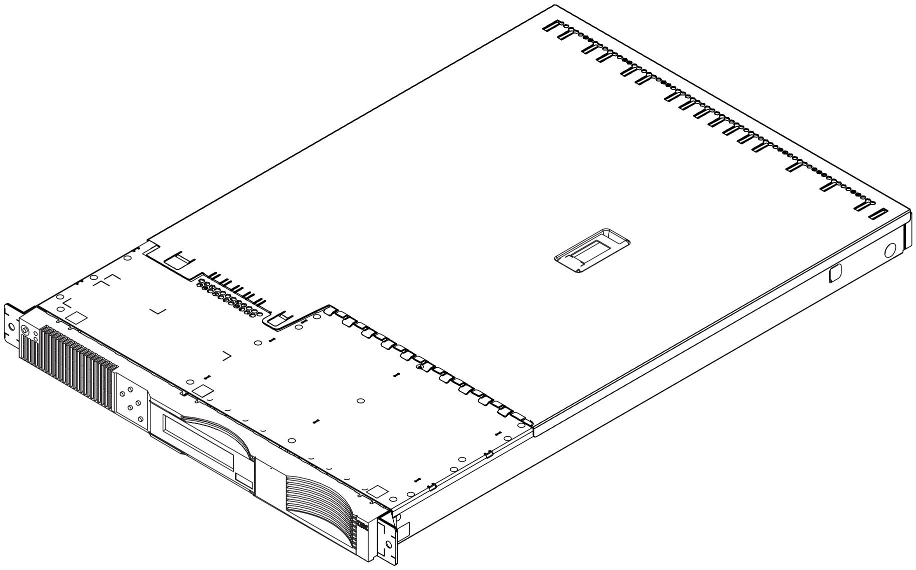


Figure 3. SAN Volume Controller 2145-4F2 node

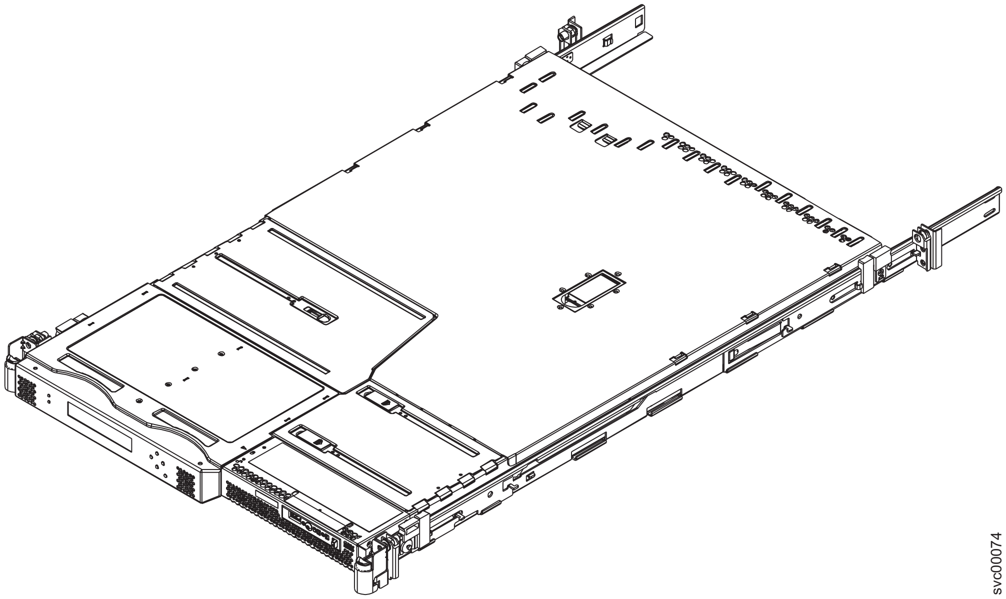


Figure 4. SAN Volume Controller 2145-8F2 and SAN Volume Controller 2145-8F4 node

All I/O operations that are managed by the nodes in an I/O group are cached on both nodes. Each virtual volume is defined to an I/O group. To avoid any single point of failure, the nodes of an I/O group are protected by independent uninterruptible power supplies (UPSs). There are two different UPSs. The UPSs are called the 2145 uninterruptible power supply-1U (2145 UPS-1U) or 2145 uninterruptible power supply (2145 UPS) units.

A SAN Volume Controller I/O group takes the storage that is presented to the SAN by the storage subsystems as MDisks and translates that storage into logical disks, known as VDisks, that are used by applications on the hosts. Each node must reside in only one I/O group and provide access to the VDisks in that I/O group.

The SAN Volume Controller provides continuous operations and can also optimize the data path to ensure that performance levels are maintained.

Field replaceable units (FRU) can be removed and replaced on one node while the other node of the pair continues to run. This allows the attached hosts to continue to access the attached storage while a node is repaired.

Related concepts

"VDisks" on page 23
A *virtual disk (VDisk)* is a logical disk that the cluster presents to the storage area network (SAN).

Related reference

"Supported host attachments" on page 89
The IBM Web site provides up-to-date information about the supported host attachment operating systems.

SAN fabric overview

The SAN fabric is an area of the network that contains routers, gateways, hubs, and switches. A single cluster SAN contains two distinct types of zones: a host zone and a disk zone.

In the host zone, the host systems can identify and address the SAN Volume Controller nodes. You can have more than one host zone. Generally, you create one host zone for each host type. In the disk zone, the SAN Volume Controller nodes identify the disk drives. Host systems cannot operate on the disk drives directly; all data transfer occurs through the SAN Volume Controller nodes. Figure 5 on page 8 shows several host systems that are connected in a SAN fabric.

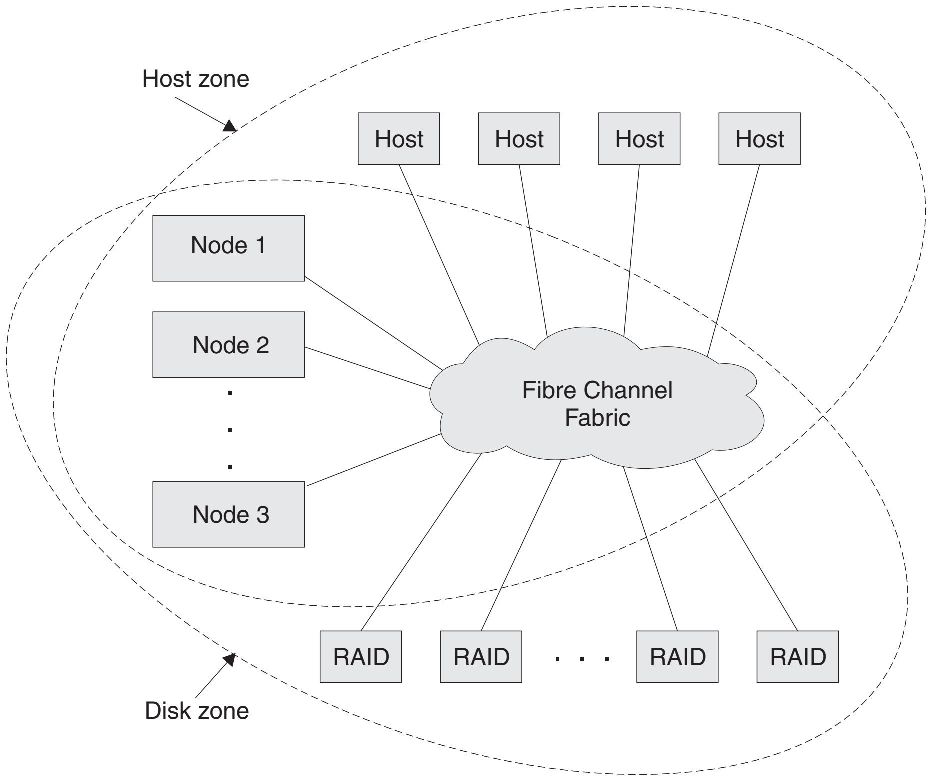


Figure 5. Example of a SAN Volume Controller in a fabric

A cluster of SAN Volume Controller nodes is connected to the same fabric and presents virtual disks (VDisks) to the host systems. You create these VDisks from units of space within a managed disk (MDisk) group. An MDisk group is a collection of MDisks that are presented by the storage subsystems (RAID controllers). The MDisk group provides a storage pool. You choose how each group is made up, and you can combine MDisks from different manufacturers' controllers in the same MDisk group.

Note: Some operating systems cannot tolerate other operating systems in the same host zone, although you might have more than one host type in the SAN fabric. For example, you can have a SAN that contains one host that runs on an AIX® operating system and another host that runs on a Windows® operating system.

You can remove one SAN Volume Controller node in each I/O group from a cluster when hardware service or maintenance is required. After you remove the SAN Volume Controller node, you can replace the field replaceable units (FRUs) in the SAN Volume Controller node. All communication between disk drives and all communication between SAN Volume Controller nodes is performed through the SAN. All SAN Volume Controller node configuration and service commands are sent to the cluster through an Ethernet network.

Each SAN Volume Controller node contains its own vital product data (VPD). Each cluster contains VPD that is common to all the SAN Volume Controller nodes in the cluster, and any system that is connected to the Ethernet network can access this VPD.

Cluster configuration information is stored on every SAN Volume Controller node that is in the cluster to allow concurrent replacement of FRUs. When a new FRU is installed and when the SAN Volume Controller node is added back into the cluster, configuration information that is required by that SAN Volume Controller node is read from other SAN Volume Controller nodes in the cluster.

SAN Volume Controller operating environment

You must set up your SAN Volume Controller operating environment using the supported multipathing software and hosts.

Minimum requirements

You must set up your SAN Volume Controller operating environment according to the following information:

- Minimum of one pair of SAN Volume Controller nodes
- Minimum of two uninterruptible power supplies
- One master console per SAN installation for configuration

Note: Depending on how you ordered your SAN Volume Controller, the master console can be preconfigured on your platform or delivered as a software-only package.

Features of a SAN Volume Controller 2145-4F2 node

The SAN Volume Controller 2145-4F2 node has the following features:

- 19-inch rack mounted enclosure
- Two 2 Gbps 2-port fibre-channel adapters (four fibre-channel ports)
- 4 GB cache memory

Features of a SAN Volume Controller 2145-8F2 node

The SAN Volume Controller 2145-8F2 node has the following features:

- 19-inch rack mounted enclosure
- Two 2 Gbps 2-port fibre-channel adapters (four fibre-channel ports)
- 8 GB cache memory

Features of a SAN Volume Controller 2145-8F4 node

The SAN Volume Controller 2145-8F4 node has the following features:

- 19-inch rack mounted enclosure
- One 4-port 4 Gbps fibre-channel adapter (four fibre-channel ports)
- 8 GB cache memory

Supported hosts

See the following Web site for a list of the supported operating systems:

http://www.ibm.com/servers/storage/software/virtualization/svc

Multipathing software

See the following Web site for the latest support and coexistence information:

http://www.ibm.com/servers/storage/software/virtualization/svc

User interfaces

The SAN Volume Controller provides the following user interfaces:

- The SAN Volume Controller Console, a Web-accessible graphical user interface (GUI) that supports flexible and rapid access to storage management information
- A command-line interface (CLI) that uses Secure Shell (SSH)

Application programming interfaces

The SAN Volume Controller provides an application programming interface called the Common Information Model (CIM) agent, which supports the Storage Management Initiative Specification (SMI-S) of the Storage Network Industry Association.

UPS

The uninterruptible power supply (UPS) provides a SAN Volume Controller node with a secondary power source if you lose power from your primary power source due to power failures, power sags, power surges, or line noise.

Unlike the traditional UPS that enables continued operation of the devices that they supply when power is lost, these UPS units are used exclusively to maintain data that is held in the SAN Volume Controller dynamic random access memory (DRAM) in the event of an unexpected loss of external power. Data is saved to the SAN Volume Controller internal disk. The UPS units are required to power the SAN Volume Controller nodes even if the input power source is uninterruptible.

| The SAN Volume Controller 2145-8F2 and SAN Volume Controller 2145-8F4 nodes
| can only operate with the 2145 UPS-1U. The SAN Volume Controller 2145-4F2
| node can operate with either the 2145 UPS or the 2145 UPS-1U.

Figure 7 on page 11 and Figure 6 provide illustrations of the two types of UPS units.

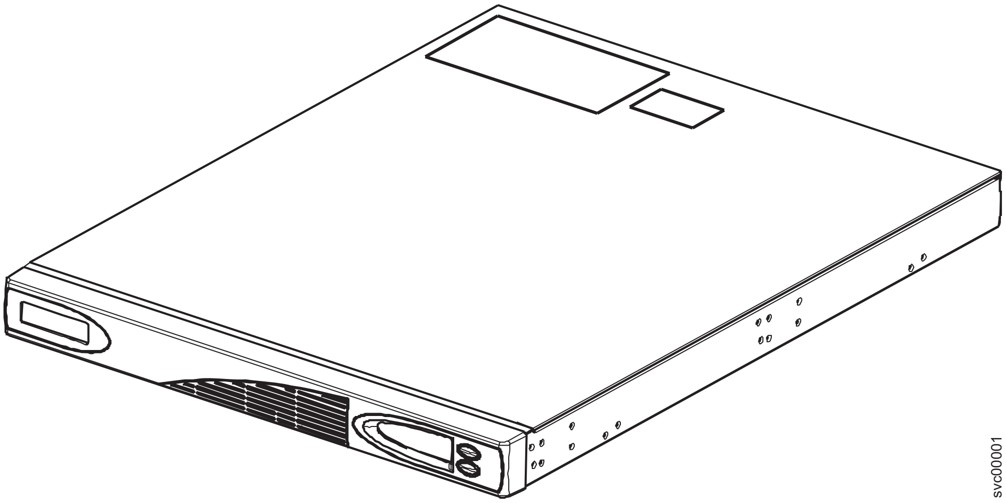


Figure 6. 2145 UPS-1U

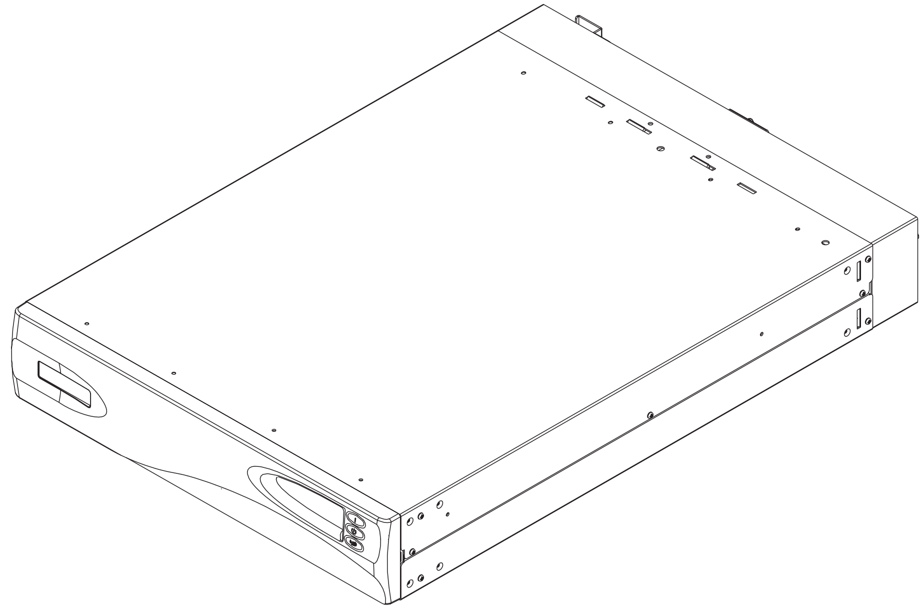


Figure 7. 2145 UPS

Note: The UPS maintains continuous SAN Volume Controller-specific communications with its attached SAN Volume Controller nodes. A SAN Volume Controller node cannot operate without the UPS. The SAN Volume Controller UPS must be used in accordance with documented guidelines and procedures and must not power any equipment other than SAN Volume Controller nodes.

UPS configuration

To provide full redundancy and concurrent maintenance, SAN Volume Controller nodes must be installed in pairs.

You must connect each SAN Volume Controller node of a pair to a different uninterruptible power supply (UPS). Each 2145 UPS can support up to two SAN Volume Controller 2145-4F2 nodes. The 2145 UPS-1U can only support one SAN Volume Controller 2145-8F4 node, one SAN Volume Controller 2145-8F2 node, or one SAN Volume Controller 2145-4F2 node. You can connect the two UPS units for the pair to different independent electrical power sources. This reduces the chance of an input power failure at both UPS units.

The UPS must be in the same rack as the nodes.

The following table provides the UPS guidelines for the SAN Volume Controller 2145-4F2:

Number of SAN Volume Controller 2145-4F2 models	Number of 2145 UPS units	Number of 2145 UPS-1U units
2	2	2
4	2	4
6	4	6
8	4	8

The following table provides the UPS guidelines for the SAN Volume Controller 2145-8F2 and the SAN Volume Controller 2145-8F4:

Number of SAN Volume Controller 2145-8F2 or SAN Volume Controller 2145-8F4 models	Number of 2145 UPS units	Number of 2145 UPS-1U units
2	Not supported	2
4	Not supported	4
6	Not supported	6
8	Not supported	8

Attention:

1. Do not connect the UPSs to an input power source that does not conform to standards.
2. Each UPS pair must power only one SAN Volume Controller cluster.

Each UPS includes power (line) cords that connect the UPS to either a rack power distribution unit (PDU), if one exists, or to an external power source.

The UPS is connected to the SAN Volume Controller nodes with a power cable and a signal cable. To avoid the possibility of power and signal cables being connected to different UPS units, these cables are wrapped together and supplied as a single field replaceable unit. The signal cables enable the SAN Volume Controller nodes to read status and identification information from the UPS.

UPS operation

Each SAN Volume Controller node monitors the operational state of the uninterruptible power supply (UPS) to which it is attached.

If the UPS reports a loss of input power, the SAN Volume Controller node stops all I/O operations and dumps the contents of its dynamic random access memory (DRAM) to the internal disk drive. When input power to the UPS is restored, the SAN Volume Controller node restarts and restores the original contents of the DRAM from the data saved on the disk drive.

A SAN Volume Controller node is not fully operational until the UPS battery charge state indicates that it has sufficient capacity to power the SAN Volume Controller node long enough to save all of its memory to the disk drive. This is in the event of a power loss. The UPS has sufficient capacity to save all the data on the SAN Volume Controller node at least twice. For a fully-charged UPS, even after battery capacity has been used to power the SAN Volume Controller node while it saves DRAM data, sufficient battery capacity remains to allow the SAN Volume Controller node to become fully operational as soon as input power is restored.

Note: If input power is disconnected from the UPS, a fully-operational SAN Volume Controller node that is connected to that UPS performs a power-down sequence. This operation, which saves the configuration and cache data to an internal disk in the SAN Volume Controller node, typically takes about three minutes, at which time power is removed from the output of the UPS. In the event of a delay in the completion of the power-down sequence, the UPS output power is removed five minutes after the power is disconnected from the UPS. Because this operation is controlled by the SAN

Volume Controller node, a UPS that is not connected to an active SAN Volume Controller node does not shut off within the five-minute required period.

Important: Data integrity can be compromised by pushing the 2145 UPS power-off button or the 2145 UPS-1U on/off button. However, in the case of an emergency, you can manually shut down the UPS by pushing the 2145 UPS power-off button or the 2145 UPS-1U on/off button. Never shut down a UPS without first shutting down the SAN Volume Controller node that it supports.

If you have two SAN Volume Controller 2145-4F2 nodes that use 2145 UPSs in the same I/O group, you must connect these nodes to different 2145 UPSs. This configuration ensures that cache and cluster state information is protected in the event of a failure of either the UPS or the mainline power source.

When SAN Volume Controller nodes are added to the cluster, you must specify the I/O group that they are joining. The configuration interfaces check the UPS units and ensure that the two SAN Volume Controller nodes in the I/O group are not connected to the same UPS units.

Cluster configuration backup functions

The SAN Volume Controller includes functions that help you to backup cluster configuration settings and business data.

To enable routine maintenance of the SAN Volume Controller clusters, the configuration settings for each cluster are stored on each node. If power fails on a cluster or if a node in a cluster is replaced, the cluster configuration settings are automatically restored when the repaired node is added to the cluster. To restore the cluster configuration in the event of a disaster (if all nodes in a cluster are lost simultaneously), plan to backup the cluster configuration settings to tertiary storage. You can use the configuration backup functions to backup the cluster configuration.

For complete disaster recovery, regularly backup the business data that is stored on virtual disks at the application server level or the host level. The SAN Volume Controller provides the following Copy Services functions that you can use to backup data: Metro Mirror and FlashCopy.

Related concepts

"Metro & Global Mirror" on page 68
The Mirror Copy Service enables you to set up a relationship between two virtual disks (VDisks), so that updates that are made by an application to one VDisk are mirrored on the other VDisk.

"FlashCopy" on page 64
FlashCopy is a Copy Service that is available with the SAN Volume Controller.

Cluster configuration backup

Cluster configuration backup is the process of extracting configuration data from a cluster and writing it to disk.

Backing up the cluster configuration enables you to restore your cluster configuration in the event that it is lost. Only the data that describes the cluster configuration is backed up. You must backup your application data using the appropriate backup methods.

Objects included in the backup

Configuration data is information about a cluster and the objects that are defined in it. Information about the following objects is included in the cluster configuration data:

- Storage subsystem
- Hosts
- Input/output (I/O) groups
- Managed disks (MDisks)
- MDisk groups
- Nodes
- Virtual disks (VDisks)
- VDisk-to-host mappings
- SSH keys
- FlashCopy mappings
- FlashCopy consistency groups
- Mirror relationships
- Mirror consistency groups

Related concepts

"Clusters" on page 58
All of your configuration and service tasks are performed at the cluster level. Therefore, after configuring your cluster, you can take advantage of the virtualization and the advanced features of the SAN Volume Controller.

Chapter 2. Object descriptions

The SAN Volume Controller is based on a group of virtualization concepts. Before setting up the system, you should understand the concepts and the objects in the system.

The smallest processing unit in a SAN Volume Controller is a single *node*. Nodes are deployed in pairs to make up a cluster. A cluster can consist of one to four pairs of nodes. Each pair of nodes is known as an *I/O group*. Each node can be in only one I/O group.

Virtual disks (VDisks) are logical disks that are presented by clusters. Each VDisk is associated with a particular I/O group. The nodes in the I/O group provide access to the VDisks in the I/O group. When an application server performs I/O to a VDisk, it can access the VDisk with either of the nodes in the I/O group. Because each I/O group only has two nodes, the distributed cache that the SAN Volume Controller provides is only two-way.

Each node does not contain any internal battery backup units and therefore must be connected to an *uninterruptible power supply (UPS)* which provides data integrity in the event of a cluster wide power failure. In such situations, the UPS maintains power to the nodes while the contents of the distributed cache are dumped to an internal drive.

The nodes in a cluster see the storage presented by back-end *disk controllers* as a number of disks, known as *managed disks (MDisks)*. Because the SAN Volume Controller does not attempt to provide recovery from physical disk failures within the back-end disk controllers, an MDisk is usually, but not necessarily, a RAID.

Each MDisk is divided up into a number of *extents* which are numbered, from 0, sequentially from the start to the end of the MDisk.

MDisks are collected into groups, known as *managed disk groups (MDisk group)*. VDisks are created from the extents contained by a MDisk group. The MDisks that constitute a particular VDisk must all come from the same MDisk group.

At any one time, a single node in the cluster can manage configuration activity. This *configuration node* manages a cache of the information that describes the cluster configuration and provides a focal point for configuration.

The SAN Volume Controller detects the fibre-channel ports that are connected to the SAN. These correspond to the worldwide port names (WWPNs) of the host bus adapter (HBA) fibre-channels that are present in the application servers. The SAN Volume Controller allows you to create logical host objects that group WWPNs that belong to a single application server or to a set of them.

Application servers can only access VDisks that have been allocated to them. VDisks can be mapped to a host object. Mapping a VDisk to a host object makes the VDisk accessible to the WWPNs in that host object, and hence the application server itself.

The SAN Volume Controller provides block-level aggregation and volume management for disk storage within the SAN. In simpler terms, this means that the SAN Volume Controller manages a number of back-end storage controllers and maps the physical storage within those controllers into logical disk images that can be seen by application servers and workstations in the SAN. The SAN is configured in such a way that the application servers cannot see the back-end physical storage. This prevents any possible conflict between the SAN Volume Controller and the application servers both trying to manage the back-end storage.

Figure 8 illustrates the objects that are described in this section and their logical positioning in a virtualized system. To simplify the example, VDisk to host mappings are not shown.

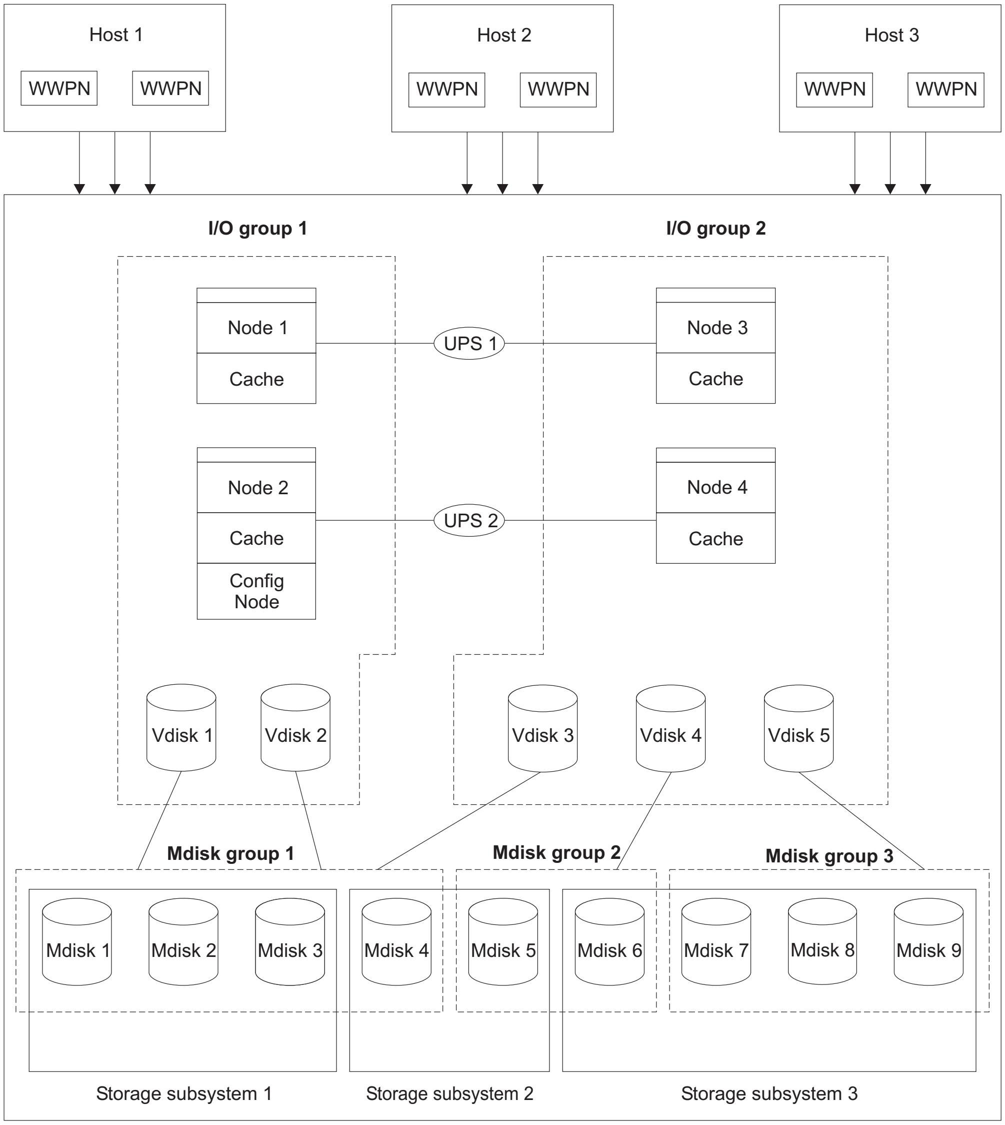


Figure 8. Objects in a virtualized system

Storage subsystems

A *storage subsystem* is a device that coordinates and controls the operation of one or more disk drives. A storage subsystem also synchronizes the operation of the drives with the operation of the system as a whole.

Storage subsystems that are attached to the SAN fabric provide the physical storage devices that the cluster detects as managed disks (MDisks). These are called RAID because the SAN Volume Controller does not attempt to provide recovery from physical disk failures within the storage subsystem. The nodes in the cluster are connected to one or more fibre-channel SAN fabrics.

Storage subsystems reside on the SAN fabric and are addressable by one or more fibre-channel ports (target ports). Each port has a unique name known as a worldwide port name (WWPN).

The exported storage devices are detected by the cluster and reported by the user interfaces. The cluster can also determine which MDisks each storage subsystem is presenting, and can provide a view of MDisks that is filtered by the storage subsystem. This allows you to associate the MDisks with the RAID that the subsystem exports.

The storage subsystem can have a local name for the RAID or single disks that it is providing. However it is not possible for the nodes in the cluster to determine this name, because the namespace is local to the storage subsystem. The storage subsystem makes the storage devices visible with a unique ID, called the logical unit number (LUN). This ID, along with the storage subsystem serial number or numbers (there can be more than one controller in a storage subsystem), can be used to associate the MDisks in the cluster with the RAID exported by the subsystem.

Storage subsystems export storage to other devices on the SAN. The physical storage that is associated with a subsystem is normally configured into RAID that provide recovery from physical disk failures. Some subsystems also allow physical storage to be configured as RAID-0 arrays (striping) or as JBODs (just a bunch of disks). However, this does not provide protection against a physical disk failure and, with virtualization, can lead to the failure of many virtual disks (VDisks). To avoid this failure, do not configure your physical storage as RAID-0 arrays or JBODs.

Many storage subsystems allow the storage that is provided by a RAID to be divided up into many SCSI logical units (LUs) that are presented on the SAN. With the SAN Volume Controller, ensure that the storage subsystems are configured to present each RAID as a single SCSI LU that are recognized by the SAN Volume Controller as a single MDisk. The virtualization features of the SAN Volume Controller can then be used to divide up the storage into VDisks.

Some storage subsystems allow the exported storage to be increased in size. The SAN Volume Controller does not use this extra capacity. Instead of increasing the size of an existing MDisk, add a new MDisk to the MDisk group and the extra capacity that are available for the SAN Volume Controller to use.

Attention: If you delete a RAID that is being used by the SAN Volume Controller, the MDisk group goes offline and the data in that group is lost.

The cluster detects and provides a view of the storage subsystems that the SAN Volume Controller supports. The cluster can also determine which MDisks each subsystem has and can provide a view of MDisks that are filtered by the device. This view enables you to associate the MDisks with the RAID that the subsystem presents.

Note: The SAN Volume Controller supports storage that is internally configured as a RAID. However, it is possible to configure a storage subsystem as a non-RAID device. RAID provides redundancy at the disk level. For RAID devices, a single physical disk failure does not cause an MDisk failure, an MDisk group failure, or a failure in the VDisks that were created from the MDisk group.

MDisks

A *managed disk (MDisk)* is a logical disk (typically a RAID or partition thereof) that a storage subsystem has exported to the SAN fabric to which the nodes in the cluster are attached.

An MDisk might, therefore, consist of multiple physical disks that are presented as a single logical disk to the SAN. An MDisk always provides usable blocks of physical storage to the cluster even if it does not have a one-to-one correspondence with a physical disk.

Each MDisk is divided into a number of extents, which are numbered, from 0, sequentially from the start to the end of the MDisk. The extent size is a property of MDisk groups. When an MDisk is added to an MDisk group, the size of the extents that the MDisk is divided into depends on the attribute of the MDisk group to which it has been added.

Access modes

The access mode determines how the cluster uses the MDisk. The following list provides the three types of possible access modes:

Unmanaged
The MDisk is not used by the cluster.

Managed
The MDisk is assigned to an MDisk group and provides extents that virtual disks (VDisks) can use.

Image The MDisk is assigned directly to a VDisk with a one-to-one mapping of extents between the MDisk and the VDisk.

Attention: If you add an MDisk that contains existing data to an MDisk group while the MDisk is in unmanaged or managed mode, you lose the data that it contains. The *image mode* is the only mode that preserves this data.

Figure 9 on page 19 shows physical disks and MDisks.

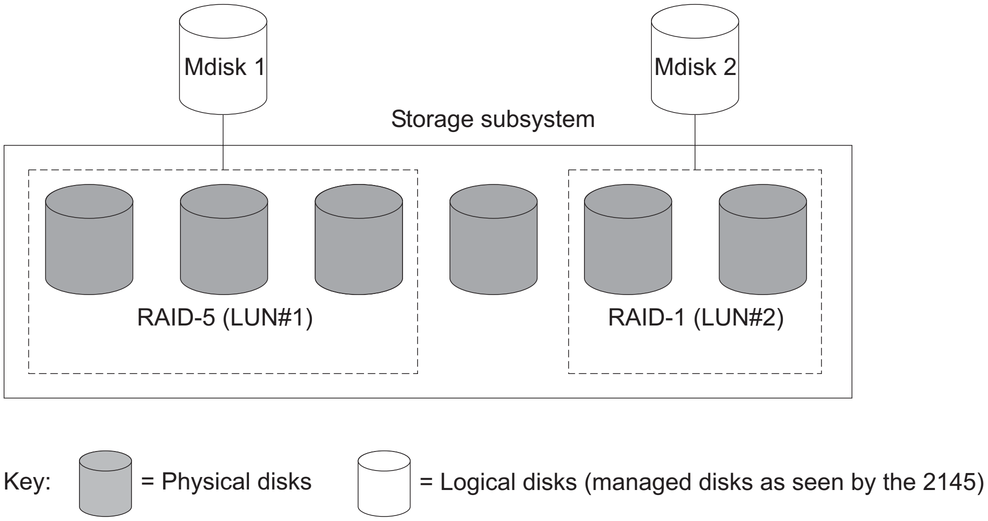


Figure 9. Controllers and MDisks

Table 1 describes the operational states of an MDisk.

Table 1. MDisk status

Status	**Description**
Online	The MDisk can be accessed by all online nodes. That is, all the nodes that are currently working members of the cluster can access this MDisk. The MDisk is online when the following conditions are met: • All timeout error recovery procedures complete and report the disk as online. • Logical unit number (LUN) inventory of the target ports correctly reported the MDisk. • Discovery of this LUN completed successfully. • All of the MDisk target ports report this LUN as available with no fault conditions.
Degraded	The MDisk cannot be accessed by all the online nodes. That is, one or more (but not all) of the nodes that are currently working members of the cluster cannot access this MDisk. The MDisk can be partially excluded; that is, some of the paths to the MDisk (but not all) have been excluded.
Excluded	The MDisk has been excluded from use by the cluster after repeated access errors. Run the Directed Maintenance Procedures to determine the problem.
Offline	The MDisk cannot be accessed by any of the online nodes. That is, all of the nodes that are currently working members of the cluster cannot access this MDisk. This state can be caused by a failure in the SAN, the storage subsystem, or one or more physical disks connected to the storage subsystem. The MDisk is reported as offline if all paths to the disk fail.

Extents

Each MDisk is divided into chunks of equal size called *extents*. Extents are a unit of mapping that provides the logical connection between MDisks and VDisks.

Attention: If you have observed intermittent breaks in links or if you have been replacing cables or connections in the SAN fabric, you might have one or more MDisks in degraded status. If an I/O operation is attempted when a link is broken and the I/O operation fails several times, the system partially excludes the MDisk and it changes the status of the MDisk to degraded. You must include the MDisk to resolve the problem. You can include the MDisk by either selecting **Work with Managed Disks** → **Managed Disk** → **Include an MDisk** in the SAN Volume Controller Console, or by issuing the following command in the command-line interface (CLI):

```
svctask includemdisk mdiskname/id
```

Where *mdiskname/id* is the name or ID of your MDisk.

MDisk path

Each MDisk has an online path count, which is the number of nodes that have access to that MDisk; this represents a summary of the I/O path status between the cluster nodes and the storage device. The maximum path count is the maximum number of paths that have been detected by the cluster at any point in the past. If the current path count is not equal to the maximum path count, the MDisk might be degraded. That is, one or more nodes might not see the MDisk on the fabric.

MDisk groups

A *managed disk (MDisk) group* is a collection of MDisks that jointly contain all the data for a specified set of virtual disks (VDisks).

Figure 10 shows an MDisk group containing four MDisks.

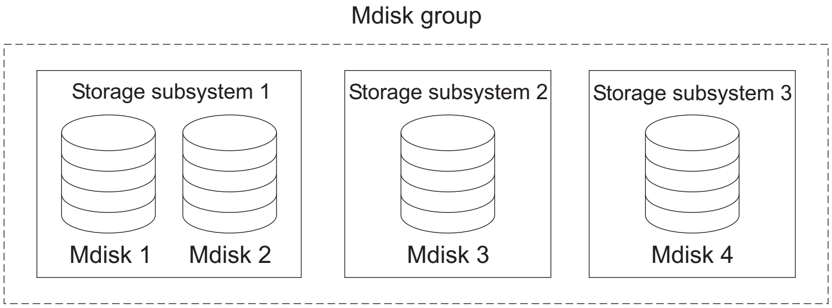


Figure 10. MDisk group

All MDisks in a group are split into extents of the same size. VDisks are created from the extents that are available in the group. You can add MDisks to an MDisk group at any time either to increase the number of extents that are available for new VDisks or to expand existing VDisks.

Note: RAID partitions on HP StorageWorks subsystems are only supported in single-port attach mode. MDisk groups that consist of single-port attached subsystems and other storage subsystems are not supported.

You can add only MDisks that are in unmanaged mode. When MDisks are added to a group, their mode changes from unmanaged to managed.

You can delete MDisks from a group under the following conditions:

- VDisks are not using any of the extents that are on the MDisk.
- Enough free extents are available elsewhere in the group to move any extents that are in use from this MDisk.

Attention: If you delete an MDisk group, you destroy all the VDisks that are made from the extents that are in the group. If the group is deleted, you cannot recover the mapping that existed between extents that are in the group or the extents that the VDisks use. The MDisks that were in the group are returned to unmanaged mode and can be added to other groups. Because the deletion of a group can cause a loss of data, you must force the deletion if VDisks are associated with it.

Table 2 describes the operational states of an MDisk group.

Table 2. MDisk group status

Status	**Description**
Online	The MDisk group is online and available. All the MDisks in the group are available.
Degraded	The MDisk group is available; however, one or more nodes cannot access all the MDisks in the group.
Offline	The MDisk group is offline and unavailable. No nodes in the cluster can access the MDisks. The most likely cause is that one or more MDisks are offline or excluded.

Attention: If a single MDisk in an MDisk group is offline and therefore cannot be seen by any of the online nodes in the cluster, then the MDisk group of which this MDisk is a member goes offline. This causes *all* the VDisks that are being presented by this MDisk group to go offline. Take care when you create MDisk groups to ensure an optimal configuration.

Consider the following guidelines when you create MDisk groups:

- Allocate your image-mode VDisks between your MDisk groups.
- Ensure that all MDisks that are allocated to a single MDisk group are the same RAID type. This ensures that a single failure of a physical disk in the storage subsystem does not take the entire group offline. For example, if you have three RAID-5 arrays in one group and add a non-RAID disk to this group, you lose access to all the data striped across the group if the non-RAID disk fails. Similarly, for performance reasons, you must not mix RAID types. The performance of all VDisks is reduced to the lowest performer in the group.
- If you intend to keep the VDisk allocation within the storage exported by a storage subsystem, ensure that the MDisk group that corresponds with a single subsystem is presented by that subsystem. This also enables nondisruptive migration of data from one subsystem to another subsystem and simplifies the decommissioning process if you want to decommission a controller at a later time.
- Except when you migrate between groups, you must associate a VDisk with just one MDisk group.
- An MDisk can be associated with just one MDisk group.

Extents

To track the space that is available on an MDisk, the SAN Volume Controller divides each MDisk into chunks of equal size. These chunks are called *extents* and are indexed internally. Extent sizes can be 16, 32, 64, 128, 256, or 512 MB.

You specify the extent size when you create a new MDisk group. You cannot change the extent size later; it must remain constant throughout the lifetime of the MDisk group.

Ensure that your MDisk groups *do not* have different extent sizes. Different extent sizes place restrictions on the use of data migration. You cannot use the SAN Volume Controller data migration function to move a VDisk between MDisk groups that have different extent sizes.

You can use Copy Services to copy a VDisk between MDisk groups that have different extent sizes using the following options:

- FlashCopy to copy a VDisk between a source and a destination MDisk group that have different extent sizes.
- Intra-cluster Metro or Global Mirror to copy a VDisk between a source and a destination MDisk group that have different extent sizes.

The choice of extent size affects the total amount of storage that is managed by the cluster. Table 3 shows the maximum amount of storage that can be managed by a cluster for each extent size.

Table 3. Capacities of the cluster given extent size

Extent size	Maximum storage capacity of cluster
16 MB	64 TB

Table 3. Capacities of the cluster given extent size (continued)

Extent size	Maximum storage capacity of cluster
32 MB	128 TB
64 MB	256 TB
128 MB	512 TB
256 MB	1 PB
512 MB	2 PB

A cluster can manage 4 million extents (4 x 1024 x 1024). For example, with a 16 MB extent size, the cluster can manage up to 16 MB x 4 MB = 64 TB of storage.

When you choose an extent size, consider your future needs. For example, if you currently have 40 TB of storage, and you specify an extent size of 16 MB, the capacity of the MDisk group is limited to 64 TB of storage in the future. If you select an extent size of 64MB, the capacity of the MDisk group is 256 TB.

Using a larger extent size can waste storage. When a VDisk is created, the storage capacity for the VDisk is rounded to a whole number of extents. If you configure the system to have a large number of small VDisks and you use a large extent size, this can cause storage to be wasted at the end of each VDisk.

VDisks

A *virtual disk (VDisk)* is a logical disk that the cluster presents to the storage area network (SAN).

Application servers on the SAN access VDisks, not managed disks (MDisks). VDisks are created from a set of extents in an MDisk group. There are three types of VDisks: striped, sequential, and image.

Types

You can create the following types of VDisks:

Striped

A VDisk that has been striped is at the extent level. One extent is allocated, in turn, from each MDisk that is in the group. For example, an MDisk group that has 10 MDisks takes one extent from each MDisk. The 11th extent is taken from the first MDisk, and so on. This procedure, known as a round-robin, is similar to RAID-0 striping.

You can also supply a list of MDisks to use as the stripe set. This list can contain two or more MDisks from the MDisk group. The round-robin procedure is used across the specified stripe set.

Attention: By default, striped VDisks are striped across all MDisks in the group. If some of the MDisks are smaller than others, the extents on the smaller MDisks are used up before the larger MDisks run out of extents. Manually specifying the stripe set in this case might result in the VDisk not being created.

If you are unsure if there is sufficient free space to create a striped VDisk, select one of the following options:

- Check the free space on each MDisk in the group using the **svcinfo lsfreeextents** command.
- Let the system automatically create the VDisk by not supplying a specific stripe set.

Figure 11 shows an example of an MDisk group that contains three MDisks. This figure also shows a striped VDisk that is created from the extents that are available in the group.

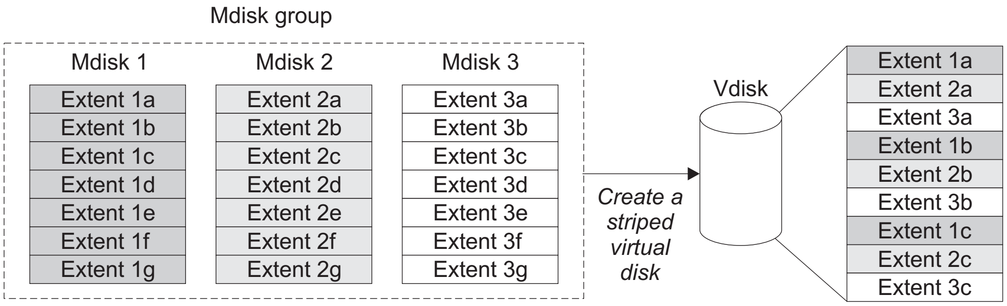


Figure 11. MDisk groups and VDisks

Sequential
When extents are selected, they are allocated sequentially on one MDisk to create the VDisk if enough consecutive free extents are available on the chosen MDisk.

Image Image-mode VDisks are special VDisks that have a direct relationship with one MDisk. If you have an MDisk that contains data that you want to merge into the cluster, you can create an image-mode VDisk. When you create an image-mode VDisk, a direct mapping is made between extents that are on the MDisk and extents that are on the VDisk. The MDisk is not virtualized. The logical block address (LBA) *x* on the MDisk is the same as LBA *x* on the VDisk.

When you create an image-mode VDisk, you must assign it to an MDisk group. An image-mode VDisk must be at least one extent in size. The minimum size of an image-mode VDisk is the extent size of the MDisk group to which it is assigned.

The extents are managed in the same way as other VDisks. When the extents have been created, you can move the data onto other MDisks that are in the group without losing access to the data. After you move one or more extents, the VDisk becomes a virtualized disk, and the mode of the MDisk changes from image to managed.

Attention: If you add a managed mode MDisk to an MDisk group, any data on the MDisk is lost. Ensure that you create image-mode VDisks from the MDisks that contain data before you start adding any MDisks to groups.

MDisks that contain existing data have an initial mode of unmanaged, and the cluster cannot determine if it contains partitions or data.

A VDisk can be in one of three states: online, offline, and degraded. Table 4 describes the different states of a VDisk.

Table 4. VDisk status

Status	Description
Online	The VDisk is online and available if both nodes in the I/O group can access the VDisk. A single node can only access a VDisk if it can access all the MDisks in the MDisk group that are associated with the VDisk.
Offline	The VDisk is offline and unavailable if both nodes in the I/O group are missing or none of the nodes in the I/O group that are present can access the VDisk. The VDisk can also be offline if the VDisk is the secondary of a Mirror relationship that is not synchronized.
Degraded	The status of the VDisk is degraded if one node in the I/O group is online and the other node is either missing or cannot access the VDisk.

You can use more sophisticated extent allocation policies to create VDisks. When you create a striped VDisk, you can specify the same MDisk more than once in the list of MDisks that are used as the stripe set. This is useful if you have an MDisk group in which not all the MDisks are of the same capacity. For example, if you have an MDisk group that has two 18 GB MDisks and two 36 GB MDisks, you can create a striped VDisk by specifying each of the 36 GB MDisks twice in the stripe set so that two-thirds of the storage is allocated from the 36 GB disks.

If you delete a VDisk, you destroy access to the data that is on the VDisk. The extents that were used in the VDisk are returned to the pool of free extents that is in the MDisk group. The deletion might fail if the VDisk is still mapped to hosts. The deletion might also fail if the VDisk is still part of a FlashCopy or a Mirror mapping. If the deletion fails, you can specify the force-delete flag to delete both the VDisk and the associated mappings to hosts. Forcing the deletion deletes the Copy Services relationship and mappings.

Cache modes

You can select to have read and write operations stored in cache by specifying a cache mode. You must specify the cache mode when you create the VDisk. After the VDisk is created, you cannot change the cache mode.

Table 5 describes the two types of cache modes for a VDisk.

Table 5. VDisk cache modes

Cache mode	Description
readwrite	All read and write I/O operations that are performed by the VDisk are stored in cache. This is the default cache mode for all VDisks.

Table 5. VDisk cache modes (continued)

Cache mode	Description
none	All read and write I/O operations that are performed by the VDisk are not stored in cache.

Related concepts

Chapter 1, "Virtualization," on page 1
Virtualization is a concept that applies to many areas of the information technology industry.

VDisk-to-host mapping

Virtual disk (VDisk)-to-host mapping is the process of controlling which hosts have access to specific VDisks within the SAN Volume Controller cluster.

VDisk-to-host mapping is similar in concept to logical unit number (LUN) mapping or masking. LUN mapping is the process of controlling which hosts have access to specific logical units (LUs) within the disk controllers. LUN mapping is typically done at the disk controller level. VDisk-to-host mapping is done at the SAN Volume Controller level.

Application servers can only access VDisks that have been made accessible to them. The SAN Volume Controller detects the fibre-channel ports that are connected to the SAN. These correspond to the host bus adapter (HBA) worldwide port names (WWPNs) that are present in the application servers. The SAN Volume Controller enables you to create logical hosts that group together WWPNs that belong to a single application server. VDisks can then be mapped to a host. The act of mapping a VDisk to a host makes the VDisk accessible to the WWPNs in that host and the application server itself.

VDisks and host mappings

LUN masking usually requires device driver software on each host. The device driver software masks the LUNs. After the masking is complete, only some disks are visible to the operating system. The SAN Volume Controller performs a similar function, but, by default, it presents to the host only those VDisks that are mapped to that host. Therefore, you must map the VDisks to the hosts that you want to access those disks.

Each host mapping associates a VDisk with a host object and allows all HBA ports in the host object to access the VDisk. You can map a VDisk to multiple host objects. When a mapping is created, multiple paths might exist across the SAN fabric from the hosts to the SAN Volume Controller nodes that are presenting the VDisk. Most operating systems present each path to a VDisk as a separate storage device. The SAN Volume Controller, therefore, requires that multipathing software be running on the host. The multipathing software manages the many paths that are available to the VDisk and presents a single storage device to the operating system.

When you map a VDisk to a host, you can optionally specify a SCSI ID for the VDisk. This ID controls the sequence in which the VDisks are presented to the host. For example, if you present three VDisks to the host, and those VDisks have SCSI IDs of 0, 1, and 3, the VDisk that has an ID of 3 might not be found because no disk is mapped with an ID of 2. The cluster automatically assigns the next available SCSI ID if none is entered.

Figure 12 and Figure 13 show two VDisks, and the mappings that exist between the host objects and these VDisks.

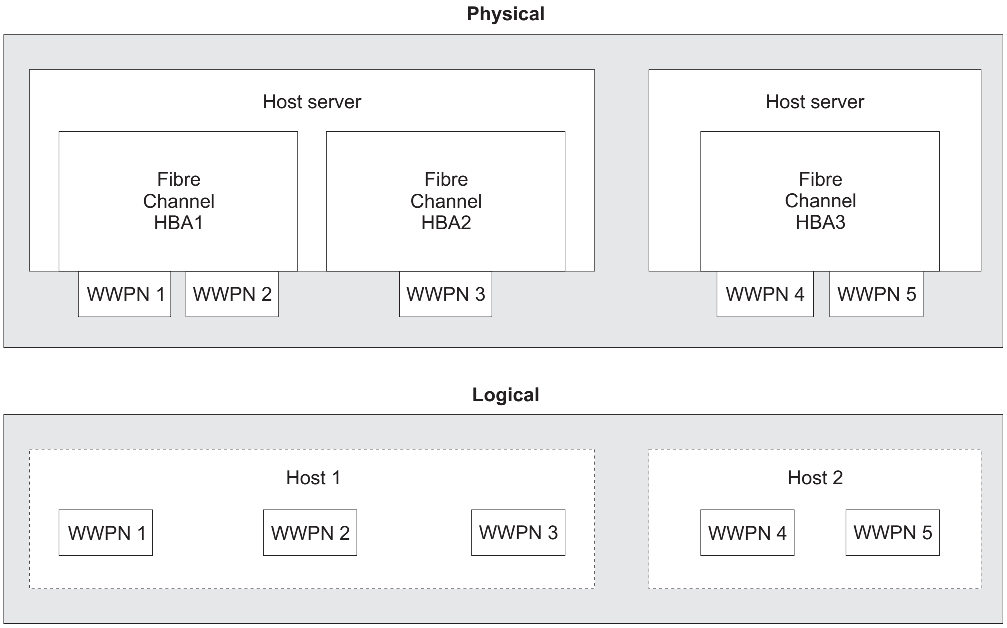


Figure 12. Hosts, WWPNs, and VDisks

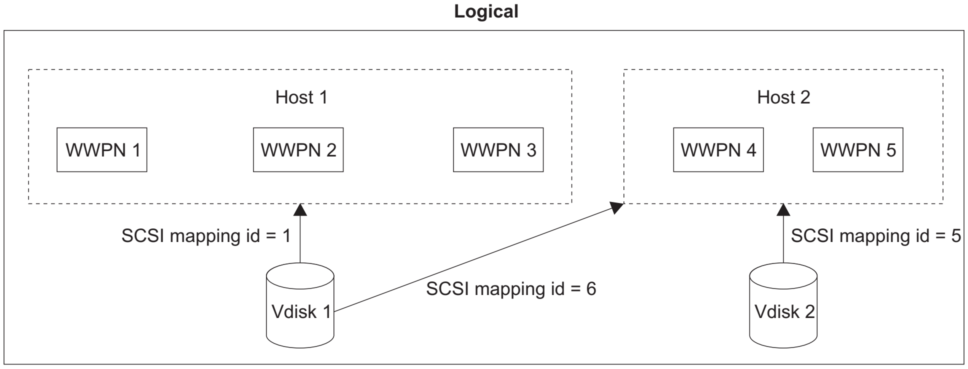


Figure 13. Hosts, WWPNs, VDisks and SCSI mappings

Host objects

A *host system* is an open-systems computer that is connected to the switch through a fibre-channel interface.

A *host object* is a logical object that groups one or more worldwide port names (WWPNs) of the host bus adapters (HBAs) that the cluster has detected on the SAN. A typical configuration has one host object for each host that is attached to the SAN. If a cluster of hosts accesses the same storage, you can add HBA ports from several hosts to one host object to make a simpler configuration.

The cluster does not automatically present virtual disks (VDisks) on the fibre-channel ports. You must map each VDisk to a particular set of ports to enable the VDisk to be accessed through those ports. The mapping is made between a host object and a VDisk.

When you create a new host object, the configuration interfaces provide a list of unconfigured WWPNs. These WWPNs represent the fibre-channel ports that the cluster has detected.

The cluster can detect only ports that are logged into the fabric. Some HBA device drivers do not let the ports remain logged in if no disks are visible on the fabric. This condition causes a problem when you want to create a host because, at this time, no VDisks are mapped to the host. The configuration interface provides a method that allows you to manually type the port names.

Attention: You must not include a node port in a host object.

A port can be added to only one host object. When a port has been added to a host object, that port becomes a configured WWPN, and is not included in the list of ports that are available to be added to other hosts.

Port masks

You can use a port mask to control the node target ports that a host can access. The port mask applies to logins from the host initiator port that are associated with the host object.

For each login between a host HBA port and node port, the node examines the port mask that is associated with the host object for which the host HBA is a member and determines if access is allowed or denied. If access is denied, the node responds to SCSI commands as if the HBA port is unknown.

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.

Multiple target ports

When you create a VDisk-to-host mapping, the host ports that are associated with the host object can see the LUN that represents the VDisk on up to eight fibre-channel ports. Nodes follow the ANSI FC standards for SCSI LUs that are accessed through multiple node ports. However, you must coordinate the nodes in an I/O group to present a consistent SCSI LU across all ports that can access it. The ANSI FC standards do not require that the same LUN is used on all ports; however, nodes always present the LU that represents a specific VDisk with the same LUN on all ports in an I/O group.

Node login counts

The number of nodes that can see each port is reported on a per node basis and is known as the node login count. If the count is less than the number of nodes in the cluster, there is a fabric problem, and not all nodes can see the port.

Standard and persistent reserves

The SCSI Reserve command and the SCSI Persistent Reserve command are specified by the SCSI standards. Servers can use these commands to prevent HBA ports in other servers from accessing the LUN.

This prevents accidental data corruption that is caused when a server overwrites data on another server. The Reserve and Persistent Reserve commands are often used by clustering software to control access to SAN Volume Controller virtual disks (VDisks).

If a server is not shut down or removed from the server cluster in a controlled way, the server reserves and persistent reserves are maintained. This prevents other servers from accessing data that is no longer in use by the server that holds the reservation. In this situation, you might want to break the reservation and allow a new server to access the VDisk.

When possible, you should have the server that holds the reservation explicitly release the reservation to ensure that the server cache is flushed and the server software is aware that access to the VDisk is lost. In circumstances where this is not possible, you can use operating system specific tools to remove reservations. Consult the operating system documentation for details.

When you use the **svctask rmvdiskhostmap** CLI command to remove VDisk-to-host mappings, SAN Volume Controller nodes with a software level of 4.1.0 or later can remove the reservations and persistent reservations that the host has on the VDisk.

Chapter 3. Planning for your SAN Volume Controller installation

Before the service representative can start to set up your SAN Volume Controller, you must verify that the prerequisite conditions for the SAN Volume Controller and uninterruptible power supply (UPS) installation are met.

1. Does your physical site meet the environment requirements for the SAN Volume Controller and UPS?
2. Do you have adequate rack space for your hardware? Ensure you have the following rack space for your components:
   - SAN Volume Controller: One Electrical Industries Association (EIA) unit high for each node.
   - 2145 uninterruptible power supply (2145 UPS): Two EIA units high for each 2145 UPS.
   - 2145 uninterruptible power supply-1U (2145 UPS-1U): One EIA unit high for each 2145 UPS-1U.
3. Do you have power distribution units in the rack to provide power to the UPS units?

   A clearly visible and accessible emergency power off switch is required.
4. Ensure that you provide appropriate connectivity by preparing your environment.

Preparing your SAN Volume Controller environment

Before installing the SAN Volume Controller, you must prepare the physical environment.

Preparing your SAN Volume Controller 2145-8F2 or SAN Volume Controller 2145-8F4 environment

The following four tables list the physical dimensions and weight of the node, as well as other environmental requirements that you must consider before you install your SAN Volume Controller 2145-8F2 or SAN Volume Controller 2145-8F4:

Dimensions and weight

Height	Width	Depth	Maximum weight
43 mm (1.69 in.)	440 mm (17.32 in.)	686 mm (27 in.)	12.7 kg (28 lb)

Additional space requirements

Location	Additional space requirements	Reason
Left and right sides	50 mm (2 in.)	Cooling air flow
Back	Minimum: 100 mm (4 in.)	Cable exit

AC input-voltage requirements

Power supply assembly type	Voltage	Frequency
200 to 240 V	88 to 264 V ac	50 or 60 Hz

Environment

Environment	Temperature	Altitude	Relative humidity	Maximum wet bulb temperature
Operating in lower altitudes	10°C to 35°C (50°F to 95°F)	0 to 914 m (0 to 2998 ft)	8% to 80% noncondensing	23°C (74°F)
Operating in higher altitudes	10°C to 32°C (50°F to 88°F)	914 to 2133 m (2998 to 6988 ft)	8% to 80% noncondensing	23°C (74°F)
Powered off	10°C to 43°C (50°F to 110°F)	–	8% to 80% noncondensing	27°C (81°F)
Storing	1°C to 60°C (34°F to 140°F)	0 to 2133 m (0 to 6988 ft)	5% to 80% noncondensing	29°C (84°F)
Shipping	-20°C to 60°C (-4°F to 140°F)	0 to 10668 m (0 to 34991 ft)	5% to 100% condensing, but no precipitation	29°C (84°F)

Preparing your SAN Volume Controller 2145-4F2 environment

The following four tables list the physical dimensions and weight of the SAN Volume Controller 2145-4F2 node, as well as other environmental requirements that you must consider before you install your SAN Volume Controller 2145-4F2.

Dimensions and weight

Height	Width	Depth	Maximum weight
43 mm (1.69 in.)	440 mm (17.32 in.)	686 mm (27 in.)	12.7 kg (28 lb.)

Additional space requirements

Location	Additional space requirements	Reason
Left and right sides	50 mm (2 in.)	Cooling air flow
Back	Minimum: 100 mm (4 in.)	Cable exit

AC input-voltage requirements

Power supply assembly type	Voltage	Frequency
200 to 240 V	88 to 264 V ac	50 or 60 Hz

Environment

Environment	Temperature	Altitude	Relative humidity	Maximum wet bulb temperature
Operating in lower altitudes	10°C to 35°C (50°F to 95°F)	0 to 914 m (0 to 2998 ft.)	8% to 80% noncondensing	23°C (74°F)
Operating in higher altitudes	10°C to 32°C (50°F to 88°F)	914 to 2133 m (2998 to 6988 ft.)	8% to 80% noncondensing	23°C (74°F)
Powered off	10°C to 43°C (50°F to 110°F)	-	8% to 80% noncondensing	27°C (81°F)
Storing	1°C to 60°C (34°F to 140°F)	0 to 2133 m (0 to 6988 ft.)	5% to 80% noncondensing	29°C (84°F)
Shipping	-20°C to 60°C (-4°F to 140°F)	0 to 10668 m (0 to 34991 ft.)	5% to 100% condensing, but no precipitation	29°C (84°F)

Heat output

The heat output (maximum) is 350 watts (1195 Btu per hour).

Power cables for the 2145 UPS-1U

You must follow your country or region’s power requirements to choose the appropriate power cable for the 2145 uninterruptible power supply-1U (2145 UPS-1U).

The following table lists the power cable requirements for your country or region:

Country or region	Length	Connection type (attached plug designed for 200-240V AC input)	Part
United States of America (Chicago), Canada, Mexico	1.8 m (6 ft)	NEMA L6-15P	7842122
Bahamas, Barbados, Bermuda, Bolivia, Brazil, Canada, Cayman Islands, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Japan, Korea (South), Liberia, Mexico, Netherlands Antilles, Nicaragua, Panama, Peru, Philippines, Saudi Arabia, Suriname, Taiwan, Trinidad (West Indies), United States of America, Venezuela	2.8 m (9 ft)	NEMA L6-15P	7842123

Country or region	Length	Connection type (attached plug designed for 200-240V AC input)	Part
Antigua, Bahrain, Brunei, Channel Islands, Hong Kong S.A.R. of China, Cyprus, Dubai, Fiji, Ghana, India, Iraq, Ireland, Kenya, Kuwait, Malawi, Malaysia, Malta, Nepal, Nigeria, Polynesia, Qatar, Sierra Leone, Singapore, Tanzania, Uganda, United Kingdom, Yemen, Zambia	2.8 m (9 ft)	BS 1363/A	14F0033
Argentina, Australia, China (PRC), New Zealand, Papua New Guinea, Paraguay, Uruguay, Western Samoa	2.8 m (9 ft)	AZ/NZS C112	13F9940
Afghanistan, Algeria, Andorra, Angola, Austria, Belgium, Benin, Bulgaria, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, China (Macau S.A.R.), Czech Republic, Egypt, Finland, France, French Guiana, Germany, Greece, Guinea, Hungary, Iceland, Indonesia, Iran, Ivory Coast, Jordan, Lebanon, Luxembourg, Malagasy, Mali, Martinique, Mauritania, Mauritius, Monaco, Morocco, Mozambique, Netherlands, New Caledonia, Niger, Norway, Poland, Portugal, Romania, Senegal, Slovakia, Spain, Sudan, Sweden, Syria, Togo, Tunisia, Turkey, former USSR, Vietnam, former Yugoslavia, Zaire, Zimbabwe	2.8 m (9 ft)	CEE 7-VII	13F9979
Denmark	2.8 m (9 ft)	DK2-5a	13F9997
Bangladesh, Burma, Pakistan, South Africa, Sri Lanka	2.8 m (9 ft)	SABS 164	14F0015
Liechtenstein, Switzerland	2.8 m (9 ft)	1011-S2450 7	14F0051
Chile, Ethiopia, Italy, Libya, Somalia	2.8 m (9 ft)	CEI 23-16	14F0069
Israel	2.8 m (9 ft)	SI 32	14F0087

Power cables for the 2145 UPS

You must follow your country or region’s power requirements to choose the appropriate power cable for the 2145 uninterruptible power supply (2145 UPS).

The following table lists the power cable requirements for your country or region:

Country or region	Length	Connection type (attached plug designed for 200-240V AC input)	Part
United States of America (Chicago), Canada, Mexico	1.8 m (6 ft)	NEMA L6-15P	14F1549
Bahamas, Barbados, Bermuda, Bolivia, Brazil, Canada, Cayman Islands, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Japan, Korea (South), Liberia, Mexico, Netherlands Antilles, Nicaragua, Panama, Peru, Philippines, Saudi Arabia, Suriname, Taiwan, Trinidad (West Indies), United States of America, Venezuela	2.5 m (8 ft)	NEMA L6-15P	12J5119
Antigua, Bahrain, Brunei, Channel Islands, China (Hong Kong S.A.R.), Cyprus, Denmark, Dubai, Fiji, Ghana, India, Iraq, Ireland, Kenya, Kuwait, Malawi, Malaysia, Malta, Nepal, Nigeria, Polynesia, Qatar, Sierra Leone, Singapore, Tanzania, Uganda, United Kingdom, Yemen, Zambia	2.5 m (8 ft)	IEC 309	36L8822
Argentina, Australia, China (PRC), New Zealand, Papua New Guinea, Paraguay, Uruguay, Western Samoa	2.5 m (8 ft)	L6-20P	12J5118
Afghanistan, Albania, Algeria, Andorra, Angola, Austria, Belgium, Benin, Bulgaria, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, China (Macau S.A.R.), Czech Republic, Egypt, Finland, France, French Guiana, Germany, Greece, Guinea, Hungary, Iceland, Indonesia, Iran, Ivory Coast, Jordan, Lebanon, Luxembourg, Malagasy, Mali, Martinique, Mauritania, Mauritius, Monaco, Morocco, Mozambique, Netherlands, New Caledonia, Niger, Norway, Poland, Portugal, Romania, Senegal, Slovakia, Spain, Sudan, Sweden, Syria, Togo, Tunisia, Turkey, former USSR, Vietnam, former Yugoslavia, Zaire, Zimbabwe	2.5 m (8 ft)	CEE7	55H6643
Bangladesh, Burma, Pakistan, South Africa, Sri Lanka	2.5 m (8 ft)	SABS 164	12J5124
Thailand	2.5 m (8 ft)	NEMA 6-15P	12J5120

Preparing your UPS environment

Ensure that your physical site meets the installation requirements for the uninterruptible power supply (UPS).

The 2145 UPS-1U

When you configure the 2145 uninterruptible power supply-1U (2145 UPS-1U), the voltage that is supplied to it must be 200 – 240 V, single phase.

Note: The 2145 UPS-1U has an integrated circuit breaker and does not require external protection.

The 2145 UPS

The SAN Volume Controller 2145-8F2 and SAN Volume Controller 2145-8F4 support the 2145 uninterruptible power supply-1U (2145 UPS-1U) but do not support the 2145 uninterruptible power supply (2145 UPS). The SAN Volume Controller 2145-4F2 supports both the 2145 UPS-1U and the 2145 UPS.

Use the following considerations when configuring the 2145 uninterruptible power supply (2145 UPS):

- Each 2145 UPS must be connected to a separate branch circuit.
- A UL-listed 15 A circuit breaker must be installed in each branch circuit that supplies power to the 2145 UPS.
- The voltage that is supplied to the 2145 UPS must be 200 – 240 V, single phase.
- The frequency supplied to the 2145 UPS must be 50 or 60 Hz.

Attention: Ensure that you comply with the following requirements for UPSs:

- If the UPS is cascaded from another UPS, the source UPS must have at least three times the capacity per phase and the total harmonic distortion must be less than 5% with any single harmonic being less than 1%.
- The UPS must also have input voltage capture that has a slew rate faster than 3 Hz per second and 1 msec glitch rejection.

UPS specifications

2145 UPS-1U dimensions and weight

Height	Width	Depth	Maximum weight
44 mm (1.73 in.)	439 mm (17.3 in.)	579 mm (22.8 in.)	18.8 kg (41.4 lb)

2145 UPS dimensions and weight

Height	Width	Depth	Maximum weight
89 mm (3.5 in.)	483 mm (19 in.)	622 mm (24.5 in.)	37 kg (84 lb)

AC input-voltage requirements

	2145 UPS-1U	2145 UPS
Power Rating	750 VA/520 W	3000 VA/2700 W
Voltage	200 – 240 V	200 – 240 V
Frequency	50 or 60 Hz	50 or 60 Hz

Environment

	Operating environment	Non-operating environment	Storing environment	Shipping environment
Air temperature	0°C – 40°C (32°F – 104°F)	0°C – 40°C (32°F – 104°F)	0°C – 25°C (32°F – 77°F)	–25°C – 55°C (–13°F – 131°F)
Relative humidity	5% – 95% non-condensing	5% – 95% non-condensing	5% – 95% non-condensing	5% – 95% non-condensing

Altitude

	Operating environment	Non-operating environment	Storing environment	Shipping environment
Altitude (from sea level)	0 – 2000 m (0 – 6560 ft)	0 – 2000 m (0 – 6560 ft)	0 – 2000 m (0 – 6560 ft)	0 – 15 000 m (0 – 49212 ft)

Heat output (maximum)

The heat output parameters are the following:

- 142 watts (485 Btu per hour) during normal operation
- 553 watts (1887 Btu per hour) when power has failed and the UPS is supplying power to the nodes of the SAN Volume Controller

Ports and connections

Ensure that you are familiar with the specific ports and connections types for the SAN Volume Controller and the uninterruptible power supply (UPS).

Each SAN Volume Controller requires the following ports and connections:

- Each SAN Volume Controller node requires one Ethernet cable to connect it to an Ethernet switch or hub. A 10/100 Mb Ethernet connection is required.
- Two TCP/IP addresses are normally required for a SAN Volume Controller cluster: a cluster address and a service address.
- Each SAN Volume Controller node has four fibre-channel ports, which are supplied fitted with LC-style optical small form-factor pluggable (SFP) gigabit interface converters (GBICs) for connection to a fibre-channel switch.

Each UPS requires serial cables that connect the UPS to the SAN Volume Controller nodes. Ensure that for each node, the serial and power cables come from the same UPS.

Chapter 4. Planning the physical configuration

Before you or your service representative installs the SAN Volume Controller, uninterruptible power supply (UPS) unit, and master console, you must plan the physical configuration and the initial settings for the system.

To plan the configuration, print or photocopy the hardware location chart, cable connection table and the configuration data table and use a pencil or pen to plan the system configuration. Before you begin filling in the charts and tables, make copies of the blank charts and tables so that you can later revise the configuration or create a new one if needed. Perform the following steps to prepare for your physical configuration:

1. Use the hardware location chart to record the physical configuration of your system.
2. Use the cable connection table to record how your SAN Volume Controller, the UPS unit, and the master console are to be connected.
3. Use the configuration data table to record the data that you and the service representative need before the initial installation.

When you or your service representative have completed these tasks you can perform the physical installation.

Completing the hardware location chart

The hardware location chart represents the rack into which the SAN Volume Controller is installed. Each row of the chart represents one Electrical Industries Association (EIA) 19-inch rack space.

- Uninterruptible power supply (UPS) units are heavy. Install them as near the bottom of the rack as possible. Place them within the range of row 1 through row 8.
- Do not exceed the maximum power rating of the rack and input power supply.
- Position the SAN Volume Controller such that information on the display screen is easily viewed and the controls used to navigate the display menu are easily reached. Place the SAN Volume Controller in EIA 11-38.
- To enable easy access to the connectors on the rear of the master console, position the console, keyboard and monitor unit adjacent to each other. To permit easy access to the CD drive, locate the master console above the keyboard and monitor unit. Place the master console in EIA 17-24.
- A SAN Volume Controller is one EIA unit high. Therefore, for each SAN Volume Controller that is to be installed, fill in the row that represents the position that the SAN Volume Controller is to occupy.
- The 2145 uninterruptible power supply (2145 UPS) is two EIA units high. Therefore, for each 2145 UPS, fill in two rows.
- The 2145 uninterruptible power supply-1U (2145 UPS-1U) is one EIA unit high. Therefore, for each 2145 UPS-1U, fill in one row.
- The master console is two EIA units high: one EIA unit for the server and one EIA unit for the keyboard and monitor.
- If there are any hardware devices already in the rack, record this information on the chart.

- Fill in rows for all other units that are present in the rack, including Ethernet hubs and fibre-channel switches. Hubs and switches are usually one EIA unit high, but check with your supplier. The UPS units must be installed at the bottom of the rack because it might be necessary to relocate some other devices before the SAN Volume Controller installation is started.

Hardware location guidelines

Ensure that you are familiar with the hardware location guidelines.

When you fill in the hardware location chart, follow these guidelines:

- Install the SAN Volume Controller nodes in pairs to provide redundancy and concurrent maintenance.
- A cluster can contain no more than eight SAN Volume Controller nodes.
- Connect each SAN Volume Controller node of a pair to a different uninterruptible power supply (UPS) unit.
- To reduce the chance of a simultaneous input power failure at both UPS units, connect each UPS unit to a separate electrical power source on a separate branch circuit.
- Because UPS units are heavy, install them in the lowest available positions of the rack. If necessary, move any lighter components that are already in the rack to higher positions.
- IBM does not install the Ethernet hub or the fibre-channel switches. You must arrange for either the suppliers or someone in your organization to install those items. Provide the installer with a copy of the completed hardware location chart.

In the following example, assume that the rack is empty and you want to create a system that contains the following components:

- Four SAN Volume Controller 2145-8F4 nodes, named SAN Volume Controller 2145-8F4 1, SAN Volume Controller 2145-8F4 2, SAN Volume Controller 2145-8F4 3 and SAN Volume Controller 2145-8F4 4.
- One master console.
- Four 2145 UPS-1U units named 2145 UPS-1U 1, 2145 UPS-1U 2, 2145 UPS-1U 3 and 2145 UPS-1U 4.
- One Ethernet hub named Ethernet hub 1. For this example, it is assumed that the hub is one Electrical Industries Association (EIA) unit high.
- Two fibre-channel switches named Fibre-channel switch 1 and Fibre-channel switch 2. In this example, each switch is one EIA unit high.
- RAID controllers named RAID controller 1, RAID controller 2, RAID controller 3, RAID controller 4.

Your completed chart might look like Table 6:

Table 6. Sample of completed hardware location chart

Rack row	Component
EIA 36	Blank
EIA 35	Ethernet Hub 1
EIA 34	Blank
EIA 33	Blank
EIA 32	Blank

Table 6. Sample of completed hardware location chart (continued)

Rack row	Component
EIA 31	Fibre-channel switch 1
EIA 30	Fibre-channel switch 2
EIA 29	Blank
EIA 28	SAN Volume Controller 2145-8F4 4
EIA 27	SAN Volume Controller 2145-8F4 3
EIA 26	SAN Volume Controller 2145-8F4 2
EIA 25	SAN Volume Controller 2145-8F4 1 (See note.)
EIA 24	Master console
EIA 23	Master console keyboard and monitor
EIA 22	RAID controller 4
EIA 21	
EIA 20	
EIA 19	RAID controller 3
EIA 18	
EIA 17	
EIA 16	RAID controller 2
EIA 15	
EIA 14	
EIA 13	RAID controller 1
EIA 12	
EIA 11	
EIA 10	
EIA 9	
EIA 8	
EIA 7	
EIA 6	
EIA 5	
EIA 4	2145 UPS-1U 4
EIA 3	2145 UPS-1U 3
EIA 2	2145 UPS-1U 2
EIA 1	2145 UPS-1U 1
Note: Because the SAN Volume Controller nodes contain a user panel with a display, ensure that the SAN Volume Controller nodes are located near the middle of the range.	

You might want to put the switches between the SAN Volume Controller nodes. The UPS units must remain in the lowest position of the rack.

Hardware location chart

The hardware location chart helps you to plan the location of the hardware.

Each row of the chart in Table 7 represents one Electrical Industries Association (EIA) unit.

Table 7. Hardware location chart

Rack row	Component
EIA 36	
EIA 35	
EIA 34	
EIA 33	
EIA 32	
EIA 31	
EIA 30	
EIA 29	
EIA 28	
EIA 27	
EIA 26	
EIA 25	
EIA 24	
EIA 23	
EIA 22	
EIA 21	
EIA 20	
EIA 19	
EIA 18	
EIA 17	
EIA 16	
EIA 15	
EIA 14	
EIA 13	
EIA 12	
EIA 11	
EIA 10	
EIA 9	
EIA 8	
EIA 7	
EIA 6	
EIA 5	
EIA 4	
EIA 3	
EIA 2	
EIA 1	

Completing the cable connection table

The cable connection table helps you connect the units that will be placed in the rack.

The following terms and descriptions assist you in completing the cable connection table:

Term	Description
Uninterruptible power supply (UPS)	The UPS to which the SAN Volume Controller is connected.
Ethernet	The Ethernet hub or switch to which the SAN Volume Controller is connected.
Fibre-channel ports 1 through 4	The fibre-channel switch ports to which the four SAN Volume Controller fibre-channel ports are connected. When viewed from the back of the SAN Volume Controller, the ports are numbered 1 through 4, from left to right. Ignore the markings on the back of the SAN Volume Controller.

For the master console, use the following terms and descriptions to complete the cable connection table:

Term	Description
Ethernet Port 1	Ethernet port 1 is used for your VPN connection. This port is required if you configure your master console to enable remote support. A remote support connection can only be enabled when this port has access to an external internet connection. For added security, you can disconnect this port when a remote support connection is not being used.
Ethernet Port 2	Ethernet port 2 is used to connect the SAN Volume Controller to the network.

Cable connection table

The cable connection table helps plan the placement of units in the rack.

You must complete all columns in Table 8.

Table 8. Cable connection table

SAN Volume Controller 2145-8F4	2145 UPS-1U	Ethernet hub or switch	Fibre-channel port-1	Fibre-channel port-2	Fibre-channel port-3	Fibre-channel port-4
			Switch speed	Switch speed	Switch speed	Switch speed
			Switch speed	Switch speed	Switch speed	Switch speed
			Switch speed	Switch speed	Switch speed	Switch speed
			Switch speed	Switch speed	Switch speed	Switch speed

Note: If you want to configure the SAN Volume Controller 2145-8F4 in the following manner, the master console cannot view the SAN topology because there are only two fibre-channel connections available. You can install the IBM TotalStorage Productivity Center for Fabric (TPC for Fabric) agent on at least one host system in the SAN to enable topology viewing.

Master console	**Ethernet**	
	Public network	**VPN**

Example of a completed cable connection table

Assume that you are completing the cabling details for the system.

Note: Remember that SAN Volume Controller nodes are configured in pairs and that each pair must *not* be connected to the same UPS. If you are configuring a pair of SAN Volume Controller 2145-8F2 nodes or a pair of SAN Volume Controller 2145-8F4 nodes, then you must install a pair of 2145 UPS-1U units. If you are configuring a pair of SAN Volume Controller 2145-4F2 nodes, then you must install either a pair of 2145 UPS-1U units or a pair of 2145 UPS units. 2145 UPS does not support SAN Volume Controller 2145-8F2 node or the SAN Volume Controller 2145-8F4 node.

To reduce the chance of input power failure at both UPS units, the two UPS units of a pair should not be connected to the same power source.

Assume that the pairs of 2145 UPS-1Us are node 1 and node 2, and node 3 and node 4, and that the four power sources provided by the 2145 UPS-1U units are A, B, C and D.

Note: Ensure that your 2145 UPS-1U complies with the following specifications:

- The voltage supplied to the 2145 UPS-1U must be 200 – 240 V single phase.
- The frequency supplied must be 50 or 60 Hz.

For the Ethernet connection, you must use Ethernet port 1 of the SAN Volume Controller node. *Do not* use any other Ethernet port, because the software is configured for Ethernet port 1 only.

Note: Connect all SAN Volume Controller nodes that are part of the same cluster to the same Ethernet subnet. Otherwise, TCP/IP address failover does not work.

Table 9 on page 45 illustrates this example:

Table 9. Example of cable connection table

SAN Volume Controller 2145-8F4	2145 UPS-1U	Ethernet hub or switch	Fibre-channel port-1	Fibre-chanel port-2	Fibre-channel port-3	Fibre-channel port-4
Node 1	2145 UPS-1U A	Hub or switch 1, Port 1	Fibre-channel switch 1, Port 1	Fibre-channel switch 2, Port 1	Fibre-channel switch 1, Port 2	Fibre-channel switch 2, Port 2
			Switch speed - 2 Gbps	Switch speed - 4 Gbps	Switch speed - 1 Gbps	Switch speed - 4 Gbps
Node 2	2145 UPS-1U B	Hub or switch 1, Port 2	Fibre-channel switch 1, Port 3	Fibre-channel switch 2, Port 3	Fibre-channel switch 1, Port 4	Fibre-channel switch 2, Port 4
			Switch speed - 4 Gbps	Switch speed - 4 Gbps	Switch speed - 4 Gbps	Switch speed - 4 Gbps
Node 3	2145 UPS-1U C	Hub or switch 1, Port 3	Fibre-channel switch 1, Port 5	Fibre-channel switch 2, Port 5	Fibre-channel switch 1, Port 6	Fibre-channel switch 2, Port 6
			Switch speed - 4 Gbps	Switch speed - 2 Gbps	Switch speed - 2 Gbps	Switch speed - 2 Gbps
Node 4	2145 UPS-1U D	Hub or switch 1, Port 4	Fibre-channel switch 1, Port 7	Fibre-channel switch 2, Port 7	Fibre-channel switch 1, Port 8	Fibre-channel switch 2, Port 8
			Switch speed - 2 Gbps	Switch speed - 2 Gbps	Switch speed - 4 Gbps	Switch speed - 4 Gbps

Master console	Ethernet	
	Public network	VPN
Master console	Ethernet hub 1, Port 5	Ethernet hub 1, Port 6

Completing the configuration data table

The configuration data table helps you to plan the initial settings for the cluster configuration. You must fill out the configuration data table.

Include the following initial settings for the cluster:

Term	Description
Language	The national language in which you want the messages displayed on the front panel. This option applies only to service messages. The default setting is English.
Cluster IP address	The address that is used for all typical configuration and service access to the cluster.
Service IP address	The address that is used for emergency access to the cluster.

Term	Description
Gateway IP address	The IP address for the default local gateway for the cluster.
Subnet mask	The subnet mask of the cluster.

Include the following information for the master console:

Term	Description
Machine name	The name that you want to call the master console. This must be a fully qualified DNS name. The default setting is *mannode* (not fully qualified).
Master console IP addresses	The addresses that are used for access to the master console. The default settings are: Port 1 = 192.168.1.11 Port 2 = 192.168.1.12
Master console gateway IP address	The IP address for the local gateway for the master console. The default setting is 192.168.1.1.
Master console subnet mask	The default subnet mask for the master console is 255.255.255.0.

Configuration data table

Use the configuration data table to plan the initial settings for the cluster configuration. You must fill out the configuration data table.

Cluster		
Language		
Cluster IP address		
Service IP address		
Gateway IP address		
Subnet mask		
Master console		
Machine name		
	Ethernet port 1	**Ethernet port 2**
Master console IP address		
Master console gateway IP address		
Master console subnet mask		

Chapter 5. Preparing to use the SAN Volume Controller in a SAN environment

Ensure that you perform the required preparation steps to ensure proper use of the SAN Volume Controller in a SAN environment.

Follow these preparation steps to set up your SAN Volume Controller environment:

1. Plan your configuration.
2. Plan your SAN environment.
3. Plan your fabric setup.
4. Create the RAID resources that you intend to virtualize.
5. Determine if you have a RAID that contains data that you want to merge into the cluster.
6. Determine if you will migrate data into the cluster or keep them as image-mode virtual disks (VDisks).
7. Determine if you will use Copy Services. Copy Services are provided for all supported hosts that are connected to the SAN Volume Controller.

   Related concepts

   "Image mode virtual disk migration" on page 64
   Image mode virtual disks (VDisks) have a special property that allows the last extent in the VDisk to be a partial extent.

Preparing to install the SAN Volume Controller into an existing SAN environment

Ensure your environment meets the necessary requirements to support the use of a SAN Volume Controller.

To install a SAN Volume Controller into an existing SAN that will be in use during installation, you must first ensure that the switch zoning is set to isolate the new SAN Volume Controller connections from the active part of the SAN.

See the following Web site for specific firmware levels and the latest supported hardware:

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

- Consider the design of the SAN according to your requirement for high availability.
- Identify the operating system for each host system that will be connected to the SAN Volume Controller, ensuring compatibility and suitability by performing the following steps:
  1. Specify the host bus adapters (HBAs) for each host.
  2. Define the performance requirements.
  3. Determine the total storage capacity.
  4. Determine the storage capacity per host.
  5. Determine the host LUN sizes.

6. Determine the total number of ports and bandwidth that are needed between the host and the SAN Volume Controller.
7. Determine if your SAN has enough ports to connect all hosts and back-end storage.

- Ensure that the existing SAN components meet the requirements for the SAN Volume Controller by performing the following steps:
  1. Determine the host system versions.
  2. Ensure that the HBAs, switches, and controllers are at or above the minimum requirements.
  3. Identify any components that must be upgraded.

Switch zoning for the SAN Volume Controller

Ensure that you are familiar with the constraints for zoning a switch.

Overview

The number of virtual paths to each virtual disk (VDisk) is limited. The following guidelines can help you achieve the correct number of virtual paths:

- Each host (or partition of a host) can have between one and four fibre-channel ports.
- Brocade and McData switches can be configured in Vendor Interoperability Mode or in Native Mode.
- The SAN Volume Controller supports the Interoperability Modes of the Cisco MDS 9000 family of switch and director products with the following restrictions:
  - The Cisco MDS 9000 must be connected to Brocade and McData switch/director products with the multivendor fabric zones connected using MDS Interoperability Mode 1, 2 or 3.
  - All of the SAN Volume Controller nodes that are in the SAN Volume Controller cluster must be attached to the Cisco part of the counterpart fabric or they must be attached to the McData or Brocade part of the counterpart fabric to avoid having a single fabric with a SAN Volume Controller cluster that has part of the SAN Volume Controller nodes connected to Cisco switch ports and part of the SAN Volume Controller nodes connected to Brocade or McData switch ports.
- The fabric uses the following default timeout values:
  - E_A_TOV=10 seconds
  - E_D_TOV=2 seconds

  Operation with values other than these default timeout values is not supported.

You must manually set the domain IDs prior to building the multiswitch fabric and prior to zoning for the following reasons:

- When two switches are joined while they are active, they can determine if the domain ID is already in use. If there is a conflict, the domain ID cannot be changed in an active switch. This conflict causes the fabric merging process to fail.
- The domain ID identifies switch ports when zoning is implemented using the domain and switch port number. If domain IDs are negotiated at every fabric start up, there is no guarantee that the same switch will have the same ID the next time. Therefore, zoning definitions can become invalid.

- If the domain ID is changed after a SAN is set up, some host systems might have difficulty logging back in with the switch, and it might be necessary to reconfigure the host in order to detect devices on the switch again.

The maximum number of paths from the nodes to a host is eight. The maximum number of host bus adapter (HBA) ports is four.

Example 1

Consider the SAN environment in the following example:

- Two nodes (nodes A and B)
- Nodes A and B each have four ports
  1. Node A has ports A0, A1, A2, and A3
  2. Node B has ports B0, B1, B2, and B3
- Four hosts called P, Q, R, and S
- Each of the four hosts has four ports, as described in Table 10.

Table 10. Four hosts and their ports

P	**Q**	**R**	**S**
P0	Q0	R0	S0
P1	Q1	R1	S1
P2	Q2	R2	S2
P3	Q3	R3	S3

- Two switches called X and Y
- One storage controller
- The storage controller has four ports on it called I0, I1, I2, and I3

The following is an example configuration:

```
1.  Attach ports 1 (A0, B0, P0, Q0, R0, and S0) and 2 (A1, B1, P1, Q1, R1, and S1)
    of each node and host to switch X.
2.  Attach ports 3 (A2, B2, P2, Q2, R2, and S2) and 4 (A3, B3, P3, Q3, R3, and S3)
    of each node and host to switch Y.
3.  Attach ports 1 and 2 (I0 and I1) of the storage controller to switch X.
4.  Attach ports 3 and 4 (I2 and I3) of the storage controller to switch Y.
```

Create the following host zones on switch X:

```
5.  Create a host zone containing ports 1 (A0, B0, P0, Q0, R0, and S0)
    of each node and host.
6.  Create a host zone containing ports 2 (A1, B1, P1, Q1, R1, and S1)
    of each node and host.
```

Create the following host zones on switch Y:

```
7.  Create a host zone on switch Y containing ports 3
    (A2, B2, P2, Q2, R2, and S2) of each node and host.
8.  Create a host zone on switch Y containing ports 4
    (A3, B3, P3, Q3, R3, and S3) of each node and host.
```

Create the following storage zone:

```
9. Create a storage zone that is configured on each switch.
Each storage zone contains all the SAN Volume Controller and
storage ports on that switch.
```

Example 2

The following example describes a SAN environment that is similar to the previous example except for the addition of two hosts that have two ports each.

- Two nodes called A and B
- Nodes A and B have four ports each
  1. Node A has ports A0, A1, A2, and A3
  2. Node B has ports B0, B1, B2, and B3
- Six hosts called P, Q, R, S, T and U
- Four hosts have four ports each and the other two hosts have two ports each as described in Table 11.

Table 11. Six hosts and their ports

P	**Q**	**R**	**S**	**T**	**U**
P0	Q0	R0	S0	T0	U0
P1	Q1	R1	S1	T1	U1
P2	Q2	R2	S2	—	—
P3	Q3	R3	S3	—	—

- Two switches called X and Y
- One storage controller
- The storage controller has four ports on it called I0, I1, I2, and I3

The following is an example configuration:

```
1.  Attach ports 1 (A0, B0, P0, Q0, R0, S0 and T0) and 2
    (A1, B1, P1, Q1, R1, S1 and T0) of each node and host to switch X.
2.  Attach ports 3 (A2, B2, P2, Q2, R2, S2 and T1) and 4
    (A3, B3, P3, Q3, R3, S3 and T1) of each node and host to switch Y.
3.  Attach ports 1 and 2 (I0 and I1) of the storage controller to switch X.
4.  Attach ports 3 and 4 (I2 and I3) of the storage controller to switch Y.
```

Attention: Hosts T and U (T0 and U0) and (T1 and U1) are zoned to different SAN Volume Controller ports so that each SAN Volume Controller port is zoned to the same number of host ports.

Create the following host zones on switch X:

```
5.  Create a host zone containing ports 1 (A0, B0, P0, Q0, R0, S0 and T0) of each
    node and host.
6. Create a host zone containing ports 2 (A1, B1, P1, Q1, R1, S1 and U0) of each
    node and host.
```

Create the following host zones on switch Y:

```
7.  Create a host zone on switch Y containing ports 3
    (A2, B2, P2, Q2, R2, S2 and T1) of each node and host.
8.  Create a host zone on switch Y containing ports 4
    (A3, B3, P3, Q3, R3, S3 and U1) of each node and host.
```

Create the following storage zone:

```
9.  Create a storage zone configured on each switch.
Each storage zone contains all the SAN Volume Controller
and storage ports on that switch.
```

Related reference

"Fibre-channel switches" on page 82
Ensure that you are familiar with the configuration rules for fibre-channel

switches. You must follow the configuration rules for fibre-channel switches to ensure that you have a valid configuration.

Zoning guidelines

Ensure that you are familiar with the following zoning guidelines.

Paths to hosts

- The number of paths through the network from the SAN Volume Controller nodes to a host must not exceed 8. Configurations in which this number is exceeded are not supported.
  - Each node has four ports and each I/O group has two nodes. Therefore, without any zoning, the number of paths to a VDisk would be 8 × the number of host ports.
  - This rule exists to limit the number of paths that must be resolved by the multipathing device driver.

If you want to restrict the number of paths to a host, zone the switches so that each HBA port is zoned with one SAN Volume Controller port for each node in the cluster. If a host has multiple HBA ports, zone each port to a different set of SAN Volume Controller ports to maximize performance and redundancy.

Controller zones

Switch zones that contain controller ports must not have more than 40 ports. A configuration that exceeds 40 ports is not supported.

SAN Volume Controller zones

The switch fabric must be zoned so that the SAN Volume Controller nodes can see the back-end storage and the front end host HBAs. Usually, the front-end host HBAs and the back-end storage are not in the same zone. The exception to this is where split host and split controller configuration is in use.

It is possible to zone the switches in such a way that a SAN Volume Controller port is used solely for internode communication, or for communication to host, or for communication to back-end storage. This is possible because each node contains 4 ports. Each node must still remain connected to the full SAN fabric. Zoning cannot be used to separate the SAN into two parts.

With Metro Mirror configurations, additional zones are required that contain only the local nodes and the remote nodes. It is valid for the local hosts to see the remote nodes, or for the remote hosts to see the local nodes. Any zone that contains the local and the remote back-end storage and local nodes or remote nodes, or both, is not valid.

If a node can see another node through multiple paths, use zoning where possible to ensure that the SAN Volume Controller to SAN Volume Controller communication does not travel over an ISL. If a node can see a storage controller through multiple paths, use zoning to restrict communication to those paths that do not travel over ISLs.

The SAN Volume Controller zones must ensure that every port on each node can see at least one port that belongs to every other node in the cluster.

The SAN Volume Controller zones must ensure that the nodes in the local cluster can only see nodes that are in the remote cluster. You can have one or two nodes that are not members of any cluster zoned to see all of the clusters. This allows you to use the command-line interface (CLI) to add the node to the cluster in the event that you must replace a node.

Host zones

The configuration rules for host zones are different depending upon the number of hosts that will access the cluster. For smaller configurations of less than 64 hosts per cluster, the SAN Volume Controller supports a simple set of zoning rules which enable a small set of host zones to be created for different environments. For larger configurations of more than 64 hosts, the SAN Volume Controller supports a more restrictive set of host zoning rules.

Zoning that contains host HBAs must not contain either host HBAs in dissimilar hosts or dissimilar HBAs in the same host that are in separate zones. Dissimilar hosts means that the hosts are running different operating systems or are different hardware platforms; thus different levels of the same operating system are regarded as similar.

Clusters with less than 64 hosts

For clusters with less than 64 hosts attached, zones that contain host HBAs must contain no more than 40 initiators including the SAN Volume Controller ports that act as initiators. A configuration that exceeds 40 initiators is not supported. A valid zone can be 32 host ports plus 8 SAN Volume Controller ports. You *should* place each HBA port in a host that connects to a node into a separate zone. You *should* also include exactly one port from each node in the I/O groups which are associated with this host. This type of host zoning is not mandatory, but is preferred for smaller configurations.

Note: If the switch vendor recommends fewer ports per zone for a particular SAN, the more strict rules that are imposed by the fibre-channel vendor takes precedence over the SAN Volume Controller rules.

Clusters with 64 to 256 hosts

For clusters with 64 to 256 hosts attached, each HBA port in a host that connects to a node *must* be placed into a separate zone. In this separate zone, you must also include exactly one port from each node in the I/O groups that are associated with this host.

The SAN Volume Controller does not specify the number of host fibre-channel ports or HBAs that a host or a partition of a host can have. The number of host fibre-channel ports or HBAs are specified by the host multipathing device driver. The SAN Volume Controller supports this number; however it is subject to the other configuration rules that are specified here.

To obtain the best performance from a host with multiple fibre-channel ports, the zoning must ensure that each fibre-channel port of a host is zoned with a different group of SAN Volume Controller ports.

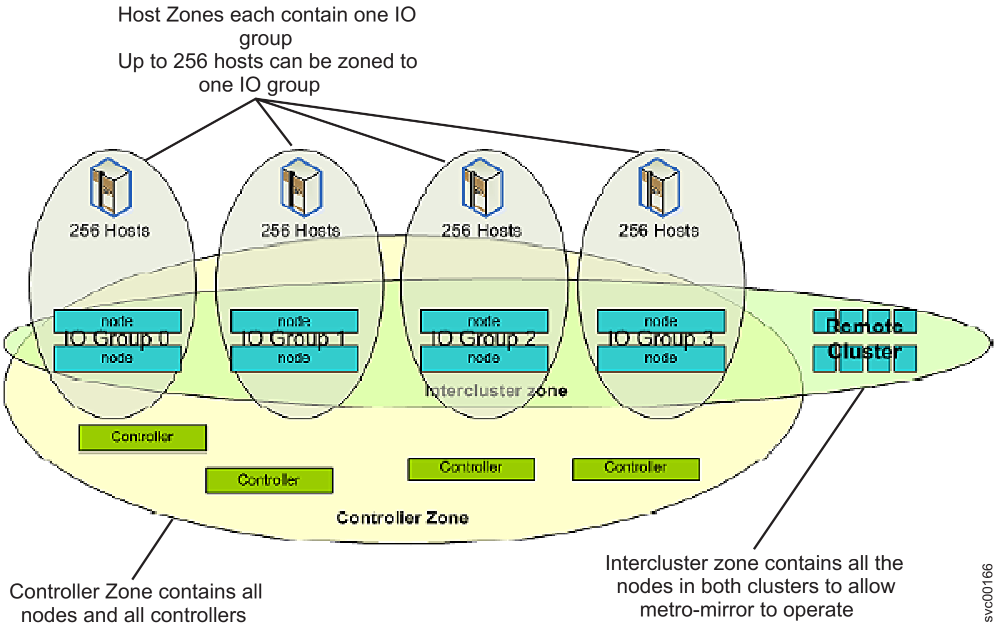
To obtain the best overall performance of the subsystem and to prevent overloading, the workload to each SAN Volume Controller port must be equal.

This can typically involve zoning approximately the same number of host fibre-channel ports to each SAN Volume Controller fibre-channel port.

Clusters with 256 to 1024 hosts

For clusters with 256 to 1024 hosts attached, the SAN must be zoned so that each HBA port in a host that connects to a node can only see one SAN Volume Controller port for each node in the I/O group that is associated with the host. If you have 1024 hosts, each host must be associated with only one I/O group and each I/O group must only be associated with up to 256 hosts.

Figure 14 provides an example configuration for zoning 1024 hosts. In this example, the hosts are arranged into four groups of 256 hosts and each group of 256 hosts is zoned to one I/O group. You must zone each group of 256 hosts separately so they cannot see other hosts that are in different I/O groups. The controller zone contains all eight of the nodes and all four of the controllers. The intercluster zone contains all of the nodes that are in both clusters to allow you to use Metro Mirror.


Host Zones each contain one IO group
Up to 256 hosts can be zoned to one IO group

Controller Zone contains all nodes and all controllers

Intercluster zone contains all the nodes in both clusters to allow metro-mirror to operate

Figure 14. Zoning a 1024 host configuration

You can have up to 512 fibre-channel logins per fibre-channel port. The following logins are counted towards the 512 login maximum:

- Host port logins
- Storage controller logins
- SAN Volume Controller node logins
- Fibre-channel name server logins

If any port has more than 512 logins, the node logs an `ID 073006` error. You can use the **svcinfo lsfabric** command-line interface (CLI) command to list the logins that are seen by each SAN Volume Controller port.

Zoning considerations for Metro Mirror

Ensure that you are familiar with the constraints for zoning a switch to support the Metro Mirror service.

SAN configurations that use the Metro Mirror feature between two clusters require the following additional switch zoning considerations:

- Additional zones for Metro Mirror. For Metro Mirror operations involving two clusters, these clusters must be zoned so that the nodes in each cluster can see the ports of the nodes in the other cluster.
- Use of extended fabric settings in a switched fabric.
- Use of interswitch link (ISL) trunking in a switched fabric.
- Use of redundant fabrics.

Note: These considerations do not apply if the simpler intracluster mode of Metro Mirror operation is in use and only a single cluster is needed.

For intracluster Metro Mirror relationships, no additional switch zones are required. For intercluster Metro Mirror relationships, you must perform the following steps:

1. Form a SAN that contains both clusters that are to be used in the Metro Mirror relationships. If cluster A originally is in SAN A and cluster B is originally in SAN B, there must be at least one fibre-channel connection between SAN A and SAN B. This connection consists of one or more interswitch links. The fibre-channel switch ports that are associated with these interswitch ports must not appear in any zone.
2. Ensure that each switch has a different domain ID before you connect the two SANs. Form a single SAN from the combination of SAN A and SAN B prior to the connection of the two SANs.
3. After the switches in SAN A and SAN B are connected, configure the switches so that they operate as a single group of switches. Each cluster must retain the same set of zones that were required to operate in the original single SAN configuration.

   Note: You do not have to perform this step if you are using routing technologies to connect the two SANs.
4. Add a new zone that contains all the switch ports that are connected to SAN Volume Controller ports. This zone contains switch ports that were originally in SAN A and in SAN B.
5. This step is optional because in some cases, this view of both clusters can complicate the way that you operate the overall system. Therefore, unless it is specifically needed, avoid implementing this view. Adjust the switch zoning so that the hosts that were originally in SAN A can recognize cluster B. This allows a host to examine data in both the local and remote cluster, if required.
6. Verify that the switch zoning is such that cluster A cannot recognize any of the back-end storage that is owned by cluster B. Two clusters cannot share the same back-end storage devices.

The following zones are needed in a typical intercluster Metro Mirror configuration:

- A zone in the local cluster that contains all the ports in the SAN Volume Controller nodes in that local cluster and the ports on the back-end storage that are associated with that local cluster. These zones are required whether Metro Mirror is in use.

- A zone in the remote cluster that contains all the ports in the SAN Volume Controller nodes in that remote cluster and the ports on the back-end storage that are associated with that remote cluster. These zones are required even if Metro Mirror is not in use.
- A zone that contains all the ports in the SAN Volume Controller nodes in both the local and remote cluster. This zone is required for intercluster communication and is specifically required by Metro Mirror operations.
- Additional zones that contain ports in host HBAs and selected ports on the SAN Volume Controller nodes in a particular cluster. These are the zones that allow a host to recognize VDisks that are presented by an I/O group in a particular cluster. These zones are required even if Metro Mirror is not in use.

Note:

1. While it is normal to zone a server connection so that it is only visible to the local or remote cluster, it is also possible to zone the server so that the host HBA can see nodes in both the local and remote cluster at the same time.
2. Intracluster Metro Mirror operation does not require any additional zones, over those that are needed to run the cluster itself.

Switch operations over long distances

Some SAN switch products provide features that allow the users to tune the performance of I/O traffic in the fabric in a way that can affect Metro Mirror performance. The two most significant features are ISL trunking and extended fabric.

The following table provides a description of the ISL trunking and the extended fabric features:

Feature	Description
ISL trunking	Trunking enables the switch to use two links in parallel and still maintain frame ordering. It does this by routing all traffic for a given destination over the same route even when there might be more than one route available. Often trunking is limited to certain ports or port groups within a switch. For example, in the IBM 2109-F16 switch, trunking can only be enabled between ports in the same quad (for example, same group of four ports). For more information on trunking with the MDS, refer to ″Configuring Trunking″ on the Cisco Systems Web site. Some switch types can impose limitations on concurrent use of trunking and extended fabric operation. For example, with the IBM 2109-F16 switch, it is not possible to enable extended fabric for two ports in the same quad. Thus, extended fabric and trunking cannot be used together. Although it is possible to enable extended fabric operation one link of a trunked pair, this does not offer any performance advantages and adds complexity to the configuration setup. Therefore, do not use mixed mode operations.

Feature	Description
Extended fabric	Extended fabric operation allocates extra buffer credits to a port. This is important over long links that are usually found in intercluster Metro Mirror operation. Because of the time that it takes for a frame to traverse the link, it is possible to have more frames in transmission at any instant in time than is possible over a short link. The additional buffering is required to allow for the extra frames. For example, the default license for the IBM 2109-F16 switch has two extended fabric options: Normal and Extended Normal. • The Normal option is suitable for short links. • The Extended Normal option provides significantly better performance for the links up to 10 km long. **Note:** With the additional extended fabric license, the user has two extra options: Medium, up to 10 - 50 km and Long, 50 - 100 km. Do not use Medium and Long settings in the intercluster Metro Mirror links that are currently supported.

Cluster configuration using SAN fabrics with long distance fibre links

A SAN Volume Controller cluster using SAN fabric switches can connect to application hosts, storage controllers, or other SAN Volume Controller clusters, through the use of short or long wave optical fibre-channel connections.

The maximum distance between the cluster and host or the cluster and the storage controller is 300 m for short wave and 10 km for long wave optical connections. Longer distances are supported between clusters that use intercluster Metro Mirror.

Follow these guidelines when you use long wave optical fibre-channel connections:

- For disaster recovery, each cluster must be regarded as a single entity. This includes the back-end storage that provides the quorum disks for the cluster. Therefore, the cluster and quorum disks must be co-located. Do not place components of a single cluster in different physical locations.
- All nodes within a cluster must be located in the same set of racks. There can be a large optical distance between the nodes in the same cluster; however, the nodes must be physically co-located to permit effective service and maintenance.
- All nodes in a cluster must be on the same IP subnet. This allows the nodes to assume the same cluster or service IP address.
- A node must be on the same rack as the uninterruptible power supply from which it receives power.

Note: Do not split cluster operation over a long optical distance, otherwise you will only be able to use asymmetric disaster recovery and it will have substantially reduced performance. Instead, use two cluster configurations for all production disaster recovery systems.

Related reference

"Switch operations over long distances" on page 55
Some SAN switch products provide features that allow the users to tune the performance of I/O traffic in the fabric in a way that can affect Metro Mirror performance. The two most significant features are ISL trunking and extended fabric.

"Zoning considerations for Metro Mirror" on page 54
Ensure that you are familiar with the constraints for zoning a switch to support the Metro Mirror service.

Performance of fibre-channel extenders

When you are planning to use fibre-channel extenders, be aware that the performance of the link to the remote location decreases as the distance to the remote location increases.

For fibre-channel IP extenders, throughput is limited by latency and bit error rates. Typical I/O latency can be expected to be 10 microseconds per kilometer. Bit error rates vary depending on the quality of the circuit that is provided.

You must review the total throughput rates that might be expected for your planned configuration with the vendor of your fibre-channel extender and your network provider.

Related reference

"Supported fibre-channel extenders" on page 90
The supported hardware for the SAN Volume Controller frequently changes.

Nodes

A SAN Volume Controller *node* is a single processing unit within a SAN Volume Controller cluster.

For redundancy, nodes are deployed in pairs to make up a cluster. A cluster can have one to four pairs of nodes. Each pair of nodes is known as an I/O group. Each node can be in *only* one I/O group. A maximum of four I/O groups each containing two nodes is supported.

At any one time, a single node in the cluster manages configuration activity. This configuration node manages a cache of the configuration information that describes the cluster configuration and provides a focal point for configuration commands. If the configuration node fails, another node in the cluster takes over its responsibilities.

Table 12 describes the operational states of a node.

Table 12. Node state

State	Description
Adding	The node was added to the cluster but is not yet synchronized with the cluster state (see Note). The node state changes to Online after synchronization is complete.
Deleting	The node is in the process of being deleted from the cluster.
Online	The node is operational, assigned to a cluster, and has access to the fibre-channel SAN fabric.
Offline	The node is not operational. The node was assigned to a cluster but is not available on the fibre-channel SAN fabric. Run the Directed Maintenance Procedures to determine the problem.

Table 12. Node state (continued)

State	Description
Pending	The node is transitioning between states and, in a few seconds, will move to one of the other states.
Note: A node can stay in the Adding state for a long time. You should wait for at least 30 minutes before taking further action, but if after 30 minutes, the node state is still Adding, you can delete the node and add it again. If the node that has been added is at a lower code level than the rest of the cluster, the node is upgraded to the cluster code level, which can take up to 20 minutes. During this time, the node is shown as adding.	

Clusters

All of your configuration and service tasks are performed at the cluster level. Therefore, after configuring your cluster, you can take advantage of the virtualization and the advanced features of the SAN Volume Controller.

A cluster can consist of two nodes, with a maximum of eight nodes. Therefore, you can assign up to eight SAN Volume Controller nodes to one cluster.

All configurations are replicated across all nodes in the cluster, however, only some service actions can be performed at the node level. Because configuration is performed at the cluster level, an IP address is assigned to the cluster instead of each node.

Cluster state

The state of the cluster holds all of the configuration and internal data.

The cluster state information is held in nonvolatile memory. If the mainline power fails, the uninterruptible power supply units maintain the internal power long enough for the cluster state information to be stored on the internal SCSI disk drive of each node. The read and write cache information, which is also held in memory, is stored on the internal SCSI disk drives of the nodes in the input/output (I/O) group that are using that information. Similarly, if the power fails to a node, configuration and cache data for that node is lost and the partner node attempts to flush the cache. The cluster state is still maintained by the other nodes on the cluster.

Figure 15 on page 59 shows an example of a cluster that contains four nodes. The cluster state shown in the grey box does not actually exist, instead each node holds a copy of the entire cluster state.

The cluster contains a single node that is elected as the configuration node. The configuration node can be thought of as the node that controls the updating of cluster state. For example, a user request is made (*1*), that results in a change being made to the configuration. The configuration node controls updates to the cluster (*2*). The configuration node then forwards the change to all nodes (including Node 1), and they all make the state-change at the same point in time (*3*). Using this state-driven model of clustering ensures that all nodes in the cluster know the exact cluster state at any one time.

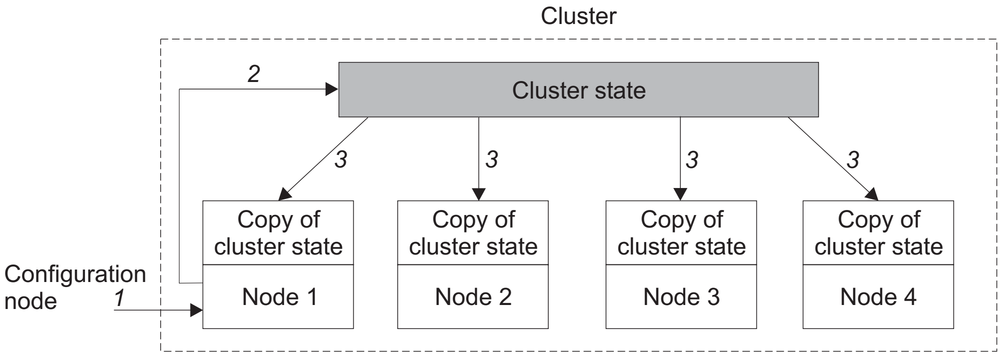


Figure 15. Cluster, nodes, and cluster state.

Each node in the cluster maintains an identical copy of the cluster state. When a change is made to the configuration or internal cluster data, then the same change is applied to all nodes. For example, a user configuration request is made to the configuration node. The configuration node forwards the request to all nodes in the cluster and they all make the change to the cluster state at the same point in time. This ensures that all nodes are aware of the configuration change. If the configuration node fails, then the cluster can elect a new node to take over its responsibilities.

Cluster operation and quorum disks

The cluster must contain at least half of its nodes to function.

Nodes are deployed in pairs known as input/output (I/O) groups, and one to four I/O groups comprise a cluster. In order to function, one node in each I/O group must be operational. If both of the nodes in an I/O group are not operational, access is lost to the virtual disks (VDisks) that are managed by the I/O group.

Note: The cluster can continue to run without loss of access to data as long as one node from each I/O group is available.

A tie-break situation can occur if exactly half the nodes in a cluster fail at the same time, or if the cluster is divided so that exactly half the nodes in the cluster cannot communicate with the other half. For example, in a cluster of four nodes, if any two nodes fail at the same time or any two cannot communicate with the other two, a tie-break exists.

The cluster automatically chooses three managed disks (MDisks) to be candidate *quorum disks* and assigns them quorum indexes of 0, 1, and 2. One of these disks is used to settle a tie-break condition.

If a tie-break occurs, the first half of the cluster to access the quorum disk after the split has occurred locks the disk and continues to operate. The other side stops. This action prevents both sides from becoming inconsistent with each other.

You can change the assignment of quorum disk at any time by issuing the following command:

```
svctask setquorum
```

I/O groups and UPS

Each pair of nodes is known as an *input/output (I/O)* group.

Each node can only be in one I/O group. The I/O groups are connected to the SAN so that all back-end storage and all application servers are visible to all of the I/O groups. Each pair of nodes has the responsibility to serve I/O operations on a particular virtual disk (VDisk).

VDisks are logical disks that are presented to the SAN by SAN Volume Controller nodes. VDisks are also associated with an I/O group. The SAN Volume Controller does not contain any internal battery backup units and therefore must be connected to an uninterruptible power supply (UPS) to provide data integrity in the event of a cluster wide power failure.

When an application server performs I/O to a VDisk, it can access the VDisk with either of the nodes in the I/O group. A VDisk can have a preferred node specified when the VDisk is created. Otherwise, if the preferred node is not specified, after the VDisk is created, a preferred node will be assigned. The preferred node is the node through which a VDisk can be accessed.

Each I/O group only has two nodes. The distributed cache inside the SAN Volume Controller is replicated across both nodes in the I/O group. When I/O is performed to a VDisk, the node that processes the I/O duplicates the data onto the partner node that is in the I/O group. I/O traffic for a particular VDisk is, at any one time, managed exclusively by the nodes in a single I/O group. Thus, although a cluster can have many nodes within it, the nodes handle I/O in independent pairs. This means that the I/O capability of the SAN Volume Controller scales well, since additional throughput can be obtained by adding additional I/O groups.

Figure 16 on page 61 shows an example I/O group. A write operation from a host is shown (item *1*), that is targeted for VDisk A. This write is targeted at the preferred node, Node 1 (item *2*). The write is cached and a copy of the data is made in the partner node, Node 2's cache (item *3*). The write is now complete so far as the host is concerned. At some time later the data is written, or de-staged, to storage (item *4*). The figure also shows two UPS units (1 and 2) correctly configured so that each node is in a different power domain.

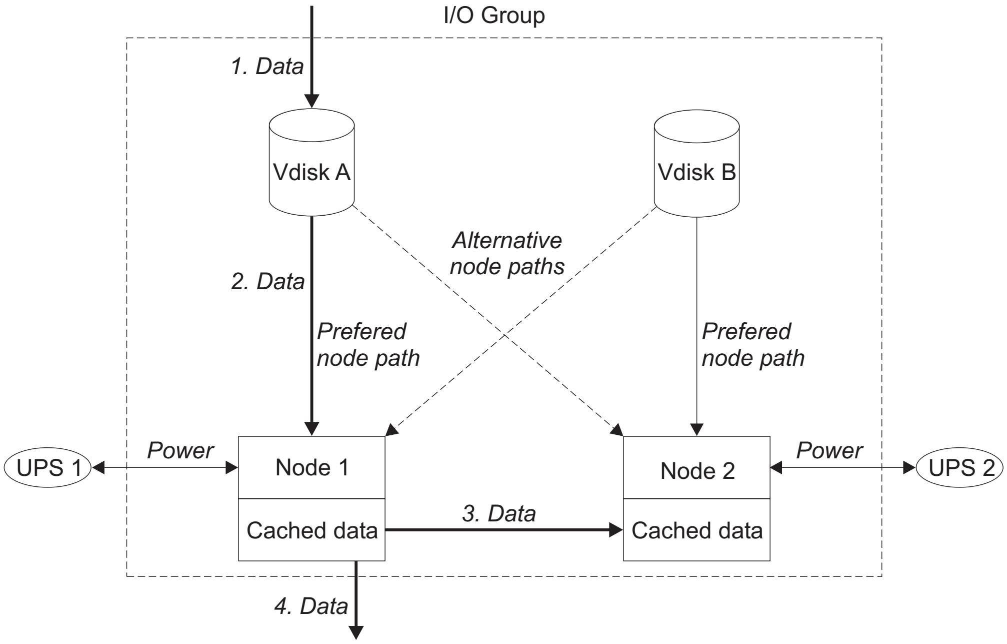


Figure 16. I/O group and UPS

When a node fails within an I/O group, the other node in the I/O group takes over the I/O responsibilities of the failed node. Data loss during a node failure is prevented by mirroring the I/O read/write data cache between the two nodes in an I/O group.

If only one node is assigned to an I/O group or if a node has failed in an I/O group, the cache goes into write-through mode. Therefore, any writes for the VDisk that are assigned to this I/O group are not cached, but are sent directly to the storage device. If both nodes in an I/O group go offline, the VDisk that are assigned to the I/O group cannot be accessed.

When a VDisk is created, you must specify the I/O group that will provide access to the VDisk. However, VDisks can be created and added to I/O groups that contain offline nodes. I/O access is not possible until at least one of the nodes in the I/O group is online.

The cluster also provides a recovery I/O group, which is used when both nodes in the I/O group have multiple failures. This allows you to move the VDisks to the recovery I/O group and then into a working I/O group. I/O access is not possible when VDisks are assigned to the recovery I/O group.

UPS and power domains

An uninterruptible power supply (UPS) protects the cluster against power failures.

If the mainline power fails to one or more nodes in the cluster, the UPS maintains the internal power long enough for the cluster state information to be stored on the internal SCSI disk drive of each node.

It is required that each node in the cluster be connected to a UPS. This allows the cluster to continue to work in degraded mode if one UPS fails.

It is very important that the two nodes in an I/O group are not both connected to the same power domain. Each node of an I/O group must be connected to a different UPS. This configuration ensures that the cache and cluster state information is protected against the failure of the UPS or of the mainline power source. If possible, each UPS should be connected to a different power source. Otherwise, a power source failure will result in the I/O group being taken offline.

The following UPS units can be used with the SAN Volume Controller:

- 2145 uninterruptible power supply-1U (2145 UPS-1U). This unit can support one node.
- 2145 uninterruptible power supply (2145 UPS). This unit can support two nodes.

Note: The 2145 UPS does *not* support the SAN Volume Controller 2145-8F2 or the SAN Volume Controller 2145-8F4 nodes.

These units can be used in combination in a cluster. Table 13 shows the required number of UPS units for the number of nodes in a cluster.

Table 13. Required UPS units

Number of nodes	**Number of required 2145 UPS units**	**Number of required 2145 UPS-1U units**
2 nodes	1 2145 UPS units	2 2145 UPS-1U unit
4 nodes	2 2145 UPS units	4 2145 UPS-1U units
6 nodes	4 2145 UPS units	6 2145 UPS-1U units
8 nodes	4 2145 UPS units	8 2145 UPS-1U units

When nodes are added to the cluster, you must specify the I/O group that they will join. The configuration interfaces will also check the UPS units and ensure that the two nodes in the I/O group are not connected to the same UPS units.

Figure 17 shows a cluster of four nodes, with two I/O groups and two UPS units.

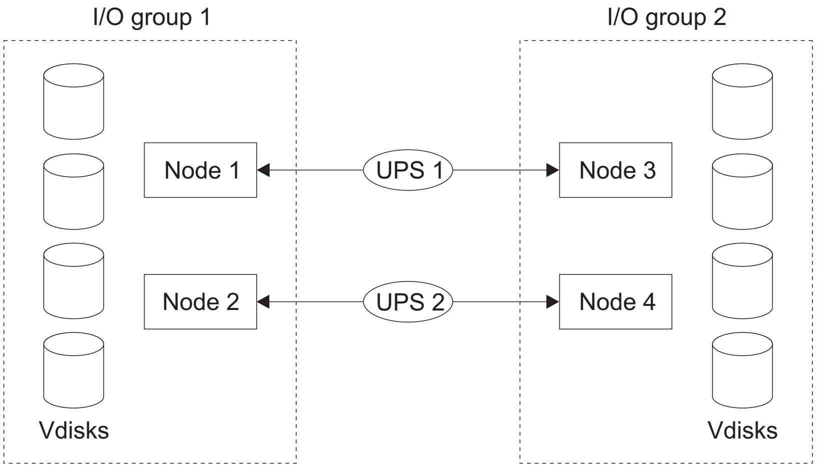


Figure 17. Relationship between I/O groups and UPS units

Attention: Do not connect two clusters to the same pair of UPS units. Both clusters will be lost in the event that a power failure occurs on both of the units.

Disk controllers

A *disk controller* is a device that coordinates and controls the operation of one or more disk drives and synchronizes the operation of the drives with the operation of the system as a whole.

Disk controllers provide the storage that the cluster detects as managed disks (MDisks).

When configuring disk controllers, ensure that you configure and manage the disk controllers and its devices for optimal performance.

The supported RAID controllers are detected by the cluster and reported by the user interfaces. The cluster can determine which MDisks each controller has and can also provide a view of MDisks that is filtered by controller. This view enables you to associate the MDisks with the RAID that the controller presents.

Note: The SAN Volume Controller supports RAID controllers, but it is possible to configure a controller as a non-RAID controller. RAID controllers provide redundancy at the disk level. Therefore, a single physical disk failure does not cause an MDisk failure, an MDisk group failure, or a failure in the virtual disks (VDisks) that were created from the MDisk group.

The controller can have a local name for the RAID or single disks that it is providing. However it is not possible for the nodes in the cluster to determine this name because the namespace is local to the controller. The controller will surface these disks with a unique ID, the controller LUN or LU number. This ID, along with the controller serial number or numbers (there can be more than one controller), can be used to associate the MDisks in the cluster with the RAID presented by the controllers.

To minimize data loss, virtualize only those RAID that provide some form of redundancy, that is, RAID 1, RAID 10, RAID 0+1 or RAID 5. Do not use RAID 0 because a single physical disk failure might cause the failure of many VDisks.

Unsupported disk controller systems (generic controllers)

When a disk controller system is detected on the SAN, the SAN Volume Controller attempts to recognize it using its inquiry data. If the disk controller system is recognized as one of the explicitly supported storage models, then the SAN Volume Controller uses error recovery programs that can be tailored to the known needs of the disk controller system. If the storage controller is not recognized, then the SAN Volume Controller configures the disk controller system as a generic controller. A generic controller might not function correctly when it is addressed by a SAN Volume Controller. The SAN Volume Controller does not regard accessing a generic controller as an error condition and, consequently, does not log an error. MDisks that are presented by generic controllers are not eligible to be used as quorum disks.

Related concepts

"MDisks" on page 18
A *managed disk (MDisk)* is a logical disk (typically a RAID or partition thereof) that a storage subsystem has exported to the SAN fabric to which the nodes in the cluster are attached.

Data migration

Data migration affects the mapping of the extents for a virtual disk (VDisk) to the extents for a managed disk (MDisk).

The host can access the VDisk during the data migration process.

Applications for data migration

The following lists the different applications for data migration:

- Redistribution of workload within a cluster across MDisks. You can redistribute the workload, by one of the following ways:
  - Moving workload onto newly installed storage
  - Moving workload from old or failing storage, prior to replacing it
  - Moving workload to rebalance workload that has changed
- Migrating data from legacy disks to disks that are managed by the SAN Volume Controller.

Image mode virtual disk migration

Image mode virtual disks (VDisks) have a special property that allows the last extent in the VDisk to be a partial extent.

You can migrate striped and sequential Vdisks to an image mode VDisk.

Copy Services

The SAN Volume Controller provides Copy Services that enable you to copy virtual disks (VDisks).

The following Copy Services options are available for all supported hosts that are connected to the SAN Volume Controller:

FlashCopy
Makes an instant, point-in-time copy from a source VDisk to a target VDisk.

Metro Mirror
Provides a consistent copy of a source VDisk on a target VDisk. Data is written to the target VDisk synchronously after it is written to the source VDisk, so the copy is continuously updated.

Global Mirror
Provides a consistent copy of a source VDisk on a target VDisk. Data is written to the target VDisk asynchronously, so the copy is continuously updated, but might not contain the last few updates in the event that a disaster recovery operation is performed.

FlashCopy

FlashCopy is a Copy Service that is available with the SAN Volume Controller.

FlashCopy copies the contents of a source virtual disk (VDisk) to a target VDisk. Any data that existed on the target disk is lost and is replaced by the copied data. After the copy operation has been completed, the target VDisks contain the contents of the source VDisks as they existed at a single point in time unless target writes have been performed. Although the copy operation takes some time to complete, the resulting data on the target is presented in such a way that the copy appears to have occurred immediately. FlashCopy is sometimes described as an instance of a time-zero copy (T 0) or point-in-time copy technology. Although the FlashCopy operation takes some time, this time is several orders of magnitude less than the time that would be required to copy the data using conventional techniques.

It is difficult to make a consistent copy of a data set that is constantly updated. Point-in-time copy techniques help solve this problem. If a copy of a data set is created using a technology that does not provide point-in-time techniques and the data set changes during the copy operation, the resulting copy might contain data which is not consistent. For example, if a reference to an object is copied earlier than the object itself and the object is moved before it is copied, the copy contains the referenced object at its new location but the copied reference still points to the old location.

Source VDisks and target VDisks must meet the following requirements:

- They must be the same size.
- The same cluster must manage them.

Related concepts

"VDisks" on page 23
A *virtual disk (VDisk)* is a logical disk that the cluster presents to the storage area network (SAN).

FlashCopy mappings

A FlashCopy mapping defines the relationship between a source virtual disk (VDisk) and a target VDisk.

Because FlashCopy copies one VDisk to another VDisk, the SAN Volume Controller must be aware of the mapping relationship. A VDisk can be the source or target of only one mapping. For example, you cannot make the target of one mapping the source of another mapping.

FlashCopy makes an instant copy of a VDisk at the time that it is started. To create a FlashCopy of a VDisk, you must first create a mapping between the source VDisk (the disk that is copied) and the target VDisk (the disk that receives the copy). The source and target must be of equal size.

To copy a VDisk, it must be part of a FlashCopy mapping or of a consistency group.

A FlashCopy mapping can be created between any two VDisks in a cluster. It is not necessary for the VDisks to be in the same I/O group or managed disk (MDisk) group. When a FlashCopy operation is started, a checkpoint is made of the source VDisk. No data is actually copied at the time a start occurs. Instead, the checkpoint creates a bitmap that indicates that no part of the source VDisk has been copied. Each bit in the bitmap represents one region of the source VDisk. Each region is called a *grain*.

After a FlashCopy operation starts, read operations to the source VDisk continue to occur. If new data is written to the source or target VDisk, the existing data on the source is copied to the target VDisk before the new data is written to the source or target VDisk. The bitmap is updated to mark that the grain of the source VDisk has been copied so that later write operations to the same grain do not recopy the data.

During a read operation to the target VDisk, the bitmap is used to determine if the grain has been copied. If the grain has been copied, the data is read from the target VDisk. If the grain has not been copied, the data is read from the source VDisk.

When you create a mapping, you specify the background copy rate. The background copy rate determines the priority that is given to the background copy process. If you want to end with a copy of the whole source at the target (so that the mapping can be deleted, but the copy can still be accessed at the target), you must copy all the data that is on the source VDisk to the target VDisk.

When a mapping is started and the background copy rate is greater than zero (or a value other than NOCOPY), the unchanged data is copied to the target, and the bitmap is updated to show that the copy has occurred. After a time, the length of which depends on the priority given and the size of the VDisk, the whole VDisk is copied to the target. The mapping returns to the idle/copied state. You can restart the mapping at any time to create a new copy at the target.

If the background copy rate is zero (or NOCOPY), only the data that changes on the source is copied to the target. The target never contains a copy of the whole source unless every extent is overwritten at the source. You can use this copy rate when you require a temporary copy of the source.

You can stop the mapping at any time after it has been started. This action makes the target inconsistent and the target VDisk is taken offline. You must restart the mapping to correct the target.

FlashCopy mapping states

At any point in time, a FlashCopy mapping is in one of the following states:

Idle or copied
The source and target VDisks act as independent VDisks even if a FlashCopy mapping exists between the two. Read and write caching is enabled for both the source and the target.

Copying
The copy is in progress.

Prepared
The mapping is ready to start. The target VDisk is online, but not accessible. The target VDisk cannot perform read or write caching. Read and write caching is failed by the SCSI front-end as a hardware error.

Preparing
The target VDisk is online, but not accessible. The target VDisk cannot perform read or write caching. Read and write caching is failed by the SCSI front-end as a hardware error. Any changed write data for the source VDisk is flushed from the cache. Any read or write data for the target VDisk is discarded from the cache.

Stopped
The mapping is stopped because either you issued a command or an I/O error occurred. Preparing and starting the mapping again can restart the copy.

Suspended

The mapping started, but it did not complete. The source VDisk might be unavailable, or the copy bitmap might be offline. If the mapping does not return to the copying state, stop the mapping to reset the mapping.

Before you start the mapping, you must prepare it. Preparing the mapping ensures that the data in the cache is de-staged to disk and a consistent copy of the source exists on disk. At this time, the cache goes into write-through mode. Data that is written to the source is not cached in the SAN Volume Controller nodes; it passes straight through to the MDisks. The prepare operation for the mapping might take you a few minutes; the actual length of time depends on the size of the source VDisk. You must coordinate the prepare operation with the operating system. Depending on the type of data that is on the source VDisk, the operating system or application software might also cache data write operations. You must flush, or synchronize, the file system and application program before you prepare for and start the mapping.

Note: The **svctask startfcmap** command can take some time to process.

If you do not want to use consistency groups, the SAN Volume Controller allows a FlashCopy mapping to be treated as an independent entity. In this case the FlashCopy mapping is known as a stand-alone mapping. For FlashCopy mappings which have been configured in this way, the **svctask prestartfcmap** and **svctask startfcmap** commands are directed at the FlashCopy mapping name rather than the consistency group ID.

Veritas Volume Manager

For FlashCopy target VDisks, the SAN Volume Controller sets a bit in the inquiry data for those mapping states where the target VDisk could be an exact image of the source VDisk. Setting this bit enables the Veritas Volume Manager to distinguish between the source and target VDisks and provide independent access to both.

Related concepts

"VDisks" on page 23
A *virtual disk (VDisk)* is a logical disk that the cluster presents to the storage area network (SAN).

FlashCopy consistency groups

A *consistency group* is a container for mappings. You can add many mappings to a consistency group.

The consistency group is specified when the mapping is created. You can also change the consistency group later. When you use a consistency group, you prepare and trigger that group instead of the various mappings. This ensures that a consistent copy is made of all the source virtual disks (VDisks). Mappings that you want to control at an individual level are known as stand alone mappings. Stand alone mappings should not be placed into a consistency group, otherwise they are controlled as part of that consistency group.

To copy a VDisk, it must be part of a FlashCopy mapping or of a consistency group.

When you copy data from one VDisk to another, the data might not include all that you need to enable you to use the copy. Many applications have data that spans multiple VDisks and requires that data integrity is preserved across VDisks. For example, the logs for a particular database usually reside on a different VDisk than the VDisk that contains the data.

Consistency groups address the problem when applications have related data that spans multiple VDisks. In this situation, FlashCopy must be performed in a way that preserves data integrity across the multiple VDisks. One requirement for preserving the integrity of data being written is to ensure that dependent writes are run in the intended sequence of the application.

FlashCopy consistency group states

At any point in time, a FlashCopy consistency group is in one of the following states:

Idle or copied
The source and target VDisks act independently even if a FlashCopy consistency group exists. Read and write caching is enabled for the source VDisks and target VDisks.

Copying
The copy is in progress.

Prepared
The consistency group is ready to start. While in this state, the target VDisks are offline.

Preparing
Any changed write data for the source VDisks is flushed from the cache. Any read or write data for the target VDisks is discarded from the cache.

Stopped
The consistency group is stopped because either you issued a command or an I/O error occurred. Preparing and starting the consistency group can restart the copy.

Suspended
The consistency group was started, but it did not complete. The source VDisks might be unavailable, or the copy bitmap might be offline. If the consistency group does not return to the copying state, stop the consistency group to reset the consistency group.

Related concepts

“VDisks” on page 23
A *virtual disk (VDisk)* is a logical disk that the cluster presents to the storage area network (SAN).

Metro & Global Mirror

The Mirror Copy Service enables you to set up a relationship between two virtual disks (VDisks), so that updates that are made by an application to one VDisk are mirrored on the other VDisk.

Although the application only writes to a single VDisk, the SAN Volume Controller maintains two copies of the data. If the copies are separated by a significant distance, the Mirror copy can be used as a backup for disaster recovery. A prerequisite for the SAN Volume Controller Mirror operations between two clusters is that the SAN fabric to which they are attached provides adequate bandwidth between the clusters.

There are two types of Mirror Copy Services: Metro Mirror and Global Mirror. For both copy types, One VDisk is designated the primary and the other VDisk is designated the secondary. Host applications write data to the primary VDisk, and updates to the primary VDisk are copied to the secondary VDisk. Normally, host applications do not perform I/O operations to the secondary VDisk.

Metro Mirror provides a synchronous-copy process. When a host writes to the primary VDisk, it does not receive confirmation of I/O completion until the write operation has completed for the copy on both the primary VDisk and the secondary VDisk. This ensures that the secondary VDisk is always up-to-date with the primary VDisk in the event that a failover operation must be performed. However, the host is limited to the latency and bandwidth limitations of the communication link to the secondary VDisk.

Global Mirror provides an asynchronous-copy process. When a host writes to the primary VDisk, confirmation of I/O completion is received before the write operation has completed for the copy on the secondary VDisk. If a failover operation is performed, the application must recover and apply any updates that were not committed to the secondary VDisk.

The Mirror Copy Service supports the following features:

- Intracluster copying of a VDisk, in which both VDisks belong to the same cluster and I/O group within the cluster.
- Intercluster copying of a VDisk, in which one VDisk belongs to a cluster and the other VDisk belongs to a different cluster.

  Note: A cluster can only participate in active Mirror relationships with itself and one other cluster.
- Intercluster and intracluster Mirror can be used concurrently within a cluster.
- The intercluster link is bidirectional. This means that it can copy data from clusterA to clusterB for one pair of VDisks while copying data from clusterB to clusterA for a different pair of VDisks.
- The copy direction can be reversed for a consistent relationship.
- Consistency groups are supported to manage a group of relationships that must be kept synchronized for the same application. This also simplifies administration, because a single command that is issued to the consistency group is applied to all the relationships in that group.

Related concepts

“Metro Mirror” on page 70
The Metro Mirror Copy Service provides a *synchronous* copy, which means that the primary virtual disk (VDisk) is always an exact match of the secondary VDisk.

“Mirror consistency groups” on page 70
The Mirror Copy Service allows you to group a number of Mirror relationships into a consistency group so that they can be updated at the same time. A command that is issued to the consistency group is simultaneously applied to all of the relationships in the group.

“VDisks” on page 23
A *virtual disk (VDisk)* is a logical disk that the cluster presents to the storage area network (SAN).

Metro Mirror

The Metro Mirror Copy Service provides a *synchronous* copy, which means that the primary virtual disk (VDisk) is always an exact match of the secondary VDisk.

The host application writes data to the primary VDisk but does not receive confirmation that the write operation is complete until the data is written to the secondary VDisk. For disaster recovery, this mode is the only practical mode of operation because a synchronous copy of the data is maintained. Metro Mirror is constrained by the latency time and bandwidth limitations that are imposed by the communication link to the secondary site.

Related concepts

"Mirror consistency groups"
The Mirror Copy Service allows you to group a number of Mirror relationships into a consistency group so that they can be updated at the same time. A command that is issued to the consistency group is simultaneously applied to all of the relationships in the group.

"VDisks" on page 23
A *virtual disk (VDisk)* is a logical disk that the cluster presents to the storage area network (SAN).

Global Mirror

The Global Mirror Copy Service provides an asynchronous copy because the secondary virtual disk (VDisk) is not an exact match of the primary VDisk at every point in time.

The host application writes data to the primary VDisk and receives confirmation that the write operation is complete before the data is actually written to the secondary VDisk. The functionality is comparable to a continuous backup process in which the last few updates are always missing. Therefore, Global Mirror is more suited for data migration and backup than it is for disaster recovery.

If I/O operations on the primary VDisk are paused for a significant length of time, the secondary VDisk can become an exact match of the primary VDisk.

Mirror consistency groups

The Mirror Copy Service allows you to group a number of Mirror relationships into a consistency group so that they can be updated at the same time. A command that is issued to the consistency group is simultaneously applied to all of the relationships in the group.

Relationships can be based on "loose" or "tight" associations. A more significant use arises when the relationships contain virtual disks (VDisks) with a tight association. A simple example of a tight association is the spread of data for an application across more than one VDisk. A more complex example is when multiple applications run on different host systems. Each application has data on different VDisks, and these applications exchange data with each other. In both examples, specific rules exist as to how the relationships can be updated. This ensures that the set of secondary VDisks contain usable data. The key property is that these relationships are consistent.

Relationships can only belong to one consistency group; however, they do not have to belong to a consistency group. Relationships that are not part of a consistency group are called stand-alone relationships. A consistency group can contain zero or more relationships. All the relationships in a consistency group must have matching primary and secondary clusters, sometimes referred to as master and auxiliary clusters. All relationships in a consistency group must also have the same copy direction and state.

Metro and Global Mirror relationships cannot belong to the same consistency group. A copy type is automatically assigned to a consistency group when the first relationship is added to the consistency group. After the consistency group is assigned a copy type, only relationships of that copy type can be added to the consistency group. Each cluster can have a maximum of six different types of consistency groups. The following types of consistency groups are possible:

- Intracluster Metro Mirror
- Intercluster Metro Mirror from the local cluster to remote cluster
- Intercluster Metro Mirror from the remote cluster to local cluster
- Intracluster Global Mirror
- Intercluster Global Mirror from the local cluster to remote cluster
- Intercluster Global Mirror from the remote cluster to local cluster

States

A consistency group can be in one of the following states:

Inconsistent (stopped)
The primary VDisks are accessible for read and write I/O operations but the secondary VDisks are not accessible for either. A copy process must be started to make the secondary VDisks consistent.

Inconsistent (copying)
The primary VDisks are accessible for read and write I/O operations but the secondary VDisk are not accessible for either. This state is entered after the **svctask startrcconsistgrp** command is issued to a consistency group in the InconsistentStopped state. This state is also entered when the **svctask startrcconsistgrp** command is issued, with the force option, to a consistency group in the Idling or ConsistentStopped state.

Consistent (stopped)
The secondary VDisks contain a consistent image, but it might be out-of-date with respect to the primary VDisks. This state can occur when a relationship was in the ConsistentSynchronized state and experiences an error that forces a freeze of the consistency group. This state can also occur when a relationship is created with the CreateConsistentFlag set to `TRUE`.

Consistent (synchronized)
The primary VDisks are accessible for read and write I/O operations. The secondary VDisks are accessible for read-only I/O operations.

Idling Both the primary VDisks and the secondary VDisks are operating in the primary role. Consequently the VDisks are accessible for write I/O operations.

Idling (disconnected)
The VDisks in this half of the consistency group are all operating in the primary role and can accept read or write I/O operations.

Inconsistent (disconnected)
The VDisks in this half of the consistency group are all operating in the secondary role and cannot accept read or write I/O operations.

Consistent (disconnected)
The VDisks in this half of the consistency group are all operating in the secondary role and can accept read I/O operations but not write I/O operations.

Empty The consistency group does not contain any relationships.

Related concepts

"VDisks" on page 23
A *virtual disk (VDisk)* is a logical disk that the cluster presents to the storage area network (SAN).

Chapter 6. Planning for configuring the SAN Volume Controller

Ensure that you perform all the required and necessary planning tasks before you begin configuring the SAN Volume Controller.

When planning to configure the SAN Volume Controller, you must complete the following planning tasks:

Planning the clusters

When planning to create your SAN Volume Controller cluster, determine the following:

- Determine the number of clusters and the number of pairs of nodes. Each pair of nodes (the I/O group) is the container for one or more virtual disks (VDisks).
- Determine the number of hosts that will be used with the SAN Volume Controller. Hosts should be grouped by operating system and by type of host bus adapter (HBA).
- Determine the number of I/Os per second between the hosts and SAN Volume Controller nodes.

Planning the host groups

Host systems have access to specific logical units (LUs) within the disk controllers as a result of LUN masking. To plan a host group, gather the following information:

- List all of the worldwide port names (WWPNs) of the fibre-channel host bus adapter ports in the hosts.
- Determine the name to assign to the host or host group.
- Determine the VDisks to assign to the host.

Planning the MDisks

To plan the managed disks (MDisks), determine the logical or physical disks (logical units) in the back-end storage.

Planning the MDisk groups

Before you create managed disk (MDisk) groups, determine the following factors:

- Determine the types of back-end controllers in the system.
- If you want to create VDisks with the sequential policy, plan to create a separate MDisk group for these VDisks or ensure that you create these VDisks before creating VDisks with the striped policy.
- Plan to create MDisk groups for the back-end controllers that provide the same level of performance or reliability, or both. For example, you can group all of the managed disks that are RAID 10 in one MDisk group and all of the MDisks that are RAID 5 in another group.

Planning the VDisks

An individual VDisk is a member of one MDisk group and one I/O group. The MDisk group defines which MDisks provide the back-end storage that makes up the VDisk. The I/O group defines which SAN Volume Controller nodes provide I/O access to the VDisk. Determine the following information before creating a VDisk:

- Determine whether there is data on the volume that needs to be preserved.
- Determine the name you want to assign to the VDisk.
- Determine the I/O group to which the VDisk will be assigned.
- Determine the MDisk group to which the VDisk will be assigned.
- Determine the capacity of the VDisk.

Maximum configuration

Ensure that you are familiar with the maximum configurations of the SAN Volume Controller.

See the following Web site for the latest maximum configuration support:

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

Configuration rules and requirements

Ensure that you understand the rules and requirements when configuring the SAN Volume Controller.

Table 14 provides terms and definitions that can guide your understanding of the rules and requirements.

Table 14. Configuration terms and definitions

Term	**Definition**
ISL hop	A hop on an interswitch link (ISL). With reference to all pairs of N-ports or end-nodes that are in a fabric, the number of ISL hops is the number of links that are crossed on the shortest route between the node pair whose nodes are farthest apart from each other. The distance is measured only in terms of the ISL links that are in the fabric.
Oversubscription	The ratio of the sum of the traffic that is on the initiator N-node connections to the traffic that is on the most heavily-loaded ISLs or where more than one ISL is in parallel between these switches. This definition assumes a symmetrical network and a specific workload that is applied equally from all initiators and sent equally to all targets. A symmetrical network means that all initiators are connected at the same level and all the controllers are connected at the same level. **Note:** The SAN Volume Controller puts its back-end traffic onto the same symmetrical network. The back-end traffic can vary by workload. Therefore, the oversubscription that a 100% read hit gives is different from the oversubscription that 100% write-miss gives. If you have an oversubscription of 1 or less, the network is nonblocking.
Virtual SAN (VSAN)	A VSAN is a virtual storage area network (SAN).

Table 14. Configuration terms and definitions (continued)

Term	Definition
Redundant SAN	A SAN configuration in which if any one component fails, connectivity between the devices that are in the SAN is maintained, possibly with degraded performance. Create a redundant SAN by splitting the SAN into two independent counterpart SANs.
Counterpart SAN	A non-redundant portion of a redundant SAN. A counterpart SAN provides all the connectivity of the redundant SAN, but without the redundancy. The SAN Volume Controller is typically connected to a redundant SAN that is made out of two counterpart SANs.
Local fabric	The fabric that consists of those SAN components (switches and cables) that connect the components (nodes, hosts, and switches) of the local cluster. Because the SAN Volume Controller supports Metro and Global Mirror, significant distances might exist between the components of the local cluster and those of the remote cluster.
Remote fabric	The fabric that consists of those SAN components (switches and cables) that connect the components (nodes, hosts, and switches) of the remote cluster. Because the SAN Volume Controller supports Metro and Global Mirror, significant distances might exist between the components of the local cluster and those of the remote cluster.
Local/remote fabric interconnect	The SAN components that connect the local fabrics to the remote fabrics. There might be significant distances between the components in the local cluster and those in the remote cluster. These components might be single-mode optical fibres that are driven by Gigabit Interface Converters (GBICs), or they might be other, more advanced components, such as channel extenders
SAN Volume Controller fibre-channel port fan in	The number of hosts that can see any one port. Some controllers recommend that the number of hosts using each port be limited to prevent excessive queuing at that port. If the port fails or the path to that port fails, the host might failover to another port, and the fan in requirements might be exceeded in this degraded mode.
Invalid configuration	In an invalid configuration, an attempted operation fails and will generate an error code to indicate what caused it to become invalid.
Unsupported configuration	A configuration that might operate successfully, but for which IBM does not guarantee the solution for problems that might occur. Usually this type of configuration does not create an error log entry.
Valid configuration	A configuration that is neither invalid nor unsupported.
Degraded	A valid configuration that has had a failure, but continues to be neither invalid nor unsupported. Typically, a repair action is required to restore the degraded configuration to a valid configuration.
Fibre channel extender	A device for long distance communication connecting other SAN fabric components. Generally these might involve protocol conversion to ATM, IP or some other long distance communication protocol.
Mesh configuration	A network that contains a number of small SAN switches configured to create a larger switched network. With this configuration, four or more switches are connected together in a loop with some of the paths short circuiting the loop. An example of this configuration is to have four switches connected together in a loop with ISLs for one of the diagonals. The SAN Volume Controller does not support this configuration.

Configuration rules

Storage area network (SAN) configurations that contain SAN Volume Controller nodes can be configured in various ways.

A SAN configuration that contains SAN Volume Controller nodes must follow the rules for the following components:

- Storage subsystems
- HBAs
- Nodes
- Fibre-channel switches
- Fabrics
- Port Switches
- Zoning
- Power requirements

Storage subsystems

Follow these rules when you are planning the configuration of storage subsystems in the SAN fabric.

All SAN Volume Controller nodes in a cluster must be able to see the same set of storage subsystem ports on each device. Any operation that is in this mode in which two nodes do not see the same set of ports on the same device is degraded, and the system logs errors that request a repair action. This rule can have important effects on a storage subsystem such as an IBM System Storage DS4000 series controller, which has exclusion rules that determine to which host bus adapter (HBA) WWNNs a storage partition can be mapped.

A configuration in which a SAN Volume Controller bridges a separate host device and a RAID is not supported. Typical compatibility matrixes are shown in a document titled *Supported Hardware List* on the following Web page:

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

The SAN Volume Controller clusters must not share its storage subsystem logical units (LUs) with hosts. A storage subsystem can be shared with a host under certain conditions as described in this topic.

You can configure certain storage controllers to safely share resources between the SAN Volume Controller and direct attached hosts. This type of configuration is described as a split controller. In all cases, it is critical that you configure the controller and SAN so that the SAN Volume Controller cannot access logical units (LUs) that a host or another SAN Volume Controller can also access. This split controller configuration can be arranged by controller logical unit number (LUN) mapping and masking. If the split controller configuration is not guaranteed, data corruption can occur.

Besides a configuration where a controller is split between a SAN Volume Controller and a host, the SAN Volume Controller also supports configurations where a controller is split between two SAN Volume Controller clusters. In all cases, it is critical that you configure the controller and SAN so that the SAN Volume Controller cannot access LUs that a host or another SAN Volume Controller can also access. This can be arranged by controller LUN mapping and masking. If this is not guaranteed, data corruption can occur. Do not use this configuration because of the risk of data corruption.

Avoid configuring one storage subsystem device to present the same LU to more than one SAN Volume Controller cluster. This configuration is not supported and is very likely to cause undetected data loss or corruption.

The SAN Volume Controller must be configured to manage only LUNs that are presented by supported disk controller systems. Operation with other devices is not supported.

Unsupported storage subsystem (generic device)

When a storage subsystem is detected on the SAN, the SAN Volume Controller attempts to recognize it using its Inquiry data. If the device is recognized as one of the explicitly supported storage models, the SAN Volume Controller uses error recovery programs that are potentially tailored to the known needs of the storage subsystem. If the device is not recognized, the SAN Volume Controller configures the device as a generic device. A generic device might not function correctly when it is addressed by a SAN Volume Controller. In any event, the SAN Volume Controller does not regard accessing a generic device as an error condition and, consequently, does not log an error. Managed disks (MDisks) that are presented by generic devices are not eligible to be used as quorum disks.

Split controller configurations

The SAN Volume Controller is configured to manage LUs that are exported only by RAID controllers. Operation with other RAID controllers is illegal. While it is possible to use the SAN Volume Controller to manage JBOD (just a bunch of disks) LUs that are presented by supported RAID controllers, the SAN Volume Controller itself does not provide RAID functions, so these LUs are exposed to data loss in the event of a disk failure.

If a single RAID controller presents multiple LUs, either by having multiple RAID configured or by partitioning one or more RAID into multiple LUs, each LU can be owned by either SAN Volume Controller or a directly attached host. Suitable LUN masking must be in place to ensure that LUs are not shared between SAN Volume Controller nodes and direct attached hosts.

In a split controller configuration, a RAID presents some of its LUs to a SAN Volume Controller (which treats the LU as an MDisk) and the remaining LUs to another host. The SAN Volume Controller presents virtual disks (VDisks) that are created from the MDisk to another host. There is no requirement for the multipathing driver for the two hosts to be the same. Figure 18 on page 78 shows that the RAID controller is an IBM DS4000, with RDAC used for pathing on the directly attached host, and SDD used on the host that is attached with the SAN Volume Controller. Hosts can simultaneously access LUs that are provided by the SAN Volume Controller and directly by the device.

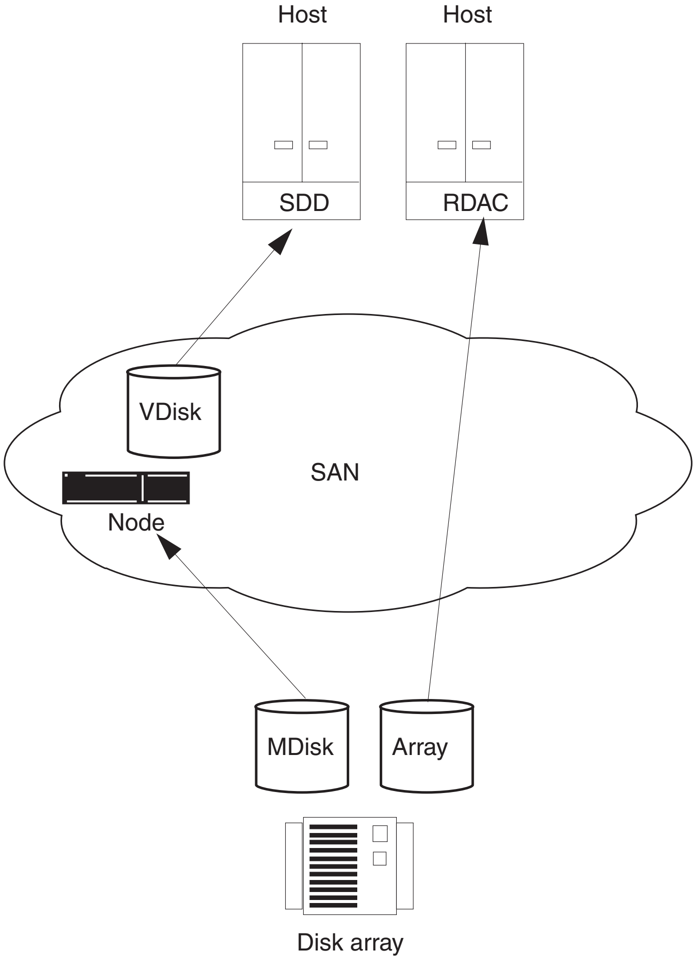


Figure 18. Disk controller system shared between SAN Volume Controller and a host

It is also possible to split a host so that it accesses some of its LUNs through the SAN Volume Controller and some directly. In this case, the multipathing software that is used by the controller must be compatible with the SAN Volume Controller nodes multipathing software. Figure 19 on page 79 is a supported configuration because the same multipathing driver is used for both direct and VDisks.

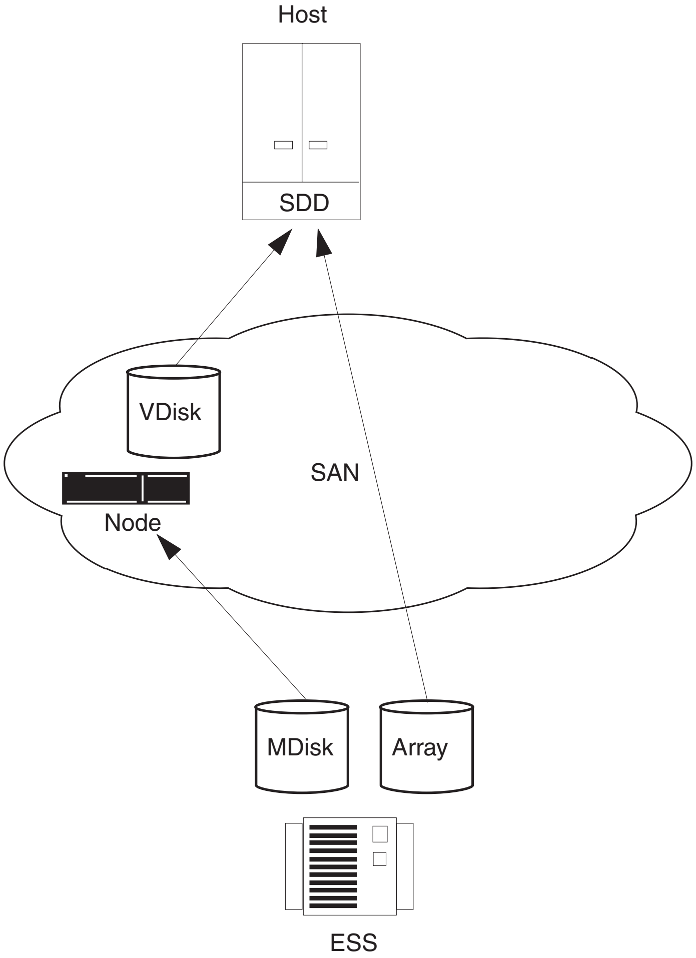


Figure 19. IBM ESS LUs accessed directly with a SAN Volume Controller

In the case where the RAID controller uses multipathing software that is compatible with SAN Volume Controller nodes multipathing software (see Figure 20 on page 80), it is possible to configure a system where some LUNs are mapped directly to the host and others are accessed through the SAN Volume Controller. An IBM TotalStorage Enterprise Storage Server (ESS) that uses the same multipathing driver as a SAN Volume Controller is one example.

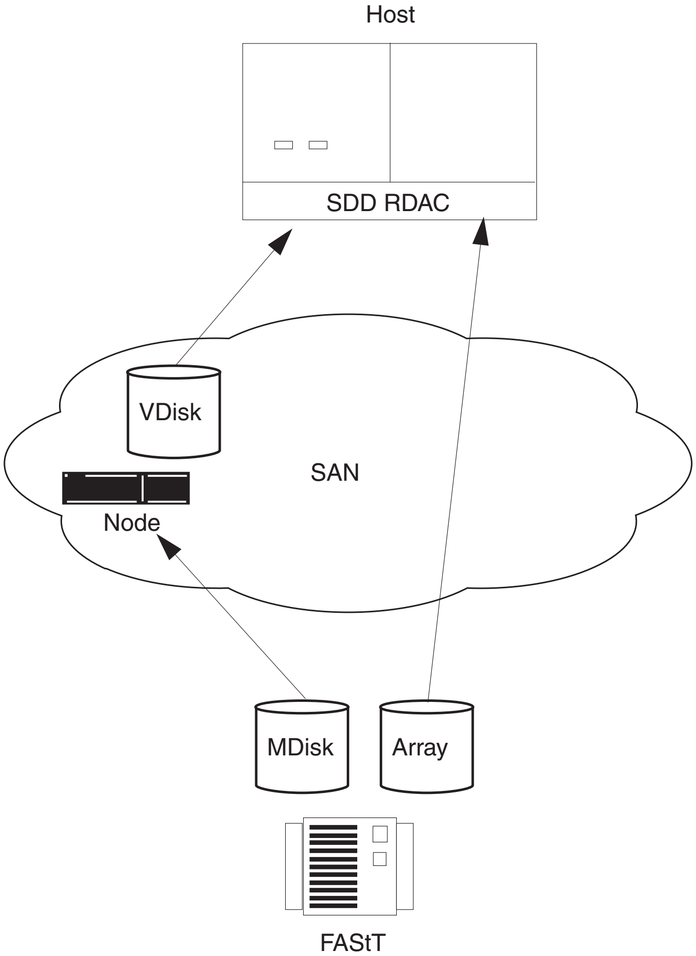


Figure 20. IBM DS4000 direct connection with a SAN Volume Controller on one host

Related concepts

"Cluster operation and quorum disks" on page 59
The cluster must contain at least half of its nodes to function.

"MDisks" on page 18
A *managed disk (MDisk)* is a logical disk (typically a RAID or partition thereof) that a storage subsystem has exported to the SAN fabric to which the nodes in the cluster are attached.

HBAs

Ensure that you are familiar with the configuration rules for host bus adapters (HBAs). You must abide by the configuration rules for HBAs to ensure that you have a valid configuration.

SAN Volume Controller 2145-4F2 and SAN Volume Controller 2145-8F2 nodes contain two 2-port HBAs. If one HBA fails, the configuration is still valid, and the SAN Volume Controller node operates in degraded mode. If an HBA is physically removed, the configuration is not supported.

SAN Volume Controller 2145-8F4 nodes contain one 4-port HBA.

HBAs that are in dissimilar hosts or dissimilar HBAs that are in the same host must be in separate zones. *Dissimilar* means that the hosts are running different operating systems or that they are different hardware platforms. For example, if you have an HP-UX host and a Windows 2000 server host, those hosts must be in separate zones. Different levels of the same operating system are considered to be similar. A configuration that breaks this requirement is not supported.

The SAN Volume Controller must be configured to export virtual disks (VDisks) only to host fibre-channel ports that are on the supported HBAs. See the following Web site for specific firmware levels and the latest supported hardware:

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

Operation with other HBAs is not supported.

The SAN Volume Controller does not specify the number of host fibre-channel ports or HBAs that a host or a partition of a host can have. The number of host fibre-channel ports or HBAs are specified by the host multipathing device driver. The SAN Volume Controller supports this number; however it is subject to the configuration rules for the SAN Volume Controller. To obtain optimal performance and to prevent overloading, the workload to each SAN Volume Controller port must be equal. You can achieve an even workload by zoning approximately the same number of host fibre-channel ports to each SAN Volume Controller fibre-channel port.

Nodes

You must follow the configuration rules for SAN Volume Controller nodes to ensure that you have a valid configuration.

HBAs

SAN Volume Controller 2145-4F2 and SAN Volume Controller 2145-8F2 nodes contain two 2-port HBAs. If one HBA fails, the configuration is still valid, and the node operates in degraded mode. If an HBA is physically removed, the configuration is not supported.

SAN Volume Controller 2145-8F4 nodes contain one 4-port HBA.

I/O groups

Nodes must always be used in pairs called I/O groups. SAN Volume Controller 2145-4F2, SAN Volume Controller 2145-8F2, and SAN Volume Controller 2145-8F4 nodes can be in the same I/O group during online upgrade procedures. If a node fails or is removed from the configuration, the remaining node in the I/O group operates in a degraded mode, but the configuration is still valid.

VDisks

Each node presents a virtual disk (VDisk) to the SAN through four ports. Each VDisk is accessible from the two nodes in an I/O group. A host HBA might recognize up to eight paths to each logical unit (LU) that is presented by the node. The hosts must run a multipathing device driver before the multiple paths can resolve to a single device.

Optical connections

Support for optical connections is based on the fabric rules that the manufacturers impose for the following connection methods:

- Host to a switch
- Backend to a switch
- Interswitch links (ISLs)

Short wave optical fibre connections must be used between a node and its switches. Clusters that use intercluster Metro Mirror can use short or long wave optical fibre connections, or distance-extender technology that is supported by the switch manufacturer.

To ensure cluster failover operations, all nodes in a cluster must be connected to the same IP subnet.

The number of paths through the network from the node to a host must not exceed eight. Configurations in which this number is exceeded are unsupported. Each node has four ports and each I/O group has two nodes. Therefore, without any zoning, the number of paths to a VDisks is eight × the number of host ports.

Port speed

The optical fibre-connections between fibre-channel switches and all SAN Volume Controller 2145-4F2 and SAN Volume Controller 2145-8F2 nodes must run at one port speed. The fibre-channel ports on SAN Volume Controller 2145-8F4 nodes auto negotiate the operational port speed independently, which allows these nodes to operate at different speeds.

UPS

Nodes must be connected to the uninterruptible power supply (UPS) using the supplied cable that joins the signal and power cables.

Power requirements

Ensure you are familiar with the configuration rules for power requirements. You must abide by the configuration rules for power requirements to ensure you have a valid configuration.

The uninterruptible power supply (UPS) must be in the same rack that contains the SAN Volume Controller node that it supplies. The SAN Volume Controller 2145-8F2 and SAN Volume Controller 2145-8F4 nodes must be connected to a 2145 uninterruptible power supply-1U (2145 UPS-1U) because these models cannot operate with a 2145 uninterruptible power supply (2145 UPS).

The combination power and signal cable for connection between the SAN Volume Controller and the UPS units is 2 m long. The SAN Volume Controller and UPS must connect with both the power and the signal cable to function correctly.

Fibre-channel switches

Ensure that you are familiar with the configuration rules for fibre-channel switches. You must follow the configuration rules for fibre-channel switches to ensure that you have a valid configuration.

The SAN must contain only supported switches.

See the following Web site for specific firmware levels and the latest supported hardware:

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

The SAN should consist of two independent switches (or networks of switches) so that the SAN includes a redundant fabric, and has no single point of failure. If one SAN fabric fails, the configuration is in a degraded mode, but it is still valid. If the SAN contains only one fabric, it is still a valid configuration, but a failure of the fabric might cause a loss of access to data. Therefore, SANs with one fabric are considered to have a possible single point of failure.

Configurations with more than four SANs are not supported.

The SAN Volume Controller nodes must always and only be connected to SAN switches. Each node must be connected to each of the counterpart SANs that are in the redundant fabric. Any configuration that uses direct connections between host and node or between controller and node is not supported. SANs that are created from a mesh of switches are not supported.

All back-end storage must always and only be connected to SAN switches. Multiple connections are permitted from the redundant controllers of the back-end storage to improve data bandwidth performance. It is not necessary to have a connection between each redundant disk controller system of the back-end storage and each counterpart SAN. For example, in an IBM System Storage DS4000 configuration in which the IBM DS4000 contains two redundant controllers, only two controller minihubs are usually used. Controller A of the IBM DS4000 is connected to counterpart SAN A, and controller B of the IBM DS4000 is connected to counterpart SAN B. Any configuration that uses a direct connection between the host and the controller is not supported.

When you attach the a node to a SAN fabric that contains core directors and edge switches, connect the node ports to the core directors and connect the host ports to the edge switches. In this type of fabric, the next priority for connection to the core directors is the storage controllers, leaving the host ports connected to the edge switches.

The switch configuration of a SAN Volume Controller SAN must observe the switch manufacturer's configuration rules. These rules might put restrictions on the switch configuration. Any configuration that runs outside the manufacturers' rules is not supported.

The switch must be configured so that the nodes can see the back-end storage and the front-end HBAs. However, the front-end HBAs and the back-end storage must not be in the same zone. Any configuration that does not follow these rules is not supported.

It is critical that you configure the controller and SAN so that a node cannot access LUs that a host or another node can also access. This can be arranged by controller LUN mapping and masking.

All nodes in a SAN Volume Controller cluster must be able to see the same set of back-end storage ports on each back-end controller. Operation in a mode where two nodes see a different set of ports on the same controller is degraded and the system logs errors that request a repair action. This can occur if inappropriate zoning was applied to the fabric or if inappropriate LUN masking is used. This rule has important implications for back-end storage, such as an IBM DS4000, which imposes exclusive rules for mappings between HBA worldwide node names (WWNNs) and storage partitions.

Because each node has four ports, the switches can be zoned so that a particular port is used only for internode communication, for communication to the host, or for communication to back-end storage. For all configurations, each node must remain connected to the full SAN fabric. You must not use zoning to split the SAN into two parts.

Operational port speed

You can change the operational port speed for SAN Volume Controller 2145-4F2 and SAN Volume Controller 2145-8F2 nodes to 1 Gbps or 2 Gbps. However, the optical-fibre connections between the fibre-channel switches and all SAN Volume Controller 2145-4F2 and SAN Volume Controller 2145-8F2 nodes in a cluster must run at the same speed. The fibre channel ports on SAN Volume Controller 2145-8F4 nodes auto negotiate the operational port speed independently, which allows these nodes to operate at different speeds. SAN Volume Controller 2145-8F4 nodes can operate at 1 Gbps, 2 Gbps or 4 Gbps. If a SAN Volume Controller 2145-8F4 node is connected to a 4 Gbps capable switch, the port attempts to operate at 4 Gbps; however, if there is a large number of link error rates, the adapter negotiates a lower speed.

Mixing manufacturer switches in a single SAN fabric

Within an individual SAN fabric, switches must have the same manufacturer, with the following exceptions:

- BladeCenter®. See the documentation that is provided with your BladeCenter for more information.
- Where one pair of counterpart fabrics (for example, Fabric A and Fabric B) provide a redundant SAN, different manufacturer’s switches can be mixed in a SAN Volume Controller configuration, provided that each fabric contains only switches from a single manufacturer. Thus, the two counterpart SANs can have different manufacturer’s switches.
- The SAN Volume Controller supports the Interoperability Modes of the Cisco MDS 9000 family of switch and director products with the following restrictions:
  - The Cisco MDS 9000 must be connected to Brocade and McData switch/director products with the multivendor fabric zones connected using MDS Interoperability Mode 1, 2 or 3.
  - All of the SAN Volume Controller nodes that are in the SAN Volume Controller cluster must be attached to the Cisco part of the counterpart fabric or they must be attached to the McData or Brocade part of the counterpart fabric to avoid having a single fabric with a SAN Volume Controller cluster that has part of the SAN Volume Controller nodes connected to Cisco switch ports and part of the SAN Volume Controller nodes connected to Brocade or McData switch ports.

Brocade core-edge fabrics

Brocade core-edge fabrics that use the M14 or M48 model can have up to 256 hosts under the following conditions:

- Each SAN Volume Controller port cannot see more than 256 node port logins.
- Each I/O group cannot be associated with more than 64 hosts.

- A host can be associated with more than one I/O group.
- Each HBA port must be in a separate zone and each zone must contain one port from each SAN Volume Controller node in the I/O group that the host accesses.
- M14, M48 or other Brocade models can be used as edge switches; however, the SAN Volume Controller ports and back-end storage must all be connected to the M14 or M48 core-edge switch.
- You can attach between one and four separate fabrics to the SAN Volume Controller cluster. If other manufacturer fabrics are also attached to the SAN Volume Controller cluster, you must follow the SAN Volume Controller support guidelines for that manufacturer.
- A host can access VDisks from different I/O groups in a Brocade SAN, but this reduces the maximum number of hosts that can be used in the SAN. For example, if the same host uses VDisks in two different I/O groups, this consumes one of the 64 hosts in each I/O group. If each host accesses VDisks in each I/O group, there can only be 64 hosts in the configuration. Alternatively, if each host accesses VDisks in two I/O groups, 32 different hosts can be attached to each I/O group. This means that 128 hosts can be used in an 8 node cluster.

Fibre-channel switches and interswitch links

The local or remote fabric must not contain more than three interswitch link (ISL) hops in each fabric. Any configuration that uses more than three ISL hops is not supported. When a local fabric is connected to a remote fabric for Metro Mirror, the ISL hop count between a local node and a remote node must not exceed seven. Therefore, some ISL hops can be used in a cascaded switch link between local and remote clusters if the internal ISL count of the local or remote cluster is less than three.

If all three allowed ISL hops are used within the local and remote fabrics, the local remote fabric interconnect must be a single ISL hop between a switch in the local fabric and a switch in the remote fabric.

The SAN Volume Controller supports the use of distance-extender technology, including DWDM (Dense Wavelength Division Multiplexing) and FCPIP extenders, to increase the overall distance between local and remote clusters. If this extender technology involves a protocol conversion, the local and remote fabrics should be regarded as independent fabrics, limited to three ISL hops each. The only restriction on the interconnection between the two fabrics is the maximum latency that is allowed in the distance extender technology.

Note: Where multiple ISL hops are used between switches, follow the fabric manufacturer's recommendations for trunking.

When ISLs are used, each ISL oversubscription must not exceed six. Any configuration that uses higher values is not supported.

With ISLs between nodes in the same cluster, the ISLs are considered a single point of failure. This is illustrated in Figure 21 on page 86.

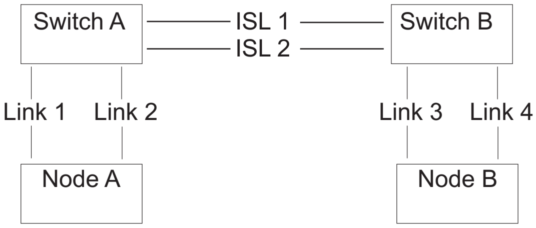


Figure 21. Fabric with ISL between nodes in a cluster

If Link 1 or Link 2 fails, the cluster communication does not fail.

If Link 3 or Link 4 fails, the cluster communication does not fail.

If ISL 1 or ISL 2 fails, the communication between Node A and Node B fails for a period of time, and the node is not recognized, even though there is still a connection between the nodes.

To ensure that a fibre-channel link failure does not cause nodes to fail when there are ISLs between nodes, it is necessary to use a redundant configuration. This is illustrated in Figure 22.

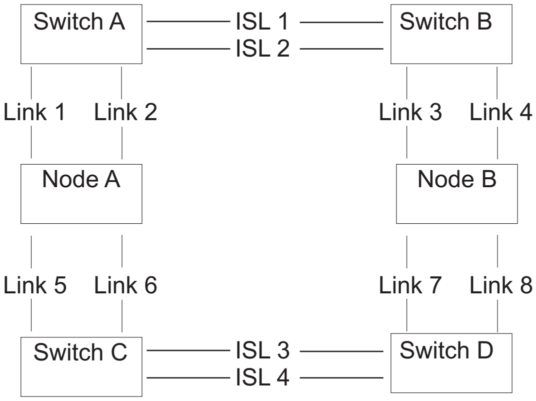


Figure 22. Fabric with ISL in a redundant configuration

With a redundant configuration, if any one of the links fails, communication on the cluster does not fail.

SAN Volume Controller in a SAN with director class switches

You can use director class switches within the SAN to connect large numbers of RAID controllers and hosts to a SAN Volume Controller cluster. Because director class switches provide internal redundancy, one director class switch can replace a SAN that uses multiple switches. However, the director class switch provides only network redundancy; it does not protect against physical damage (for example, flood or fire), which might destroy the entire function. A tiered network of smaller switches or a core-edge topology with multiple switches in the core can provide comprehensive redundancy and more protection against physical damage for a network in a wide area.

Configuration requirements

Ensure that you are familiar with the configuration requirements for the SAN Volume Controller. You must abide by the configuration requirements for the SAN Volume Controller to ensure you have a valid configuration.

You *must* perform the following steps before you configure the SAN Volume Controller.

1. Your IBM service representative must have installed the SAN Volume Controller.
2. Install and configure your disk controller systems and create the RAID resources that you intend to virtualize. To prevent loss of data, virtualize only those RAIDs that provide some kind of redundancy, that is, RAID 1, RAID 10, RAID 0+1, or RAID 5. Do *not* use RAID 0 because a single physical disk failure might cause the failure of many virtual disks (VDisks). RAID 0, like other types of RAID offers cost-effective performance by using available capacity through data striping. However, RAID 0 does not provide a parity disk drive for redundancy (RAID 5) or mirroring (RAID 10).

   When creating RAID with parity protection (for example, RAID 5), consider how many component disks to use in each array. The more disks that you use, the fewer disks that you need to provide availability for the same total capacity (one per array). However, if you use more disks, it takes longer to rebuild a replacement disk after a disk failure. If a second disk failure occurs during the rebuild period, all data on the array is lost. More data is affected by a disk failure for a larger number of member disks resulting in reduced performance while rebuilding onto a hot spare and more data being exposed if a second disk fails before the rebuild has completed. The smaller the number of disks, the more likely it is that write operations span an entire stripe (stripe size x number of members minus 1). In this case, write performance is improved because the disk write operations do not have to be preceded by disk reads. The number of disk drives that are required to provide availability might be unacceptable if the arrays are too small.

   When in doubt, create arrays with between six and eight member disks.

   If reasonably small RAIDs are used, it is easier to extend a managed disk (MDisk) group by adding a new RAID of the same type. Construct multiple RAID devices of the same type, when it is possible.

   When you create RAID with mirroring, the number of component disks in each array does not affect redundancy or performance.

   Most back-end disk controller systems enable RAID to be divided into more than one SCSI logical unit (LU). When you configure new storage for use with the SAN Volume Controller, you do not have to divide up the array. New storage is presented as one SCSI LU. This gives a one-to-one relationship between MDisks and RAID.

   Attention: Losing an array in an MDisk group can result in the loss of access to *all* MDisks in that group.
3. Install and configure your switches to create the zones that the SAN Volume Controller requires. One zone must contain all the disk controller systems and the SAN Volume Controller nodes. For hosts with more than one port, use switch zoning to ensure that each host fibre-channel port is zoned to exactly one fibre-channel port of each SAN Volume Controller node in the cluster. Set up a zone on each fibre-channel switch that includes all of the SAN Volume Controller ports that are connected to that switch.

4. If you want the SAN Volume Controller to export redundant paths to disks, you must install a multipathing device on all of the hosts that are connected to the SAN Volume Controller. Otherwise, you cannot use the redundancy inherent in the configuration. You can install the subsystem device driver (SDD) from the following Web site:

   http://www.ibm.com/server/storage/support/software/sdd.html
5. Install and configure the SAN Volume Controller master console (see the *IBM System Storage Master Console for SAN File System and SAN Volume Controller: Installation and User's Guide*). The communication between the master console and the SAN Volume Controller runs under a client-server network application called Secure Shell (SSH). The SSH server software is already installed on each SAN Volume Controller cluster. The SSH client software called PuTTY is already installed on the master console. You will need to configure the SSH client key pair using PuTTY on the master console.
   a. You can configure the SAN Volume Controller using the SAN Volume Controller Console Web-based application that is preinstalled on the master console.

      Note: You can also install the master console on another machine (which you provide) using the CD-ROM provided with the master console.
   b. You can configure the SAN Volume Controller using the CLI commands.
   c. You can install an SSH client if you only want to use the CLI commands. If you want to use the CLI from a host other than the master console, ensure that the host has an SSH client installed on it.

   Note:
   - AIX comes with an installed SSH client.
   - Linux® comes with an installed SSH client.
   - Use PuTTY for Windows.

When you and the IBM service representative have completed the initial preparation steps, you must perform the following steps:

1. Add nodes to the cluster and set up the cluster properties.
2. Create MDisk groups from the MDisks to make pools of storage from which you can create VDisks.
3. Create host objects from the host bus adapter (HBA) fibre-channel ports to which you can map VDisks.
4. Create VDisks from the capacity that is available in your MDisk groups.
5. Map the VDisks to the host objects to make the disks available to the hosts, as required.
6. Optionally, create Copy Services (FlashCopy and Mirror) objects, as required.

   Related concepts

   "MDisk groups" on page 20
   A *managed disk (MDisk) group* is a collection of MDisks that jointly contain all the data for a specified set of virtual disks (VDisks).

   Related reference

   "Fibre-channel switches" on page 82
   Ensure that you are familiar with the configuration rules for fibre-channel switches. You must follow the configuration rules for fibre-channel switches to ensure that you have a valid configuration.

Chapter 7. SAN Volume Controller supported environment

The IBM Web site provides up-to-date information about the supported environment for the SAN Volume Controller.

This includes:
- Host attachments
- Physical disk storage systems
- Host bus adapters
- Switches

See the following Web site for specific firmware levels and the latest supported hardware:

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

Supported host attachments

The IBM Web site provides up-to-date information about the supported host attachment operating systems.

For a list of supported host attachment operating systems, see the following Web site:

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

Supported storage subsystems

The IBM Web site provides up-to-date information about the supported physical disk storage systems.

For a list of supported storage systems, see the following Web site:

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

Supported fibre-channel host bus adapters

The IBM Web site provides up-to-date information about the supported host bus adapters.

Ensure that host bus adapters (HBAs) are at or above the minimum requirements.

For a list of supported HBAs, see the following Web site for specific firmware levels and the latest supported hardware:

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

Supported switches

The IBM Web site provides up-to-date information about the supported fibre-channel switches.

Ensure that switches are at or above the minimum requirements.

The SAN must contain only supported switches.

See the following Web site for the latest models and firmware levels:

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

Operation with other switches is not supported.

Supported fibre-channel extenders

The supported hardware for the SAN Volume Controller frequently changes.

See the following Web site for the latest supported hardware:

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

Accessibility

Accessibility features help a user who has a physical disability, such as restricted mobility or limited vision, to use software products successfully.

Features

These are the major accessibility features in the SAN Volume Controller master console:

- You can use screen-reader software and a digital speech synthesizer to hear what is displayed on the screen. The following screen readers have been tested: JAWS v4.5 and IBM Home Page Reader v3.0.
- You can operate all features using the keyboard instead of the mouse.

Navigating by keyboard

You can use keys or key combinations to perform operations and initiate many menu actions that can also be done through mouse actions. You can navigate the SAN Volume Controller Console and help system from the keyboard by using the following key combinations:

- To traverse to the next link, button, or topic, press Tab inside a frame (page).
- To expand or collapse a tree node, press → or ←, respectively.
- To move to the next topic node, press V or Tab.
- To move to the previous topic node, press ^ or Shift+Tab.
- To scroll all the way up or down, press Home or End, respectively.
- To go back, press Alt+←.
- To go forward, press Alt+→.
- To go to the next frame, press Ctrl+Tab.
- To move to the previous frame, press Shift+Ctrl+Tab.
- To print the current page or active frame, press Ctrl+P.
- To select, press Enter.

Accessing the publications

You can view the publications for the SAN Volume Controller in Adobe Portable Document Format (PDF) using the Adobe Acrobat Reader. The PDFs are provided at the following Web site:

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

Related reference

"SAN Volume Controller library and related publications" on page xi
A list of other publications that are related to this product are provided to you for your reference.

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 give 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 PUBLICATIONS "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 Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites 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 measurement may have been estimated through extrapolation. Actual results may vary. Users of this document may 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 may be addressed to the suppliers of those products.

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

This information is for planning purposes only. The information herein is subject to change before the products described become available.

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.

Trademarks

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

- AIX
- BladeCenter
- Enterprise Storage Server
- FlashCopy
- IBM
- IBM eServer
- IBM TotalStorage
- IBM System Storage
- System p5
- System z9
- System Storage
- TotalStorage
- xSeries

Intel and Pentium are trademarks of Intel Corporation in the United States, other countries, or both.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

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

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

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

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

Definitions of notices

Ensure that you understand the typographic conventions that are used to indicate special notices.

The notices throughout the SAN Volume Controller documentation and in the *IBM System Safety Notices* follow specific guidelines for their content.

The following notices are used throughout this library to convey specific meanings:

DANGER

These notices indicate situations that can be potentially lethal or extremely hazardous to you. A danger notice precedes the description of a potentially lethal or extremely hazardous procedural step or situation.

CAUTION:
These notices indicate situations that can be potentially hazardous to you. A caution notice precedes the description of a potentially hazardous procedural step or situation.

Attention: These notices indicate possible damage to programs, devices, or data. An attention notice appears before the instruction or the situation in which damage might occur.

Note: These notices provide important tips, guidance, or advice.

Use the reference numbers in parentheses, for example (1), at the end of each notice to find the matching translated notice. For all danger, caution, and attention notices, see the *IBM System Safety Notices*.

Glossary

Ensure you are familiar with the list of terms and their definitions used in this guide.

A

asymmetric virtualization
A virtualization technique in which the virtualization engine is outside the data path and performs a metadata-style service. The metadata server contains all the mapping and locking tables while the storage devices contain only data. See also *symmetric virtualization*.

auxiliary virtual disk
The virtual disk that contains a backup copy of the data and that is used in disaster recovery scenarios. See also *master virtual disk.*

B

blade One component in a system that is designed to accept some number of components (blades). Blades could be individual servers that plug into a multiprocessing system or individual port cards that add connectivity to a switch. A blade is typically a hot-swappable hardware device.

block A unit of data storage on a disk drive.

block virtualization
The act of applying virtualization to one or more block-based (storage) services for the purpose of providing a new aggregated, higher-level, richer, simpler, or secure block service to clients. Block virtualization functions can be nested. A disk drive, RAID system, or volume manager all perform some form of block-address to (different) block-address mapping or aggregation. See also *virtualization*.

C

cache A high-speed memory or storage device used to reduce the effective time required to read data from or write data to lower-speed memory or a device. Read cache holds data in anticipation that it will be requested by a client. Write cache holds data written by a client until it can be safely stored on more permanent storage media such as disk or tape.

cascading
The process of connecting two or more fibre-channel hubs or switches together to increase the number of ports or extend distances.

cluster
In SAN Volume Controller, a pair of nodes that provides a single configuration and service interface.

CIM See *Common Information Model.*

CLI See *command line interface.*

command line-interface (CLI)
A type of computer interface in which the input command is a string of text characters.

Common Information Model (CIM)
A set of standards developed by the Distributed Management Task Force (DMTF). CIM provides a conceptual framework for storage management and an open approach to the design and implementation of storage systems, applications, databases, networks, and devices.

connected
In a Global Mirror relationship, pertaining to the status condition that occurs when two clusters can communicate.

consistency group
A group of copy relationships between virtual disks that are managed as a single entity.

consistent copy
In a Global Mirror relationship, a copy of a secondary virtual disk (VDisk) that is identical to the primary VDisk from the viewpoint of a host system, even if a power failure occurred while I/O activity was in progress.

copied
In a FlashCopy relationship, a state that indicates that a copy has been started after the copy relationship was created. The copy process is complete and the target disk has no further dependence on the source disk.

copying
A status condition that describes the state of a pair of virtual disks (VDisks) that have a copy relationship. The copy process has been started but the two virtual disks are not yet synchronized.

Copy Services
The two services that enable you to copy virtual disks (VDisks): FlashCopy and Global Mirror.

counterpart SAN
A nonredundant portion of a redundant storage area network (SAN). A counterpart SAN provides all the connectivity of the redundant SAN but without the redundancy. Each counterpart SANs provides an alternate path for each SAN-attached device. See also *redundant SAN*.

cross-volume consistency
In SAN Volume Controller, a consistency group property that guarantees consistency between virtual disks when an application issues dependent write operations that span multiple virtual disks.

D

data migration
The movement of data from one physical location to another without disrupting I/O operations.

destage
A write command initiated by the cache to flush data to disk storage.

disk controller
A device that coordinates and controls the operation of one or more disk drives and synchronizes the operation of the drives with the operation of the system as a whole. Disk controllers provide the storage that the cluster detects as managed disks (MDisks).

E

error code
A value that identifies an error condition.

excluded
In SAN Volume Controller, the status of a managed disk that the cluster has removed from use after repeated access errors.

extent A unit of data that manages the mapping of data between managed disks and virtual disks.

F

fabric In fibre-channel technology, a routing structure, such as a switch, that receives addressed information and routes it to the appropriate destination. A fabric can consist of more than one switch. When multiple fibre-channel switches are interconnected, they are described as cascading. See also *cascading*.

failover
In SAN Volume Controller, the function that occurs when one redundant part of the system takes over the workload of another part of the system that has failed.

fibre channel
A technology for transmitting data between computer devices at a data rate of up to 4 Gbps. It is especially suited for attaching computer servers to shared storage devices and for interconnecting storage controllers and drives.

FlashCopy mapping
A relationship between two virtual disks.

FlashCopy relationship
See *FlashCopy mapping*.

FlashCopy service
In SAN Volume Controller, a copy service that duplicates the contents of a source virtual disk (VDisk) to a target VDisk. In the process, the original contents of the target VDisk are lost. See also *point-in-time copy.*

G

Global Mirror
An asynchronous copy service that enables host data on a particular source virtual disk (VDisk) to be copied to the target VDisk that is designated in the relationship.

H

HBA See *host bus adapter*.

host An open-systems computer that is connected to the SAN Volume Controller through a fibre-channel interface.

host bus adapter (HBA)
In SAN Volume Controller, an interface card that connects a host bus, such as a peripheral component interconnect (PCI) bus, to the storage area network.

host ID
In SAN Volume Controller, a numeric identifier assigned to a group of host

fibre-channel ports for the purpose of logical unit number (LUN) mapping. For each host ID, there is a separate mapping of Small Computer System Interface (SCSI) IDs to virtual disks (VDisks).

hub A communications infrastructure device to which nodes on a multi-point bus or loop are physically connected. Commonly used in Ethernet and fibre-channel networks to improve the manageability of physical cables. Hubs maintain the logical loop topology of the network of which they are a part, while creating a “hub and spoke” physical star layout. Unlike switches, hubs do not aggregate bandwidth. Hubs typically support the addition or removal of nodes from the bus while it is operating. (S) Contrast with *switch*.

I

idling

- The status of a pair of virtual disks (VDisks) that have a defined copy relationship for which no copy activity has yet been started.
- In a Global Mirror relationship, that state that indicates that the master virtual disks (VDisks) and auxiliary VDisks are operating in the primary role. Consequently, both VDisks are accessible for write I/O operations.

image mode
An access mode that establishes a one-to-one mapping of extents in the managed disk (MDisk) with the extents in the virtual disk (VDisk). See also *managed space mode* and *unconfigured mode*.

inconsistent
In a Global Mirror relationship, pertaining to a secondary virtual disk (VDisk) that is being synchronized with the primary VDisk.

input/output (I/O)
Pertaining to a functional unit or communication path involved in an input process, an output process, or both, concurrently or not, and to the data involved in such a process.

Internet Protocol (IP)
In the Internet suite of protocols, a connectionless protocol that routes data through a network or interconnected networks and acts as an intermediary between the higher protocol layers and the physical network.

interoperability
The capability to communicate, run programs, or transfer data among various functional units in a way that requires the user to have little or no knowledge of the unique characteristics of those units.

Inter-Switch Link (ISL)
A protocol for interconnecting multiple routers and switches in a storage area network.

I/O See *input/output.*

I/O group
A collection of virtual disks (VDisks) and node relationships that present a common interface to host systems.

IP See *Internet Protocol*.

ISL See *Inter-Switch Link*.

ISL hop
Considering all pairs of node ports (N-ports) in a fabric and measuring

distance only in terms of Inter-Switch Links (ISLs) in the fabric, the number of ISLs traversed is the number of ISL hops on the shortest route between the pair of nodes that are farthest apart in the fabric.

J

JBOD (just a bunch of disks)

- IBM definition: See *non-RAID*.
- HP definition: A group of single-device logical units not configured into any other container type.

L

line card
See *blade*.

local fabric
In SAN Volume Controller, those storage area network (SAN) components (such as switches and cables) that connect the components (nodes, hosts, switches) of the local cluster together.

logical unit (LU)
An entity to which Small Computer System Interface (SCSI) commands are addressed, such as a virtual disk (VDisk) or managed disk (MDisk).

logical unit number (LUN)
The SCSI identifier of a logical unit within a target. (S)

LU See *logical unit*.

LUN See *logical unit number*.

M

managed disk (MDisk)
A Small Computer System Interface (SCSI) logical unit that a redundant array of independent disks (RAID) controller provides and a cluster manages. The MDisk is not visible to host systems on the storage area network (SAN).

managed disk group
A collection of managed disks (MDisks) that, as a unit, contain all the data for a specified set of virtual disks (VDisks).

managed space mode
An access mode that enables virtualization functions to be performed. See also *image mode* and *unconfigured mode*.

Management Information Base (MIB)
Simple Network Management Protocol (SNMP) units of managed information that specifically describe an aspect of a system, such as the system name, hardware number, or communications configuration. A collection of related MIB objects is defined as a MIB.

mapping
See *FlashCopy mapping*.

master virtual disk
The virtual disk (VDisk) that contains a production copy of the data and that an application accesses. See also *auxiliary virtual disk.*

MDisk
See *managed disk*.

mesh configuration
A network that contains a number of small SAN switches configured to create a larger switched network. With this configuration, four or more switches are connected together in a loop with some of the paths short circuiting the loop. An example of this configuration is to have four switches connected together in a loop with ISLs for one of the diagonals. The SAN Volume Controller does not support this configuration.

migration
See *data migration*.

N

node One SAN Volume Controller. Each node provides virtualization, cache, and Copy Services to the storage area network (SAN).

node rescue
In SAN Volume Controller, the process by which a node that has no valid software installed on its hard disk drive can copy the software from another node connected to the same fibre-channel fabric.

non-RAID
Disks that are not in a redundant array of independent disks (RAID). HP definition: See *JBOD*.

O

offline
Pertaining to the operation of a functional unit or device that is not under the continual control of the system or of a host.

online Pertaining to the operation of a functional unit or device that is under the continual control of the system or of a host.

P

point-in-time copy
The instantaneous copy that the FlashCopy service makes of the source virtual disk (VDisk). In some contexts, this copy is known as a *T_0 copy*.

port ID
An identifier associated with a port.

primary virtual disk
In a Global Mirror relationship, the target of write operations issued by the host application.

PuTTY
A free implementation of Telnet and SSH for Windows 32-bit platforms

Q

quorum disk
A managed disk (MDisk) that contains quorum data and that a cluster uses to break a tie and achieve a quorum.

R

rack A free-standing framework that holds the devices and card enclosure.

RAID See *redundant array of independent disks*.

redundant array of independent disks
A collection of two or more disk drives that present the image of a single disk drive to the system. In the event of a single device failure, the data can be read or regenerated from the other disk drives in the array.

redundant SAN
A storage area network (SAN) configuration in which any one single component might fail, but connectivity between the devices within the SAN is maintained, possibly with degraded performance. This configuration is normally achieved by splitting the SAN into two, independent, counterpart SANs. See also *counterpart SAN*.

relationship
In Global Mirror, the association between a master virtual disk (VDisk) and an auxiliary VDisk. These VDisks also have the attributes of a primary or secondary VDisk. See also *auxiliary virtual disk, master virtual disk, primary virtual disk, and secondary virtual disk.*

remote fabric
In Global Mirror, the storage area network (SAN) components (switches and cables) that connect the components (nodes, hosts, and switches) of the remote cluster.

roles Authorization is based on roles that map to the administrator and service roles in an installation. The switch translates these roles into SAN Volume Controller administrator and service user IDs when a connection is made to the node for the SAN Volume Controller.

S

SAN See *storage area network*.

SDD See *subsystem device driver (SDD)*.

secondary virtual disk
In Global Mirror, the virtual disk (VDisk) in a relationship that contains a copy of data written by the host application to the primary VDisk.

Simple Network Management Protocol (SNMP)
In the Internet suite of protocols, a network management protocol that is used to monitor routers and attached networks. SNMP is an application-layer protocol. Information on devices managed is defined and stored in the application's Management Information Base (MIB).

SNMP
See *Simple Network Management Protocol*.

storage area network (SAN)
A network whose primary purpose is the transfer of data between computer systems and storage elements and among storage elements. A SAN consists of a communication infrastructure, which provides physical connections, and a management layer, which organizes the connections, storage elements, and computer systems so that data transfer is secure and robust. (S)

subsystem device driver (SDD)
An IBM pseudo device driver designed to support the multipath configuration environments in IBM products.

switch A network infrastructure component to which multiple nodes attach.

Unlike hubs, switches typically have internal bandwidth that is a multiple of link bandwidth, and the ability to rapidly switch node connections from one to another. A typical switch can accommodate several simultaneous full link bandwidth transmissions between different pairs of nodes. (S) Contrast with *hub*.

symmetric virtualization
A virtualization technique in which the physical storage in the form of Redundant Array of Independent Disks (RAID) is split into smaller chunks of storage known as *extents*. These extents are then concatenated, using various policies, to make virtual disks (VDisks). See also *asymmetric virtualization*.

synchronized
In Global Mirror, the status condition that exists when both virtual disks (VDisks) of a pair that has a copy relationship contain the same data.

U

unconfigured mode
A mode in which I/O operations cannot be performed. See also *image mode* and *managed space mode*.

uninterruptible power supply
A device connected between a computer and its power source that protects the computer against blackouts, brownouts, and power surges. The uninterruptible power supply contains a power sensor to monitor the supply and a battery to provide power until an orderly shutdown of the system can be performed.

V

valid configuration
A configuration that is supported.

VDisk See *virtual disk*.

virtual disk (VDisk)
In SAN Volume Controller, a device that host systems attached to the storage area network (SAN) recognize as a Small Computer System Interface (SCSI) disk.

virtual storage area network (VSAN)
A fabric within the SAN.

virtualization
In the storage industry, a concept in which a pool of storage is created that contains several disk subsystems. The subsystems can be from various vendors. The pool can be split into virtual disks that are visible to the host systems that use them.

virtualized storage
Physical storage that has virtualization techniques applied to it by a virtualization engine.

| **VLUN** See *managed disk*.

VSAN See *virtual storage area network*.

W

worldwide node name (WWNN)
An identifier for an object that is globally unique. WWNNs are used by Fibre Channel and other standards.

worldwide port name (WWPN)
A unique 64-bit identifier associated with a fibre-channel adapter port. The WWPN is assigned in an implementation- and protocol-independent manner.

WWNN
See *worldwide node name*.

WWPN
See *worldwide port name*.

Index

Readers' Comments — We'd Like to Hear from You

IBM System Storage SAN Volume Controller
Planning Guide
Version 4.1.0

Publication No. GA32-0551-00

Overall, how satisfied are you with the information in this book?

	Very Satisfied	Satisfied	Neutral	Dissatisfied	Very Dissatisfied
Overall satisfaction	□	□	□	□	□

How satisfied are you that the information in this book is:

	Very Satisfied	Satisfied	Neutral	Dissatisfied	Very Dissatisfied
Accurate	□	□	□	□	□
Complete	□	□	□	□	□
Easy to find	□	□	□	□	□
Easy to understand	□	□	□	□	□
Well organized	□	□	□	□	□
Applicable to your tasks	□	□	□	□	□

Please tell us how we can improve this book:

Thank you for your responses. May we contact you? □ Yes □ No

When you send comments to IBM, you grant IBM a nonexclusive right to use or distribute your comments in any way it believes appropriate without incurring any obligation to you. IBM or any other organizations will only use the personal information that you supply to contact you about the issues that you state on this form.

Name

Address

Company or Organization

Phone No.

Readers' Comments — We'd Like to Hear from You

GA32-0551-00

Cut or Fold
Along Line

Fold and Tape

Please do not staple

Fold and Tape

NO POSTAGE
NECESSARY
IF MAILED IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST-CLASS MAIL PERMIT NO. 40 ARMONK, NEW YORK

POSTAGE WILL BE PAID BY ADDRESSEE

International Business Machines Corporation
Information Development
Department 61C
9032 South Rita Road
Tucson, Arizona
USA 85775-4401

Fold and Tape

Please do not staple

Fold and Tape

Cut or Fold
Along Line

Printed in USA

Spine information:

IBM

IBM System Storage SAN Volume
Controller

SAN Volume Controller Planning Guide

Version 4.1.0