EMC Fibre Channel Storage System CX-Series

CONFIGURATION PLANNING GUIDE

P/N 014003113
REV A02

EMC Corporation
171 South Street
Hopkinton, MA 01748-9103
Corporate Headquarters: (508) 435-1000, (800) 424-EMC2
Fax: (508) 435-5374 *Service*: (800) SVC-4EMC

Published September 2002

Trademark Information

Contents

Chapter 6 Storage-System Hardware

Chapter 7 Storage-System Management

Figures

Preface

This planning guide introduces EMC Fibre Channel CX-Series disk-array storage-system models and offers useful background information and worksheets to help you plan.

Please read this guide

- if you are considering purchase of a CX-Series disk-array storage system and want to understand its features; or
- before you plan the installation of a storage system.

Audience for the Manual

You should be familiar with the host servers that will use the storage systems and with the operating systems of the servers. After reading this guide, you will be able to

- determine the best storage-system components for your installation
- determine your site requirements
- configure storage systems correctly

Organization of the Manual

Chapter 1	Provides background information about Fibre Channel features and explains the major types of storage.
Chapter 2	Describes the RAID Groups and the different ways they store data.
Chapter 3	Describes the optional EMC MirrorView™ remote mirroring software.
Chapter 4	Describes the optional EMC SnapView™ software to produce clones and snapshots.
Chapter 5	Helps you plan your storage-system file systems and LUNs.
Chapter 6	Explains the hardware components of storage systems.
Chapter 7	Explains storage management software.

Conventions Used in The Manual

A note presents information that is important, but not hazard-related.

Where to Get Help

For questions about technical support and service, contact your service provider.

If you have a valid EMC service contract, contact EMC Customer Service at:

United States:	**(800) 782-4362 (SVC-4EMC)**
Canada:	**(800) 543-4782 (543-4SVC)**
Worldwide:	**(508) 497-7901**

Follow the voice menu prompts to open a service call, then select CLARiiON Product Support.

Sales and Customer Service Contacts

For the list of EMC sales locations, please access the EMC home page at:

`http://www.emc.com/contact/`

For additional information on the EMC products and services available to customers and partners, refer to the EMC Powerlink Web site at:

`http://powerlink.emc.com`

Your Comments

Your suggestions will help us continue to improve the accuracy, organization, and overall quality of the user publications. Please send a message to **techpub_comments@emc.com** with your opinions of this guide.

1

About CX-Series Storage Systems and Storage Networks

This chapter introduces Fibre Channel CX-Series disk-array storage systems and storage area networks (SANs). Major sections are

Introducing CX-Series Storage Systems

EMC Fibre Channel CX-Series disk-array storage systems provide terabytes of disk storage capacity, high transfer rates, flexible configurations, and highly available data at low cost. Their hardware RAID features are provided by one or two storage processors (SPs). For high availability, two SPs are required.

There are two types of CX-Series storage system: CX600 and CX400.

A CX600 system is based on a storage processor enclosure (called an SPE). An SPE does not support disks, but it supports up to 16 two-Gbit disk-array enclosures (DAE2s). Each DAE2 has slots for 15 disks. So a single SPE storage system can support up to 240 disks in the 16 DAE2 enclosures.

A CX400 system is based on a two-Gbit disk processor enclosure, (called a DPE2). A DPE2 supports up to 15 disks within itself and up to three separate DAE2s for a total of 60 disks.

A typical CX600 storage system is shown in the following figure.

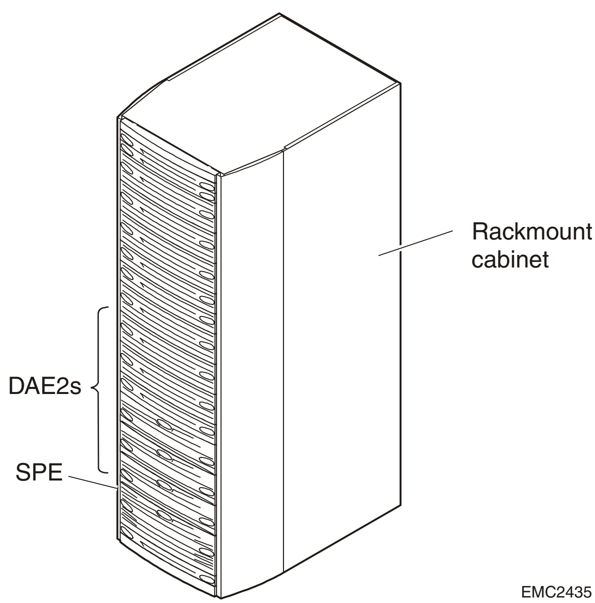


Figure 1-1 CX600 Storage Systems

A storage-system package includes storage management software, Fibre Channel interconnect hardware, and one or more storage systems. The host bus adapter driver hardware and software are available from outside vendors.

Fibre Channel Background

Fibre Channel is a high-performance serial protocol that allows transmission of both network and I/O channel data. It is a low level protocol, independent of data types, and supports such formats as SCSI and IP.

The Fibre Channel standard supports several physical topologies, including switch fabric point-to-point and arbitrated loop (FC-AL). The Fibre Channel storage systems described in this manual use switch fabric and FC-AL topologies.

A switch fabric is a set of point-to-point connections between nodes; each connection is made through one or more Fibre Channel switches. Each node may have its own unique address, but the path between nodes is governed by a switch. The nodes are connected by optical cable.

A Fibre Channel arbitrated loop is a circuit consisting of nodes. Each node has a unique address, called a Fibre Channel arbitrated loop address. The nodes connect with optical cables. An optical cable can transmit data over great distances for connections that span entire enterprises and can support remote disaster recovery systems.

Each connected device in a switched fabric or arbitrated loop is a server adapter (initiator) or a target (storage system). The switches are not considered nodes.

The following figure shows a node and initiator.

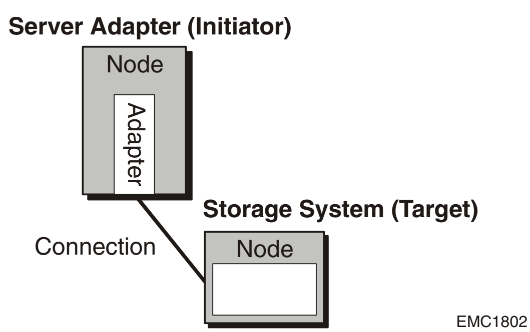


Figure 1-2 Nodes - Initiator and Target

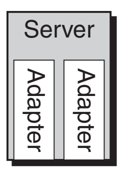
Fibre Channel Storage Components

A Fibre Channel storage system has three main components:

- Server component (host bus adapter driver with adapter and software)
- Interconnect components (cables based on Fibre Channel standards, and switches)
- Storage component (storage system with storage processors, power supply, cooling hardware, and disks)

Server Component (Host Bus Adapter and Driver)

The host bus adapter is a printed-circuit board that slides into an I/O slot in the server's cabinet. It transfers data between server memory and one or more disk-array storage systems over Fibre Channel — as controlled by the support software (adapter driver).



EMC1803

Interconnect Components

The interconnect components include the optical cables between components and any Fibre Channel switches.

The maximum length of optical cable between server and switch or storage system is 500 meters (1,640 feet) for 62.5-micron multimode cable or 10 kilometers (6.2 miles) for 50-micron single-mode cable. With extenders, optical cable can span up to 60 kilometers (36 miles) or more. This ability to span great distances is a major advantage of optical cable.

Details on cable lengths and rules appear later, in Chapter 6.

Fibre Channel Switches

A Fibre Channel switch, which is required for switched shared storage (a storage area network, SAN), connects all the nodes cabled to it using a fabric topology. A switch adds serviceability and scalability to any installation; it allows on-line insertion and removal of any device on the fabric and maintains integrity if any connected device stops participating. A switch also provides server-to-storage-system access control and point-to-point connections.

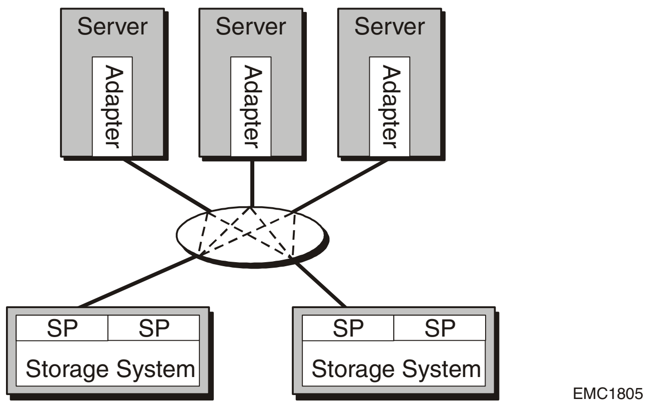


Figure 1-3 Switch and Point-to-Point Connections

You can cascade switches (connect one switch port to another switch) for additional port connections.

Switch Zoning

Switch zoning lets an administrator define paths between connected nodes based on the node's unique World Wide Name. Each zone includes a server adapter and/or one or more SPs. EMC recommends single-initiator zoning, which limits each zone to a single HBA. A switch can have as many zones as it has ports.

The current connection limits are four SP ports to one adapter port (the SPs fan in to the adapter) and 32 adapters to one SP port (the SP port fans out to the adapters).

In the following figure, Server 1 has access to one SP (SP A) in storage systems 1 and 2; it has no access to any other SP.

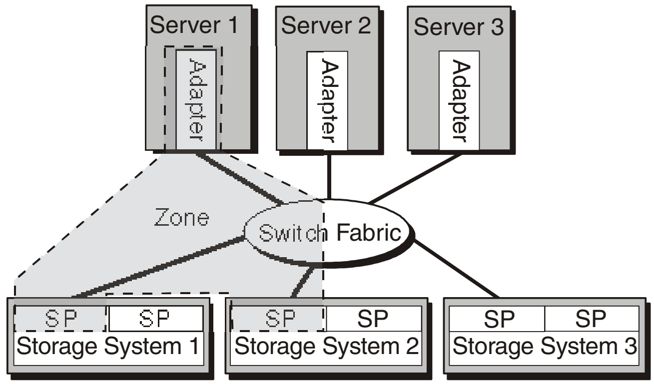


Figure 1-4 A Switch Zone

To illustrate switch zoning, Figure 1-4 shows just one HBA per server and one switch. Normally, such installations include multiple HBAs per server and two or more switches.

If you do not define a zone in a switch, all adapter ports connected to the switch can communicate with all SP ports connected to the switch. However, access to an SP does not necessarily provide access to the SP's storage. Access to storage is governed by the Storage Groups you create (explained later).

Fibre Channel switches are available with 8, 16, 32, or more ports. They are compact units that fit into a rackmount cabinets.

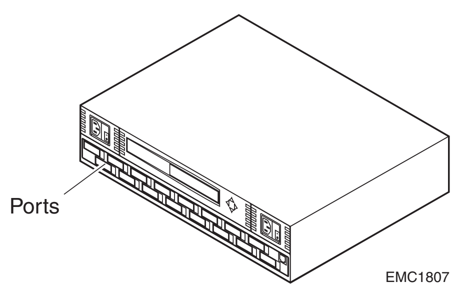


Figure 1-5 16-Port Switch, Back View

If your servers and storage systems will be far apart, you can place the switches closer to the servers or storage systems, as convenient.

A switch is technically a repeater, not a node, in a Fibre Channel loop. However, it is bound by the same cabling distance rules as a node.

Storage Component (Storage System Enclosures)

EMC CX-Series disk-array storage systems, with their storage processor (SPs), power supplies, and cooling hardware form the storage component of a Fibre Channel system. The controlling unit is a CX600 storage processor enclosure (SPE) or CX400 disk processor enclosure (DPE2). A CX600 SPE, outside of its cabinet, looks like the following figure.

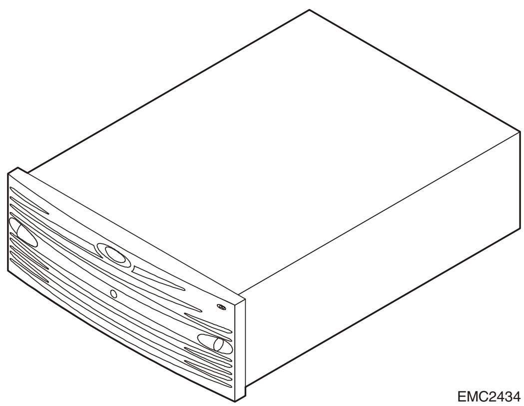


Figure 1-6 CX600 Storage Processor Enclosure (SPE)

Each CX600 and CX400 has two SPs. A CX600 SP has four front-end ports, and a CX400 two front end ports, to communicate with switches or servers and two back-end loops to communicate with disks. Hardware details appear later in Chapter 6.

Types of Storage-System Installations

You can use a storage system in any of several types of installation:

- **Unshared direct** with one server is the simplest and least costly.
- **Shared-or-clustered direct**, with a limit of two servers, lets two servers share storage resources with high availability.
- **Shared switched**, with two switch fabrics, lets two or more servers share the resources of several storage systems in a storage area network (SAN). Shared switched installations are available in high-availability versions (two HBAs per server) or with one HBA per server. Shared switched storage systems can have multiple paths to each SP, providing multipath I/O for dynamic load sharing and greater throughput.

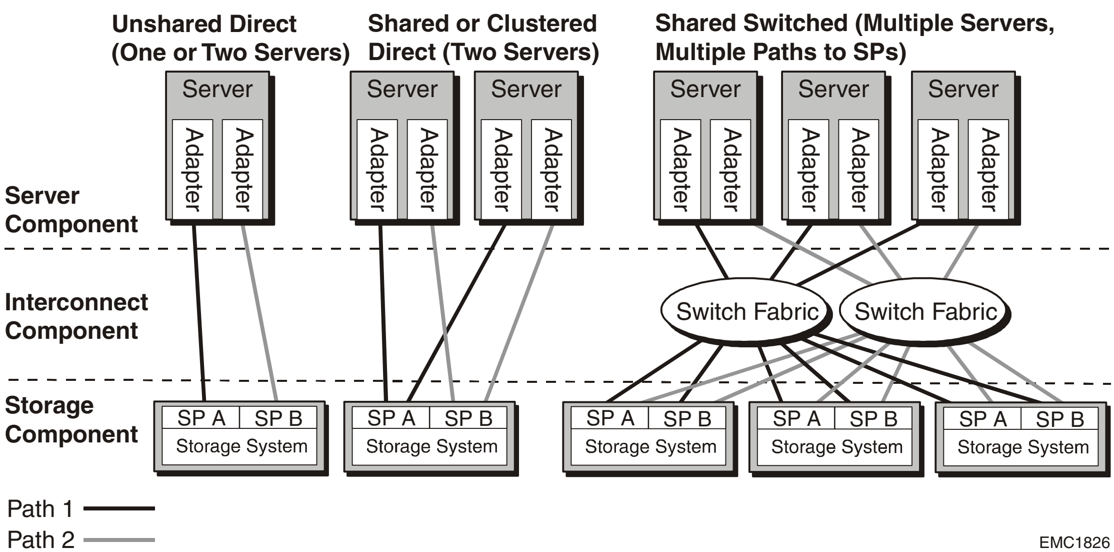


Figure 1-7 Types of Storage-System Installation

Storage systems for any shared installation require EMC Access Logix™ software to control server access to the storage-system LUNs.

The Shared-or-clustered direct installation can be either shared (that is, use Access Logix to control LUN access) or clustered (without Access Logix, but with operating system cluster software controlling LUN access).

About Switched Shared Storage and SANs (Storage Area Networks)

This section explains the features that let multiple servers share disk-array storage systems on a SAN (storage area network).

A SAN is one or more storage devices connected to servers through Fibre Channel switches to provide a central location for disk storage. Centralizing disk storage among multiple servers has many advantages, including

- highly available data
- flexible association between servers and storage capacity
- centralized management for fast, effective response to users' data storage needs
- easier file backup and recovery

An EMC SAN is based on shared storage; that is, the SAN requires EMC Access Logix software to provide flexible access control to storage-system LUNs. Within the SAN, a network connection to each SP in the storage system lets you configure and manage it.

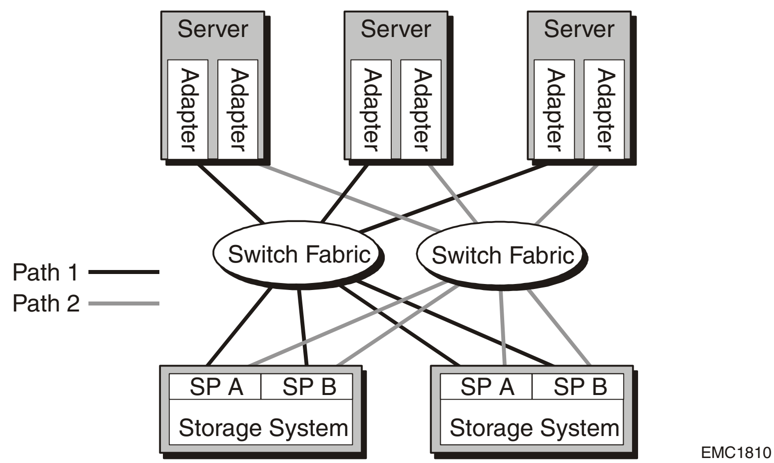


Figure 1-8 Components of a SAN

Fibre Channel switches can control data access to storage systems through the use of switch zoning, explained earlier on page 1-5.

However, switch zoning cannot selectively control data access to LUNs in a storage system, because each SP appears as a single Fibre Channel device to the switch fabric. So switch zoning can prevent or allow communication with an SP, but not with specific disks or LUNs attached to an SP. For access control with LUNs, a different solution is required: Storage Groups.

Storage Groups

A Storage Group is one or more LUNs (logical units) within a storage system that is reserved for one or more servers and is inaccessible to other servers. Storage Groups are the central component of shared storage; storage systems that are unshared do not use Storage Groups.

When you configure shared storage, you specify servers and the Storage Group(s) each server can read from and/or write to. Acess Logix software running in each storage system enforces the server-to-Storage Group permissions.

More than one server can access a Storage Group if all the servers run cluster software. The cluster software enforces orderly access to the shared Storage Group LUNs.

The following figure shows a simple shared storage configuration consisting of one storage system with two Storage Groups. One Storage Group serves a cluster of two servers running the same operating system, and the other Storage Group serves a UNIX® database server. Each server is configured with two independent paths to its data, including separate host bus adapters, switches, and SPs, so there is no single point of failure for access to its data.

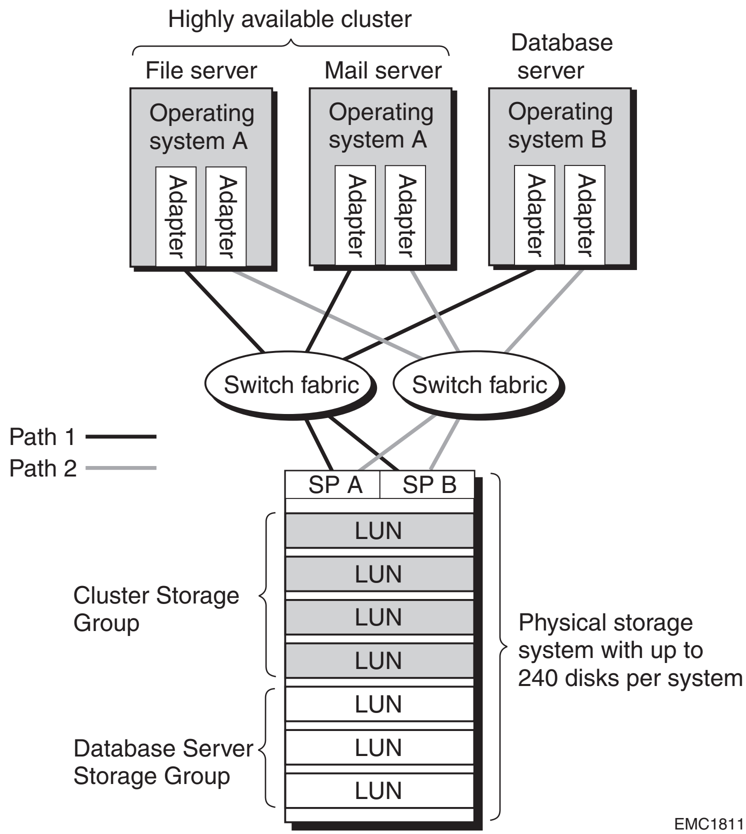


Figure 1-9 Sample Shared Storage Configuration

Access Control with Shared Storage

Access control permits or restricts a server's access to shared storage. Configuration access, the ability to configure storage systems, is governed by the storage system software security mechanism: a set of privileges based on username and password that's stored, encrypted, on one or more storage systems in a network.

Data access, the ability to read and write information to storage-system LUNs, is provided by Storage Groups. During storage-system configuration, using a management utility, the system administrator associates a server with one or more LUNs. The associated LUNs compose a Storage Group.

Each server sees its Storage Group as if it were an entire storage system, and never sees the other LUNs on the storage system. Therefore, it cannot access or modify data on LUNs that are not part of its Storage Group. However, you can define a Storage Group as sharable (accessible by more than one server), if, as shown above in Figure 1-9, the servers run cluster software.

The following figure shows access control through Storage Groups. Each server has exclusive read and write access to its designated Storage Group.

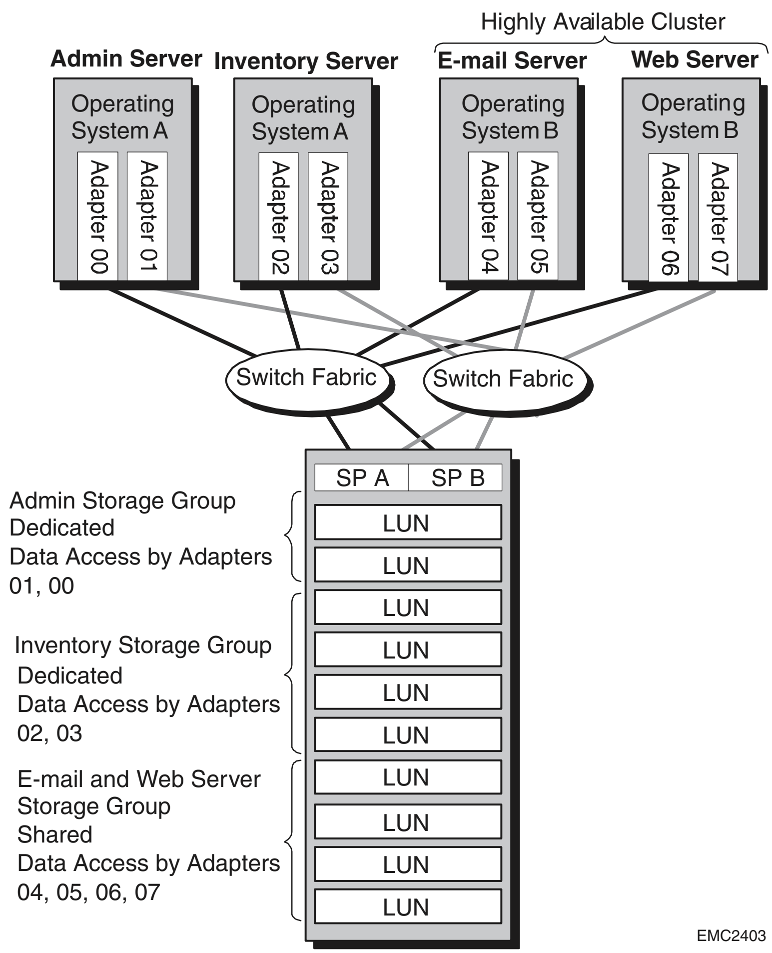


Figure 1-10 Data Access Control with Shared Storage

What Next?

For information about RAID types and RAID tradeoffs, continue to the next chapter.

For information on the MirrorView™ or SnapView™ software options, go to Chapter 3 or 4.

To plan LUNs and file systems, skip to Chapter 5.

For details on the storage-system hardware, skip to Chapter 6.

2

RAID Types and Tradeoffs

This chapter explains the RAID types you can choose for your storage-system LUNs. If you already know about RAID types and know which ones you want, you can skip this background information and go to the LUN planning chapter (Chapter 5). Topics are

Introducing RAID

The storage system uses RAID (redundant array of independent disks) technology. RAID technology groups separate disks into one logical unit (LUN) to improve reliability and/or performance.

The storage system supports five RAID levels and two other disk configurations, the individual unit and the hot spare (global spare). You group the disks into one RAID Group by *binding* them using a storage-system management utility.

Four of the RAID levels use *disk striping* and two use *mirroring*.

Disk Striping

Using disk stripes, the storage-system hardware can read from and write to multiple disks simultaneously and independently. By allowing several read/write heads to work on the same task at once, disk striping can enhance performance. The amount of information read from or written to each disk makes up the stripe element size. The stripe size is the stripe element size multiplied by the number of disks in a group. For example, assume a stripe element size of 128 sectors (the default). If the group has five disks, you would multiply five by the stripe element size of 128 to yield a stripe size of 640 sectors.

The storage system uses disk striping with most RAID types.

Mirroring

Mirroring maintains a copy of a logical disk *image* that provides continuous access if the original image becomes inaccessible. The system and user applications continue running on the good image without interruption. There are two kinds of mirroring: hardware mirroring, in which the SP synchronizes the disk images; and software mirroring, in which the operating system synchronizes the images. Software mirroring consumes server resources, since the operating system must mirror the images, and has no offsetting advantages; we mention it here only for historical completeness.

With a storage system, you can create a hardware mirror by binding disks as a RAID 1 mirrored pair or a RAID 1/0 Group (a mirrored RAID 0 Group); the hardware will then mirror the disks automatically.

With a LUN of any RAID type, a storage system can maintain a remote copy using the optional MirrorView software. MirrorView remote mirroring, primarily useful for disaster recovery, is explained in Chapter 3.

RAID Groups and LUNs

Some RAID types let you create multiple LUNs on one RAID Group. You can then allot each LUN to a different user, server, or application. For example, a five-disk RAID 5 Group that uses 73-Gbyte disks offers 292 Gbytes of space. You could bind three LUNs, with 60, 100, and 132 Gbytes of storage capacity, for temporary, mail, and customer files.

One disadvantage of multiple LUNs on a RAID Group is that I/O to each LUN may affect I/O to the others in the group; that is, if traffic to one LUN is very heavy, I/O performance with other LUNs may be degraded. The main advantage of multiple LUNs per RAID Group is the ability to divide the enormous amount of disk space provided by RAID Groups on newer, high-capacity disks.

The following figure shows three LUNs in one RAID Group.

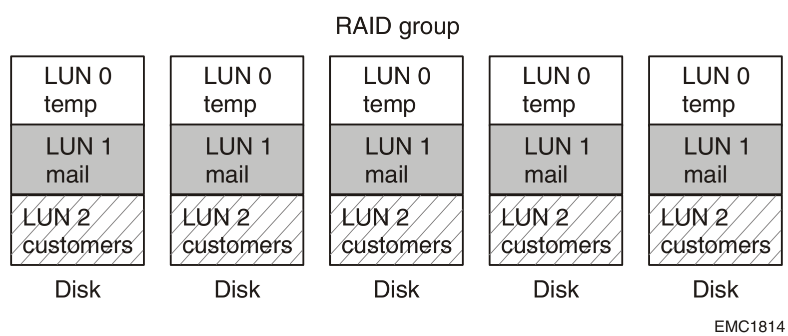


Figure 2-1 Multiple LUNs in a RAID Group

RAID Types

You can choose from the following RAID types: RAID 5, RAID 3, RAID 1, RAID 0, RAID 1/0, individual disk unit, and hot spare.

You can choose an additional type of redundant disk — a remote mirror — for any RAID type except a hot spare.

RAID 5 Group (Individual Access Array)

A RAID 5 Group usually consists of five disks (but can have three to sixteen). A RAID 5 Group uses disk striping. With a RAID 5 group, you can create up to 32 RAID 5 LUNs to apportion disk space to different users, servers, and applications.

The storage system writes parity information that lets the Group continue operating if a disk fails. When you replace the failed disk, the SP rebuilds the group using the information stored on the working disks. Performance is degraded while the SP rebuilds the group. However, the storage system continues to function and gives users access to all data, including data stored on the failed disk.

The following figure shows user and parity data with the default stripe element size of 128 sectors (65,536 bytes) in a five-disk RAID 5 group. The stripe size comprises all stripe elements. Notice that the disk block addresses in the stripe proceed sequentially from the first disk to the second, third, and fourth, then back to the first, and so on.

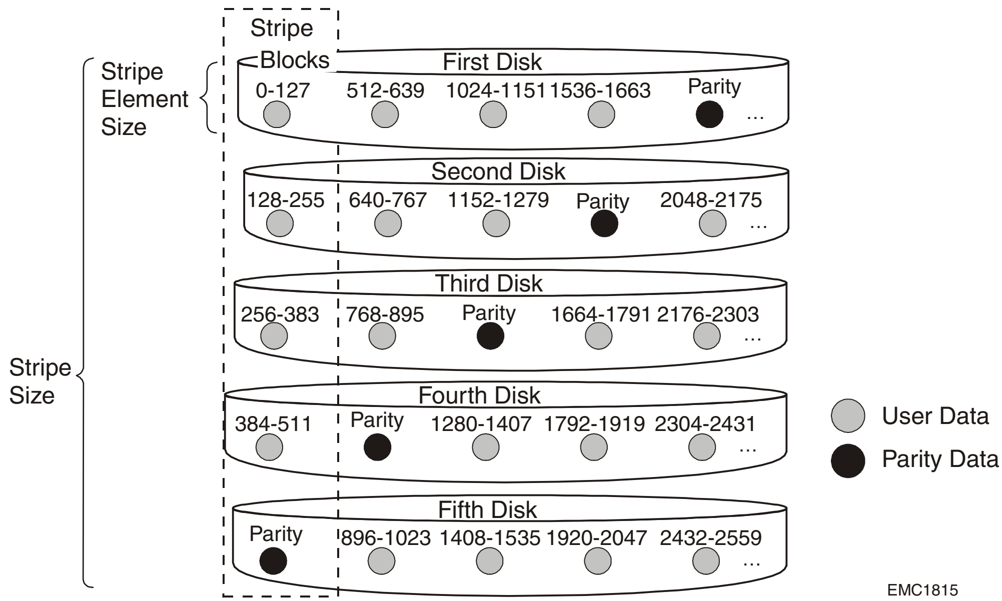


Figure 2-2 RAID 5 Group

RAID 5 Groups offer excellent read performance and good write performance. Write performance benefits greatly from storage-system caching.

RAID 3 Group (Parallel Access Array)

A RAID 3 Group consists of five or nine disks. The hardware always reads from or writes to all the disks. A RAID 3 Group uses disk striping. To maintain the RAID 3 performance, you can create only one LUN per RAID 3 group.

The storage system writes parity information that lets the Group continue operating if a disk fails. When you replace the failed disk, the SP rebuilds the group using the information stored on the working disks. Performance is degraded while the SP rebuilds the group. However, the storage system continues to function and gives users access to all data, including data stored on the failed disk.

The following figure shows user and parity data with a data block size of 2 Kbytes in a RAID 3 Group. Notice that the byte addresses proceed from the first disk to the second, third, and fourth, then the first, and so on.

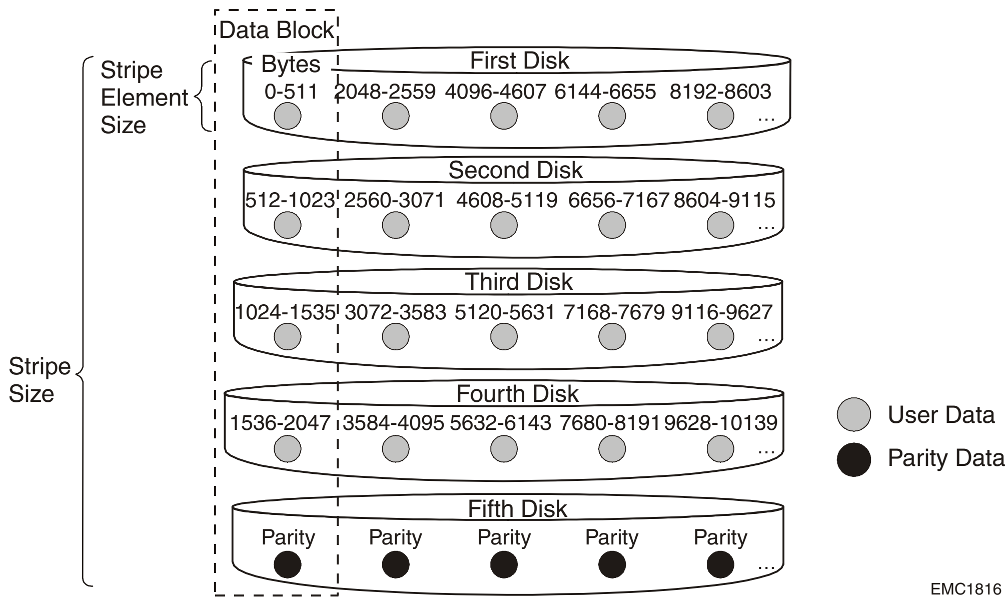


Figure 2-3 RAID 3 Group

RAID 3 differs from RAID 5 in several important ways. First, in a RAID 3 Group the hardware processes disk requests serially; whereas in a RAID 5 Group the hardware can interleave disk requests. Second, with a RAID 3 Group, the parity information is stored on one disk; with a RAID 5 Group, it is stored on all disks. Finally, with a RAID 3 Group, the I/O occurs in small units (one sector) to each disk. A RAID 3 Group works well for single-task applications that use I/Os of blocks larger than 64 Kbytes.

Each RAID 3 Group requires some dedicated SP memory (6 Mbytes recommended per group). This memory is allocated when you create the group, and becomes unavailable for storage-system caching. For top performance, we suggest that you do not use RAID 3 Groups with RAID 5, RAID 1/0, or RAID 0 Groups, since SP processing power and memory are best devoted to the RAID 3 Groups. RAID 1 mirrored pairs and individual units require less SP processing power, and therefore work well with RAID 3 Groups.

RAID 1 Mirrored Pair

A RAID 1 Group consists of two disks that are mirrored automatically by the storage-system hardware. With a RAID 1 Group, you can create multiple RAID 1 LUNs to apportion disk space to different users, servers, and applications.

RAID 1 hardware mirroring within the storage system is not the same as software mirroring, remote mirroring, or hardware mirroring for other kinds of disks. Functionally, the difference is that you cannot manually stop mirroring on a RAID 1 mirrored pair, and then access one of the images independently. If you want to use one of the disks in such a mirror separately, you must unbind the mirror (losing all data on it), rebind the disk as the type you want, and software format the newly bound LUN.

With a storage system, RAID 1 hardware mirroring has the following advantages:

- automatic operation (you do not have to issue commands to initiate it)
- physical duplication of images
- a rebuild period that you can select during which the SP recreates the second image after a failure

With a RAID 1 mirrored pair, the storage system writes the same data to both disks, as follows.

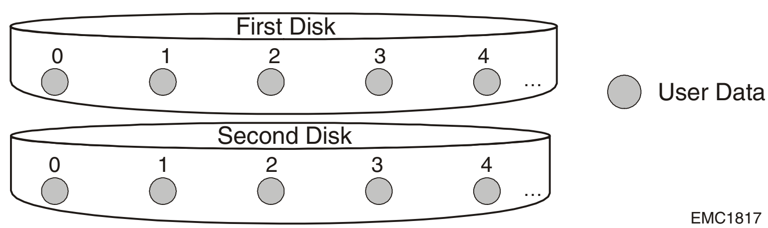


Figure 2-4 RAID 1 Mirrored Pair

RAID 0 Group (Nonresident Array)

A RAID 0 Group consists of three to a maximum of sixteen disks. A RAID 0 Group uses disk striping, in which the hardware writes to or reads from multiple disks simultaneously. You can create up to 32 LUNs per RAID 0 Group.

Unlike the other RAID levels, with RAID 0 the hardware does not maintain parity information on any disk; this type of group has no inherent data redundancy. RAID 0 offers enhanced performance through simultaneous I/O to different disks.

If the operating system supports software mirroring, you can use software mirroring with the RAID 0 Group to provide high availability. A desirable alternative to RAID 0 is RAID 1/0.

RAID 1/0 Group (Mirrored RAID 0 Group)

A RAID 1/0 Group consists of four, six, eight, ten, twelve, fourteen, or sixteen disks. These disks make up two mirror images, with each image including two to eight disks. The hardware automatically mirrors the disks. A RAID 1/0 Group uses disk striping. It combines the speed advantage of RAID 0 with the redundancy advantage of mirroring. With a RAID 1/0 Group, you can create up to 32 RAID 1/0 LUNs to apportion disk space to different users, servers, and applications.

The following figure shows the distribution of user data with the default stripe element size of 128 sectors (65,536 bytes) in a six-disk RAID 1/0 Group. Notice that the disk block addresses in the stripe proceed sequentially from the first mirrored disks (first and fourth disks) to the second mirrored disks (second and fifth disks), to the third mirrored disks (third and sixth disks), and then from the first mirrored disks, and so on.

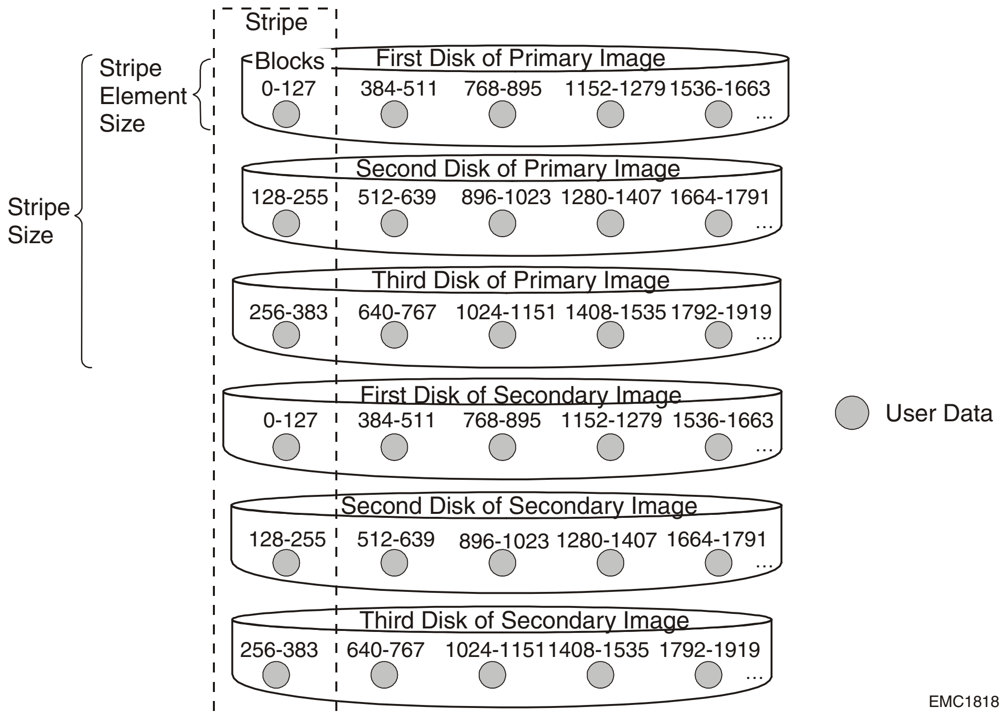


Figure 2-5 RAID 1/0 Group

A RAID 1/0 Group can survive the failure of multiple disks, providing that one disk in each image pair survives.

Individual Disk Unit

An individual disk unit is a disk bound to be independent of any other disk in the cabinet. An individual unit has no inherent high availability, but you can make it highly available by using software mirroring with another individual unit. You can create one LUN per individual disk unit. If you want to apportion the disk space, you can do so using partitions, file systems, or user directories.

Hot Spare

A hot spare is a dedicated replacement disk on which users cannot store information. A hot spare is global: if any disk in a RAID 5 Group, RAID 3 Group, RAID 1 mirrored pair, or RAID 1/0 Group fails, the SP automatically rebuilds the failed disk's structure on the hot spare. When the SP finishes rebuilding, the disk group functions as usual, using the hot spare instead of the failed disk. When you replace the failed disk, the SP copies the data from the former hot spare onto the replacement disk.

When the copy is done, the disk group consists of disks in the original slots, and the SP automatically frees the hot spare to serve as a hot spare again. A hot spare is most useful when you need the highest data availability. It eliminates the time and effort needed for someone to notice that a disk has failed, find a suitable replacement disk, and insert the disk.

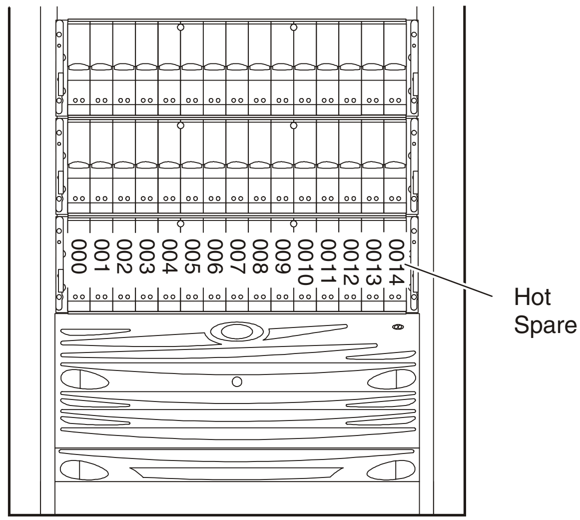
When you plan to use a hot spare, make sure the disk has the capacity to serve in any RAID Group in the storage-system chassis. A RAID Group cannot use a hot spare that is smaller than a failed disk in the group.

You can have one or more hot spares per storage-system chassis. You can make any disk in the chassis a hot spare, except a disk that stores Base Software or the write cache vault. That is, a hot spare can be any disk except the following:

SPE system without write caching:	disk IDs 000-002
SPE system with write caching:	disk IDs 000-004

If you use hot spares of different sizes, the storage system will automatically use the hot spare of the proper size in place of a failed disk.

An example of hot spare usage for a CX600 storage system follows.



1. RAID 5 Groups consist of disk modules 0-4 and 5-9; mirrored pairs are modules 10-11 and 12-13, disk module 14 is a hot spare.
2. Disk module 3 fails
3. RAID 5 Group becomes modules 0, 1, 2, 14, and 4; now no hot spare is availale.
4. System operator replaces failed midule 3 with a functional module.
5. Once again, RAID 5 Group consists of modules 0-4 and the hot spare is 14.

EMC2432

Figure 2-6 How a Hot Spare Works

RAID Benefits and Tradeoffs

This section reviews RAID types and explains their benefits and tradeoffs. You can create seven types of LUN:

- RAID 5 Group (individual access array)
- RAID 3 Group (parallel access array)
- RAID 1 mirrored pair
- RAID 1/0 Group (mirrored RAID 0 Group); a RAID 0 Group mirrored by the storage-system hardware
- RAID 0 Group (nonredundant individual access array); no inherent high-availability features
- Individual unit; no inherent high-availability features
- Hot spare; serves only as an automatic replacement for any disk in a RAID type other than 0; does not store data during normal system operations

Plan the disk unit configurations carefully. After a disk has been bound into a LUN, you cannot change the RAID type of that LUN without unbinding it, and this means losing all data on it.

The following table compares the read and write performance and relative cost per gigaabyte (Gbyte) of the RAID types. Figures shown are theoretical maximums.

Table 2-1 Performance, Availability, and Cost of RAID Types (Individual Unit = 1.0)

Disk Configuration	Relative Read Performance Without Cache	Relative Write Performance Without Cache	Relative Cost Per Gbyte
RAID 5 Group with five disks	Up to 5x with five disks (for small I/O requests that are 2 to 8 Kbytes)	Up to 1.25x with five disks (for small I/O requests that are 2 to 8 Kbytes)	1.25
RAID 3 Group with five disks	Up to 4x (for large I/O requests)	Up to 4x (for large I/O requests)	1.25
RAID 1 mirrored pair	Up to 2x	Up to 1x	2
RAID 1/0 Group with 10 disks	Up to 10x	Up to 5x	2
Individual unit	1x	1x	1
Notes: These performance numbers are not based on storage-system caching. With caching, the performance numbers for RAID 5 writes improve significantly. Performance multipliers vary with the load on yjr server and storage system.			

Performance

RAID 5, with individual access, provides high read throughput by allowing simultaneous reads from each disk in the group. RAID 5 write performance is excellent when the storage system uses write caching.

RAID 3, with parallel access, provides high throughput for sequential, large block-size requests (blocks of more than 64 Kbytes). With RAID 3, the system accesses all five disks in each request but need not read data and parity before writing – advantageous for large requests but not for small ones. RAID 3 uses SP memory without caching, which means you do not need the second SP and BBU that caching requires.

Generally, the performance of a RAID 3 Group increases as the size of the I/O request increases. Read performance increases rapidly with read requests up to 1 Mbyte. Write performance increases greatly for sequential write requests that are greater than 256 Kbytes. For applications issuing very large I/O requests, a RAID 3 LUN provides significantly better write performance than a RAID 5 LUN.

We do not recommend using RAID 3 in the same storage-system chassis with RAID 5 or RAID 1/0.

A RAID 1 mirrored pair has its disks locked in synchronization, but the SP can read data from the disk whose read/write heads are closer to it. Therefore, RAID 1 read performance can be twice that of an individual disk while write performance remains the same as that of an individual disk.

A RAID 0 Group (nonredundant individual access array) or RAID 1/0 Group (mirrored RAID 0 Group) can have as many I/O operations occurring simultaneously as there are disks in the group. Since RAID 1/0 locks pairs of RAID 0 disks the same way as RAID 1 does, the performance of RAID 1/0 equals the number of disk pairs times the RAID 1 performance number. If you want high throughput for a specific LUN, use a RAID 1/0 or RAID 0 Group. A RAID 1/0 Group requires at least four disks; a RAID 0 Group, at least three disks.

An individual unit has only one I/O operation per read or write operation.

RAID types 5, 1, 1/0, and 0 allow multiple LUNs per RAID Group. If you create multiple LUNs on a RAID Group, the LUNs share the RAID Group disks, and the I/O demands of each LUN affect the I/O service time to the other LUNs. For best performance, you may want to use one LUN per RAID Group.

Storage Flexibility

Certain RAID Group types — RAID 5, RAID 1, RAID 1/0, and RAID 0 — let you create up to 32 LUNs in each group. This adds flexibility, particularly with large disks, since it lets you apportion LUNs of various sizes to different servers, applications, and users. Conversely, with RAID 3, there can be only one LUN per RAID Group, and the group must include five or nine disks — a sizable block of storage to devote to one server, application, or user. However, the nature of RAID 3 makes it ideal for that single-threaded type of application.

Data Availability and Disk Space Usage

If data availability is critical and you cannot afford to wait hours to replace a disk, rebind it, make it accessible to the operating system, and load its information from backup, then use a redundant RAID Group: RAID 5, RAID 3, RAID 1 mirrored pair, or RAID 1/0. If data availability is not critical, or disk space usage is critical, bind an individual unit.

A RAID 1 mirrored pair or RAID 1/0 Group provides very high data availability. They are more expensive than RAID 5 or RAID 3 Groups, since only 50 percent of the total disk capacity is available for user data, as shown on page 2-13.

A RAID 5 or RAID 3 Group provides high data availability, but requires more disks than a mirrored pair. In a RAID 5 or RAID 3 Group of five disks, 80 percent of the disk space is available for user data. So RAID 5 and RAID 3 Groups use disk space much more efficiently than a mirrored pair. A RAID 5 or RAID 3 Group is usually more suitable than a RAID 1 mirrored pair for applications where high data availability, good performance, and efficient disk space usage are all of relatively equal importance.

For a LUN in any RAID Group, you can provide for disaster recovery by establishing a remote mirror at a distant site.

The following figure illustrates RAID type disk space usage.

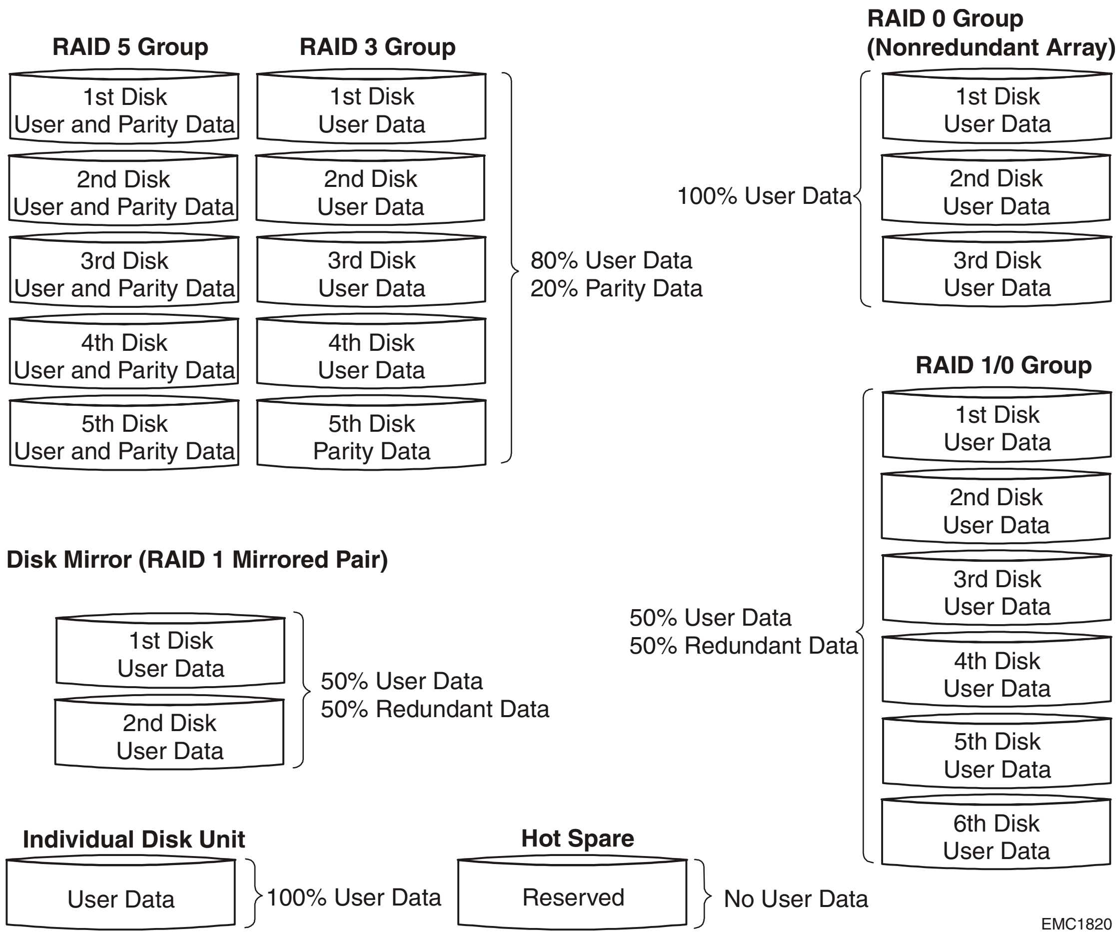


Figure 2-7 Disk Space Usage in the RAID Configuration

A RAID 0 Group (nonredundant individual access array) provides all its disk space for user files, but does not provide any high availability features. For high availability, you can use a RAID 1/0 Group instead.

A RAID 1/0 Group provides the best combination of performance and availability, at the highest cost per Gbyte of disk space.

An individual unit, like a RAID 0 Group, provides no high-availability features. All its disk space is available for user data, as shown in the figure above.

Guidelines for RAID Groups

To decide when to use a RAID 5 Group, RAID 3 Group, mirror (that is, a RAID 1 mirrored pair or RAID 1/0 Group, a RAID 0 Group, individual disk unit, or hot spare), you need to weigh these factors:

- Importance of data availability
- Importance of performance
- Amount of data stored
- Cost of disk space

The following guidelines will help you decide on RAID types.

Use a RAID 5 Group (individual access array) for applications where

- Data availability is very important.
- Large volumes of data will be stored.
- Multitask applications use I/O transfers of different sizes.
- Excellent read and good write performance is needed (write performance is very good with write caching).
- You want the flexibility of multiple LUNs per RAID Group.

Use a RAID 3 Group (parallel access array) for applications where

- Data availability is very important.
- Large volumes of data will be stored.
- A single-task application uses large I/O transfers (more than 64 Kbytes). The operating system must allow transfers aligned to start at disk addresses that are multiples of 2 Kbytes from the start of the LUN.

Use a RAID 1 mirrored pair for applications where

- Data availability is very important.
- Speed of write access is important and write activity is heavy.

Use a RAID 1/0 Group (mirrored nonredundant array) for applications where

- Data availability is critically important.
- Overall performance is very important.

Use a RAID 0 Group (nonredundant individual access array) for applications where

- High availability is not important.
- You can afford to lose access to all data stored on a LUN if a single disk fails.
- Overall performance is very important.

Use an individual unit for applications where

- High availability is not important.
- Speed of write access is somewhat important.

Use a hot spare where

- In any RAID 5, RAID 3, RAID 1/0 or RAID 1 Group, high availability is so important that you want to regain data redundancy quickly without human intervention if any disk in the group fails.
- Minimizing the degraded performance caused by disk failure in a RAID 5 or RAID 3 Group is important.

Sample Applications for RAID Types

This section describes some types of applications in which you would want to use a RAID 5 Group, RAID 3 Group, RAID 1 mirrored pair, RAID 0 Group (nonredundant array), RAID 1/0 Group, or individual unit.

RAID 5 Group (individual access array) — Useful as a database repository or a database server that uses a normal or low percentage of write operations (writes are 33 percent or less of all I/O operations). Use a RAID 5 Group where multitasking applications perform I/O transfers of different sizes. Write caching can significantly enhance the write performance of a RAID 5 Group.

For example, a RAID 5 Group is suitable for multitasking applications that require a large history database with a high read rate, such as a database of legal cases, medical records, or census information. A RAID 5 Group also works well with transaction processing applications, such as an airline reservations system, where users typically read the information about several available flights before making a reservation, which requires a write operation. You could also use a RAID 5 Group in a retail environment, such as a supermarket, to hold the price information accessed by the point-of-sale terminals. Even though the price information may be updated daily, requiring many write operations, it is read many more times during the day.

RAID 3 Group — A RAID 3 Group (parallel access array) works well with a single-task application that uses large I/O transfers (more than 64 Kbytes), aligned to start at a disk address that is a multiple of 2 Kbytes from the beginning of the logical disk. RAID 3 Groups can use SP memory to great advantage without the second SP and battery backup unit required for storage-system caching.

You might use a RAID 3 Group for a single-task application that does large I/O transfers, like a weather tracking system, geologic charting application, medical imaging system, or video storage application.

RAID 1 mirrored pair — A RAID 1 mirrored pair is useful for logging or record-keeping applications because it requires fewer disks than a RAID 0 Group (nonredundant array) and provides high availability and fast write access. Or you could use it to store daily updates to a database that resides on a RAID 5 Group, and then, during off-peak hours, copy the updates to the database on the RAID 5 Group.

RAID 0 Group (nonredundant individual access array) — Use a RAID 0 Group where the best overall performance is important. In terms of high availability, a RAID 0 Group is less available than an individual unit. A RAID 0 Group (like a RAID 5 Group) requires a minimum of three disks. A RAID 0 Group is useful for applications using short-term data to which you need quick access.

RAID 1/0 Group (mirrored RAID 0 Group) — A RAID 1/0 Group provides the best balance of performance and availability. You can use it very effectively for any of the RAID 5 applications. A RAID 1/0 Group requires a minimum of four disks.

Individual unit — An individual unit is useful for print spooling, user file exchange areas, or other such applications, where high availability is not important or where the information stored is easily restorable from backup.

The performance of an individual unit is slightly less than a standard disk not in an storage system. The slight degradation results from SP overhead.

Hot spare — A hot spare provides no data storage but enhances the availability of each RAID 5, RAID 3, RAID 1, and RAID 1/0 Group in a storage system. Use a hot spare where you must regain high availability quickly without human intervention if any disk in such a RAID Group fails. A hot spare also minimizes the period of degraded performance after a RAID 5 or RAID 3 disk fails.

What Next?

To plan LUNs and file systems, skip to Chapter 5. For details on the storage-system hardware, skip to Chapter 6.

3

About MirrorView Remote Mirroring Software

This chapter introduces EMC MirrorView software — mirroring software that works on Fibre Channel CX-Series disk-array storage systems to create a byte-for-byte copy of one or more local LUNs connected to a distant computer system.

Topics are

What Is EMC MirrorView Software?

EMC MirrorView is a software application that maintains a copy image of a logical unit (LUN) at separate locations. The images are far enough apart to provide for disaster recovery; that is, to let one image continue if a serious accident or natural disaster disables the other.

The production image (the one mirrored) is called the primary image; the copy image is called the secondary image. The primary image is connected to a server called the production host. The secondary image is maintained by a separate storage system that can be a stand-alone storage system or connected to its own server. Both storage systems are managed by the same management station, which can promote the secondary image if the primary image becomes inaccessible.

The following figure shows two sites and a primary and secondary image that includes the database of three LUNs. Notice that the storage-system SP As and SP Bs are connected.

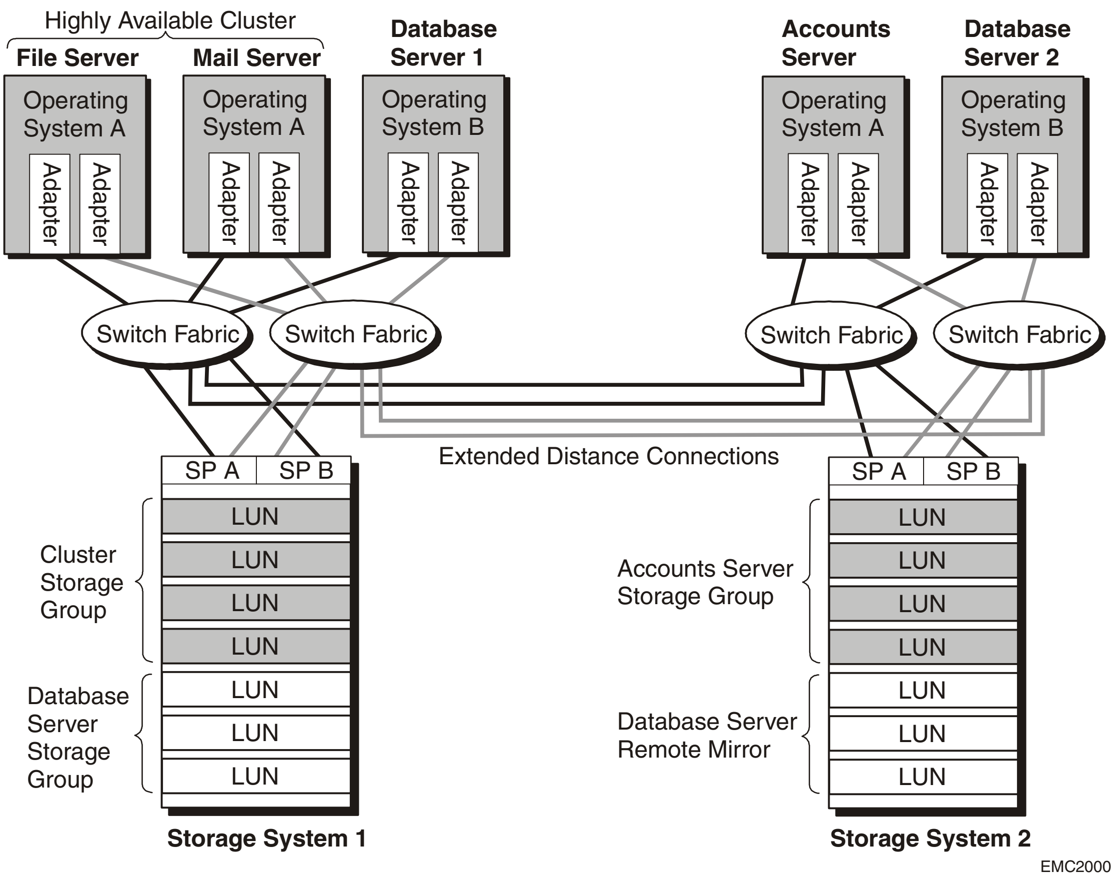


Figure 3-1 Sites with MirrorView Primary and Secondary Images

If the connections include extender boxes, then the distance between storage systems can be up to the maximum supported by the extender — generally 60 kilometers.

MirrorView can also connect to CNT UltraNet™ Edge Storage Router to interface with an IP network. Each storage system requires one router.

Without extender boxes or routers, the maximum distance is 10 kilometers.

MirrorView Features and Benefits

MirrorView mirroring adds value to customer systems by offering the following features:

- Provision for disaster recovery with minimal overhead
- Local high availability
- Bidirectional (cross) mirroring
- Integration with EMC SnapView LUN snapshot copy software

Provision for Disaster Recovery with Minimal Overhead

Provision for disaster recovery is the major benefit of MirrorView mirroring. Destruction of the primary data site would cripple or ruin many organizations. MirrorView lets data processing operations resume very quickly, often within minutes or a few hours.

MirrorView is transparent to servers and their applications. Server applications do not know that a LUN is mirrored, and the effect on performance depends on the distance and application workload.

MirrorView supports up to four connections per storage system, which means you can use it in a central backup configuration, where one storage system maintains remote mirrors for as many as four other storage systems.

MirrorView uses synchronous writes, which means that server writes are acknowledged only after all secondary storage systems commit the data. Most disaster recovery systems sold today use this type of mirroring.

MirrorView runs in storage systems, not on servers, therefore it uses no host I/O or CPU resources. The additional processing for mirroring is performed on storage systems.

Local High Availability

MirrorView operates in a highly available environment. TEach host has two host bus adapters (HBAs) and teach storage system has two SPs. If a single adapter or SP fails, the path in the surviving SP can take control of (*trespass*) any LUNs owned by the failed adapter or SP. The high-availability features of RAID protect against disk failure.

Mirrors are resilient to an SP failure in the primary or secondary storage system.

Bidirectional Mirroring

The primary or secondary role applies to just one remote mirror. A storage system can maintain the primary image for one mirror and a secondary image for another mirror. This allows the use of server resources at both sites while maintaining duplicate copies of all data at both sites.

Integration with EMC SnapView Software

EMC SnapView software allows users to create up to eight snapshot copies of an active LUN at any point in time. Each snapshot copy is a consistent image that can serve as backup while I/O continues to the original LUN. You can use SnapView in conjunction with MirrorView to make backup copies at a remote site.

A common situation for disaster recovery is to have a primary and a secondary site that are geographically separate. MirrorView ensures that the data from the primary site replicates to the secondary site. The secondary site sits idle until there is a failure of the primary site. With the addition of SnapView at the secondary site, the secondary site can take snapshot copies of the replicated images and back them up to other media, providing time-of-day snapshots of data on the production host with minimal overhead.

How MirrorView Handles Failures

When a failure occurs during normal operations, MirrorView implements several actions to recover.

Primary Image Failure

When access to the primary image fails, access to the mirror stops until you promote a secondary to primary or until the primary is repaired. If a secondary image is promoted, the old primary typically requires full synchronization before it can rejoin the mirror and be qualify as a secondary for future recovery. If the primary is repaired, then the mirror can continue as before the failure, possibly with synchronization, depending on the failure.

For fast synchronization of the images after a primary failure, MirrorView provides a write-intent log feature. The write-intent log records the current activity so that a repaired primary needs to copy only data that recently changed (instead of the entire image), thus greatly reducing the recovery time.

Secondary Image Failure

When a primary image cannot communicate with a secondary image, it marks the secondary as unreachable and stops trying to write to it. However, the secondary image remains a member of the mirror.

The primary image also attempts to minimize the amount of work required to synchronize the secondary after it recovers. It does this by *fracturing* the mirror. This means that, while the secondary is unreachable, the primary keeps track of all write requests so that only those blocks that were modified need to be copied to the secondary during recovery. When the secondary is repaired, the software writes the modified blocks to it while mirroring.

The following table shows how MirrorView might help you recover from system failure at the primary and secondary sites. It assumes that the mirror is active and is in the in-sync or consistent state.

Table 3-1 MirrorView Recovery Scenarios

Event	Result and recovery
Server or storage system running primary image fails.	Option 1 - Catastrophic failure, repair is difficult or impossible. If a host is attached to the secondary storage system, the administrator promotes the secondary image, and then takes other prearranged recovery steps required for application startup on the standby host. **Note**: Any writes in progress when the primary image fails may not propagate to the secondary image. Also, if the remote image was fractured at the time of the failure, any writes since the fracture will not have propagated.
	Option 2 -Non-catastrophic failure, repair is feasible. The administrator has the problem fixed, and then synchronizes the secondary image. The write-intent log, if used, shortens the synchronization time needed. If a write intent log is not used, and the secondary LUN was fractured at the time of failure, then a full synchronization is necessary.
Storage system running secondary image fails.	The administrator has a choice: If the secondary image can easily be fixed (for example, if someone pulled out a cable), then the administrator can have it fixed and let things resume. If the secondary cannot be fixed easily, the administrator may want to create a secondary image of the mirror on another storage system to preserve recovery option in the event of another failure.

MirrorView Example

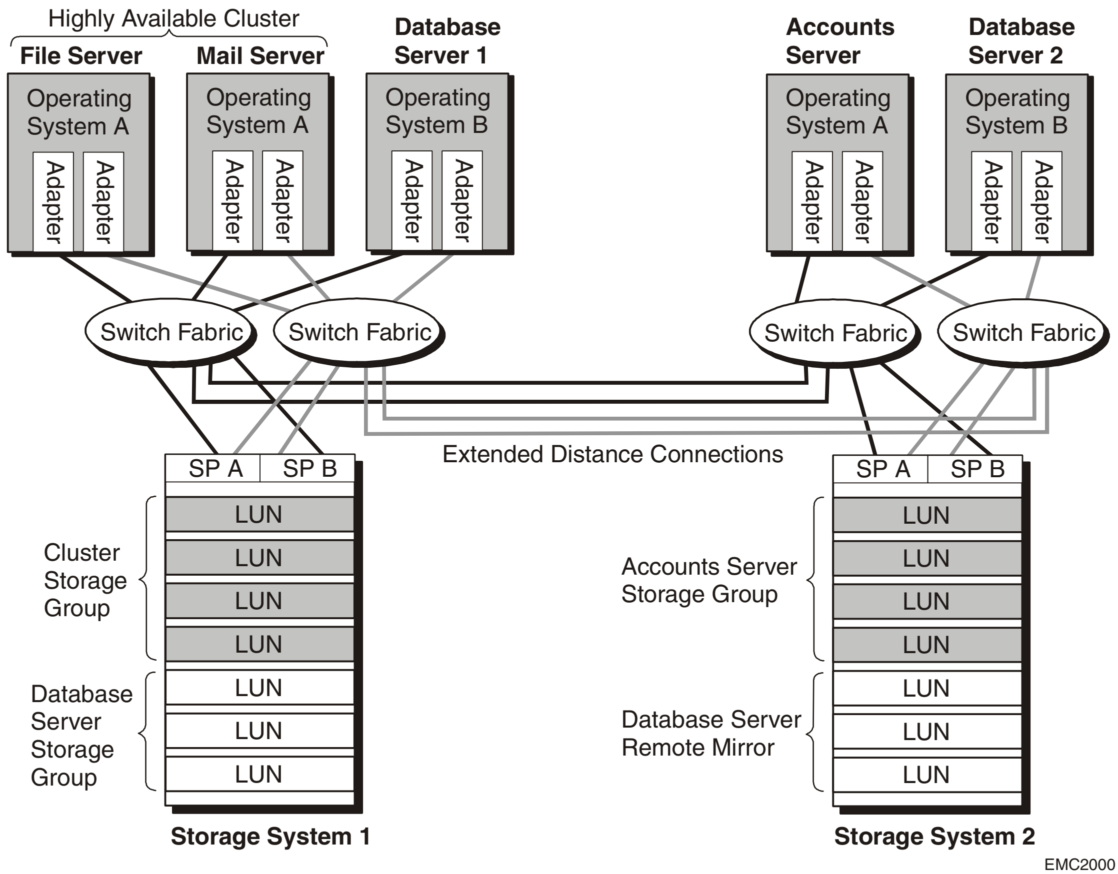


Figure 3-2 Sample MirrorView Configuration

In the figure above, Database Server 1, the production host, executes customer applications. These applications access data on Storage System 1, in the database server Storage Group. Storage System 2 is 40 kilometers away and mirrors the data on the Database Server Storage Group. The mirroring is synchronous, so that Storage System 2 always contains all data modifications that are acknowledged by Storage System 1 to the production host.

Each server has two paths — one through each SP — to each storage system. If a failure occurs in a path, then the storage-system software may switch to the path through the other SP (transparent to any applications).

The server sends a write request to an SP in Storage-System 1, which writes data to its image while sending it to the corresponding SP in Storage System 2. Storage System 2 then writes the data to its image. Storage System 1 then acknowledges the write to the production host.

Database server 2, the other host, has no direct access to the mirrored data. (The standby site does not need a server at all.) This server runs applications that access other data stored on Storage System 2. If a failure occurs in either the production host or Storage System 1, an administrator can use the management station to promote the image on Storage-System 2 to the primary image. Then the appropriate applications can start on any connected server (here, Database Server 2) with full access to the data. The mirror will be accessible in minutes, although the time needed for applications to recover will vary.

MirrorView Planning Worksheet

To plan, you must decide whether you want to use a write-intent log and, if so, which LUNs you will bind for this. You will also need to complete a MirrorView mirroring worksheet.

Note that you must assign each primary image LUN to a Storage Group (as with any normal LUN), but you must not assign a secondary image LUN to a Storage Group.

MirrorView Mirroring Worksheet

Production Host Name	Primary LUN ID, Size, and File System Name	Storage Group Number or Name	Use Write Intent Log - Y/N (256 Mbytes per Storage System)	SP (A/B)	Remote Mirror Name	Secondary Image Contact Person	Secondary Image LUN ID

What Next?

For information on SnapView snapshot copy software, continue to the next chapter. To plan LUNs and file systems, skip to Chapter 5. For details on the storage-system hardware, skip to Chapter 6.

4

About SnapView Clone and Snapshot Software

This chapter introduces the SnapView software that creates clones and snapshots you can use for independent data modeling or backup.

Major sections are

What Is EMC SnapView Software?

EMC SnapView™ is a storage-system-based software application that allows you to create up to eight copies of a LUN using either clones or snapshots.

A clone is an actual copy of a LUN and it takes time to create. A snapshot is a virtual point-in-time copy of a LUN and takes only seconds to create, since it does not actually copy data. Depending on your application needs, you can create clones, snapshots or snapshots of clones.

You can use the snapshot or clone for decision support, system backups, data modeling, software application testing, or as a base for temporary operations on the production data without the risk of damaging the original data on the source LUN.

SnapView Benefits

SnapView clones and snapshots offer the following important benefits:

- They allow full access to production data with modest impact on performance.
- For decision support or revision testing, the provide a coherent, readable and writable copy of real production data at a particular point in time.
- For backup, they practically eliminates the time that production data spends off line or in hot backup mode. And they offload the backup overhead from the production host to another host.

Clones

A clone is really a local mirror with sophisticated synchronization and management features. It is an actual of a source LUN. You specify a source LUN when you create a Clone Group. The copy of the source LUN begins when you add a clone LUN to the Clone Group. I/O can continue to the source LUN while copying to the clone proceeds.

While the clone is part of the Clone Group, any host write requests made to the source LUN are copied to the clone. Once the clone contains the required data, you can fracture the clone, which breaks it from its source and makes it available to a second host. The former clone retains the copied data and becomes a normal LUN that occupies disk space and can be used for other operations.

For all clones in a storage system, the software uses Clone Private LUNs (CPLs) to record the regions of the source LUN and the regions of the clone LUN that have changed since the clone was fractured. The CPL helps reduce the time it takes to synchronize or reverse synchronize a clone with its source, since only modified chunks need to be copied.

The following figure shows clone synchronization and usage.

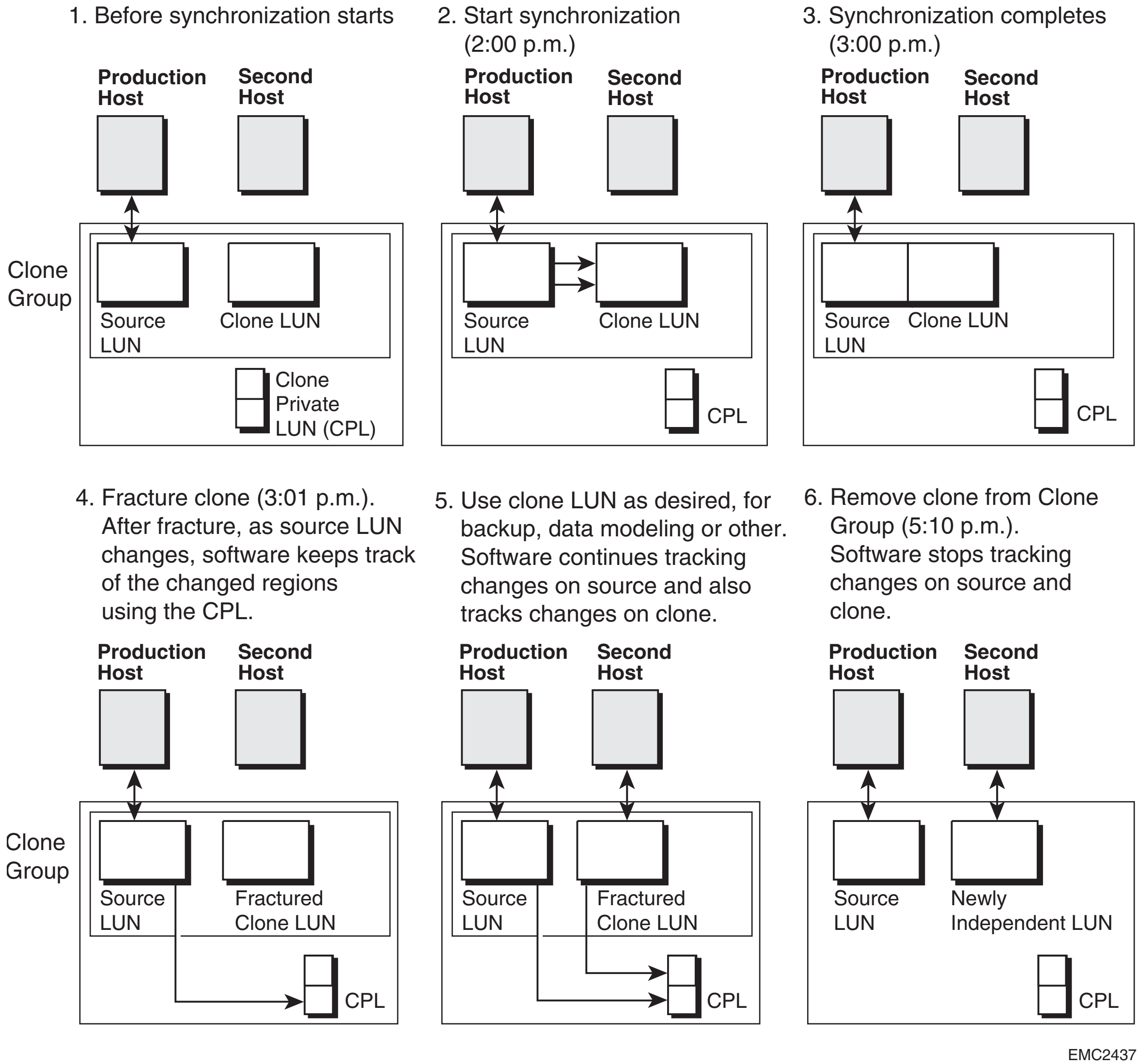


Figure 4-1 Clone Synchronization and Usage

You can start and stop clone operations using directives to EMC ControlCenter™ Navisphere® software on either host.

The SnapView software includes a host-based utility named **admsnap** whose commands activate and deactivate clones (that is, make newly created clone devices available to the host operating system and remove access to the devices after the session ends).

Snapshots

A SnapVIew snapshot is a virtual copy of a source LUN. The snapshot does not reside on disk like a conventional LUN. It does not occupy any disk space. However, the snapshot appears as a conventional LUN to another host. Any other server can access the snapshot for data processing analysis, testing, or backup.

SnapView can create or destroy a snapshot in seconds, regardless of the LUN size, since it does not actually copy data. When the original data on the source LUN is modified, the software stores a copy of that data in the snapshot cache, in blocks called chunks. (The snapshot cache is sometimes called the save area.) For any chunk, the copy happens only once, when the block is first modified. This is called copy on first write (COFW).

For data overwritten since the snapshot was started, the snapshot accesses the original chunks from the snapshot cache instead of the source LUN. As time passes, and I/O modifies the source LUN, the number of chunks stored in the snapshot cache grows.

The following figure shows how a snapshot session works.

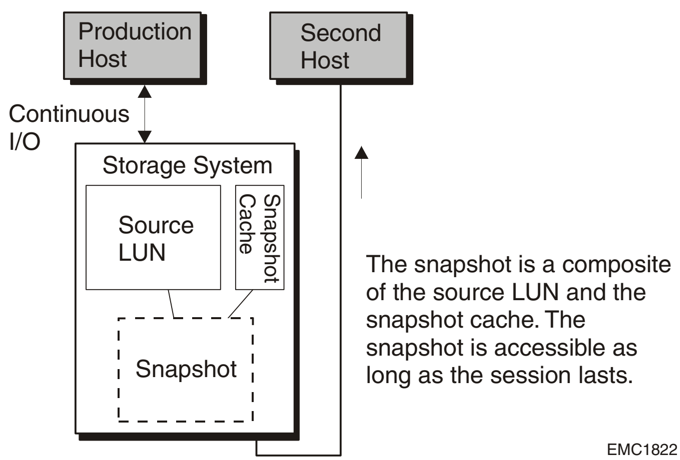


Figure 4-2 SnapView Operations Model

- The production host runs the customer applications on the LUN that you want to copy, and allows the management software to create, start, and stop snapshot sessions.
- The second host reads the snapshot during the snapshot session, and performs analysis or backup using the snapshot.

You can start and stop snapshot sessions using directives to EMC ControlCenter™ Navisphere® software on the production host.

The SnapView software includes a host-based utility named **admsnap** whose commands activate and deactivate snapshots. That is, make newly created snapshot devices available to the host operating system and remove access to the devices after the session ends.

Sample Snapshot Session

The following figure shows a sample snapshot session.

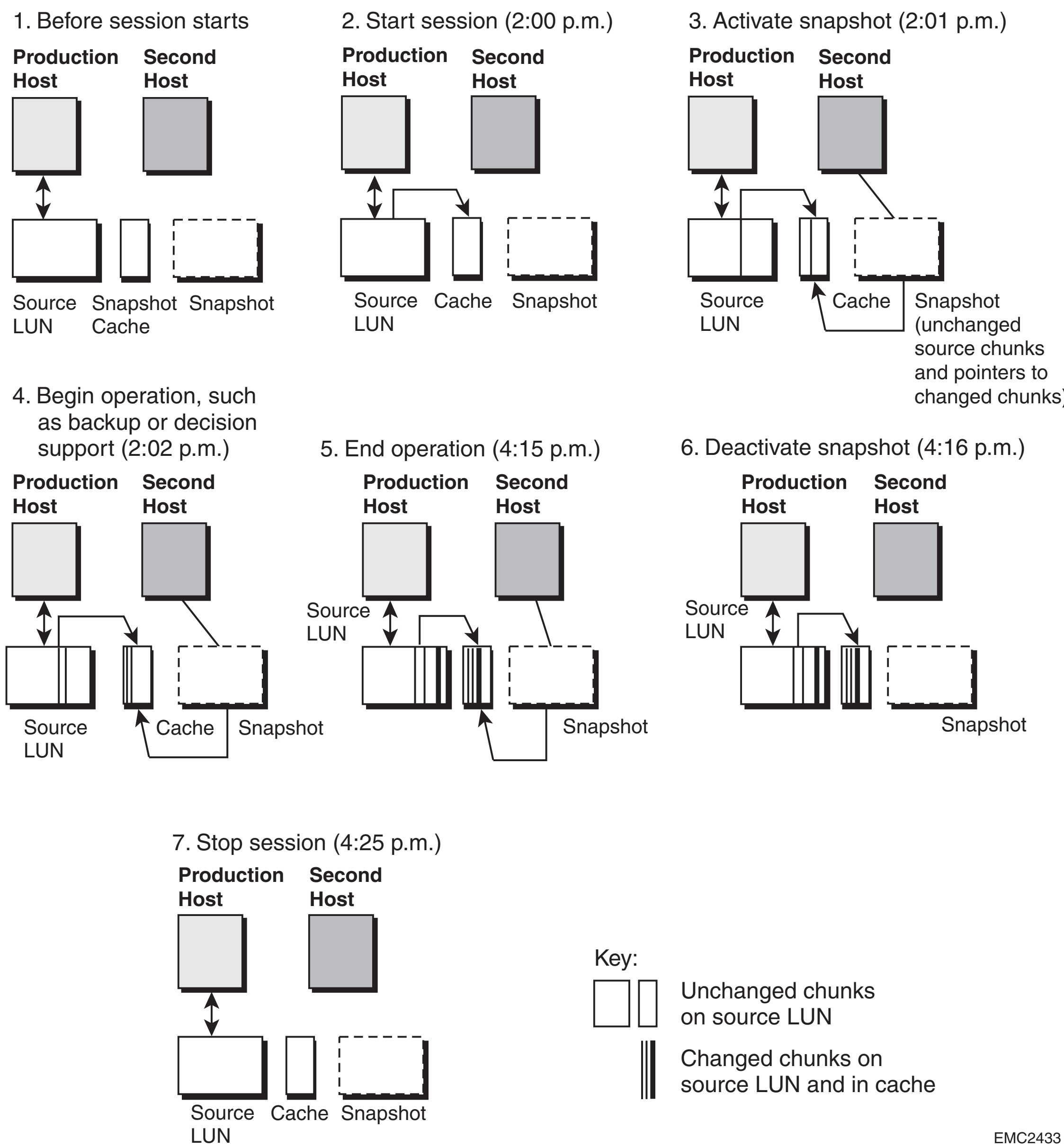


Figure 4-3 How a Snapshot Session Starts, Runs, and Stops

Snapshot and Clone Tradeoffs

The following table describes the benefits and tradeoffs of using clones, snapshots, and snapshots of clones.

	Snapshot	Clone	Snapshot of Clone
Benefits	Provides a quick and instant copy because it is a virtual LUN.	• Provides enhanced protection against critical data because it is a real LUN. • Provides immediacy in replacing the source LUN with the clone LUN, should the source become corrupted.	Provides an extra level of protection against critical data should the source LUN and clone LUN become corrupted.
Creation Time	Instantaneous	Minutes to hours. The creation time depends on the size of the source LUN.	Instantaneous
Disk Space Used	Uses snapshot cache space, which is 10% to 20% of the source LUN size.	Uses full disk space (100%), which is the same amount of disk space that the source LUN uses.	Uses snapshot cache space and full disk space, which is 110% to 120% of the source LUN size.
Data Recovery Time After Source LUN Failure	Restore time from backup medium (tape or network)	Instantaneous after initializing a reverse synchronization.	Restore time from backup tape or network if fractured clone received host I/O.
Performance Impact on the Storage System	Higher than average and constant due to copy on first write.	• After the clone is fractured, its performance is average in comparison to a normal LUN. • For the initial synchronization, the performance is below average. • For synchronizations and reverse synchronizations, the performance is the same as an initial synchronization but for a shorter period of time. Impact is also determined by the Synchronization Rate, which is set when the clone is added to the Clone Group.	Combination of both clones and snapshots.

SnapView Worksheets

This section includes clone and snapshot planning worksheets.

Clone Planning Worksheet

The following worksheet will help you plan for clones. Note that the size of the Clone Private LUNs (one for each SP) is fixed and it is quite modest, so the worksheet does not include it.

Production Host Name	LUN			
	Source LUN ID	Storage Group ID	Size (Gbytes)	Application, File System, or Database Name

Snapshot Planning Worksheet

The following information is needed for system setup to let you bind one or more LUNs for the snapshot cache.

Snapshot Cache Setup Information (For Binding)

SP	Snapshot Source LUN Size	RAID tYpe For Snapshot Cache	RAID Group ID of Parent RAID Group	LUN size (Gbytes, We Suggest 20% of Source LUN Size)	Cache LUN ID (Complete after Binding)
A					
B					

For each session, you must complete a snapshot session worksheet. Note that you must assign the LUN and snapshot to different Storage Groups. One Storage Group must include the production host and source LUN; another Storage Group must include the second host and the snapshot.

Snapshot Session Worksheet

Production Host Name	LUN				Snapshot cache			SP (Both LUN and Cache)	Time of Day To Copy	Session Name
	Source LUN ID	Storage Group ID	Size (MB)	Application, File System, or Database Name	LUN ID	Size (MB)	Chunk (Cache Write) Size			

What Next? To plan LUNs and file systems, continue to the next chapter. For details on the storage-system hardware, skip to Chapter 6.

5

Planning File Systems and LUNs

This chapter shows sample installations with sample shared switched and unshared direct storage, and then provides worksheets for planning your own storage installation. Topics are

Multiple Paths to LUNs

A shared storage-system installation includes two or more servers and one or more storage systems, each with two SPs and the Access Logix option. Often shared storage installations include two or more Fibre Channel switches.

Shared storage (switched or direct) provides at least two paths to each LUN in the storage system. The Access Logix software detects all paths and, using optional PowerPath™ failover software, can automatically switch to the other path, without disrupting applications, if a device (such as a host bus adapter or cable) fails. And with two adapters and two SPs, PowerPath can send I/O to each available path in a user-selectable sequence (multipath I/O) for load sharing and greater throughput.

Unshared storage, if the server has two adapters, provides two paths to each LUN. With two adapters, PowerPath performs the same function as with shared systems: automatically switches to the other path if a device (such as host bus adapter or cable) fails.

Sample Shared Switched Installation

The following figure shows a sample shared storage system connected to three servers: two servers in a cluster and one server running a database management program.

Disk IDs have the form *b e d*, where *b* is the back-end loop number (0, or 1), *e* is the enclosure number, set on the enclosure front panel (0 for the first — lowest — DAE2), and *d* is the disk position in the enclosure (left is 0, right is 14).

Note that figure shows a single-cabinet storage system. In a storage system with two cabinets, all the disks in the first cabinet are on loop 0 (so the Disk IDs in the first cabinet begin with 0) and the disks in the second cabinet are on loop 1 (so their Disk IDs begin with 1).

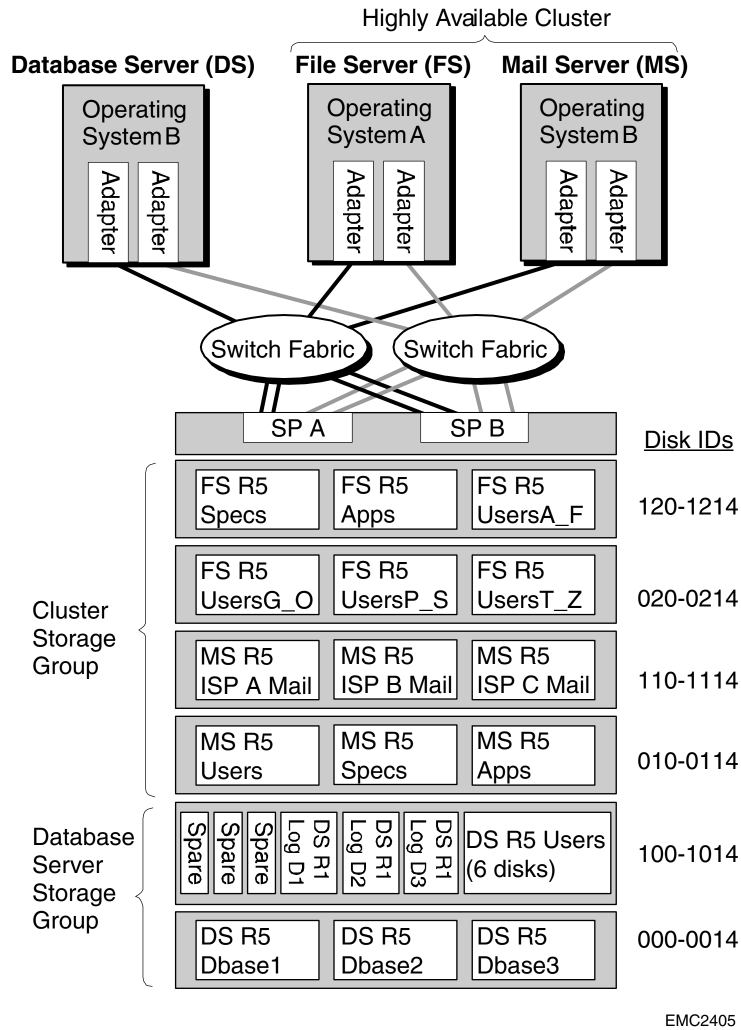


Figure 5-1 Sample Shared Switched Storage Configuration

The storage-system disk IDs, RAID types, storage types, capacities, and drive letters are as follows. The capacities shown reflect 73-Mbyte disks.

Database Server LUNs (DS) - SP A 1,440 Gbytes on Seven LUNs			Clustered System LUNs					
			Mail Server LUNs (MS) - SP A 1,460 Gbytes on Seven LUNs			File Server LUNs (FS) - SP B 1,460 Gbytes on Seven LUNs		
Disk IDs	**RAID Type**	**Storage Type, Capacity, Drive**	**Disk IDs**	**RAID Type**	**Storage Type, Capacity, Drive**	**Disk IDs**	**RAID Type**	**Storage Type, Capacity, Drive**
000- 004	RAID 5	Dbase1, 292 Mbytes, drive G	010-014	RAID 5	Users, 292 Mbytes, drive N	020-024	RAID 5	UsersG_O, 292 Mbytes, drive T
005- 009	RAID 5	Dbase2, 292 Mbytes, drive H	015-019	RAID 5	Specs, 292 Mbytes, drive O	025-029	RAID 5	UsersP_S, 292 Mbytes, drive U
0010- 0014	RAID 5	Dbase3, 292 Mbytes, drive I	0110- 0114	RAID 5	Apps, 292 Mbytes, drive P	0210- 0214	RAID 5	UsersT_Z, 292 Mbytes, drive V
103- 104	RAID 1	Log file,Dbase1, 73 Mbytes, drive J	110-114	RAID 5	ISP A Mail, 292 Mbytes, drive Q	120-124	RAID 5	Specs, 292 Mbytes, drive U
105- 106	RAID 1	Log file, Dbase2, 73 Mbytes, drive K	115-119	RAID 5	ISP B Mail, 292 Mbytes, drive R	125-129	RAID 5	Apps, 292 Mbytes, drive V
107-108	RAID 1	Log file, Dbase3, 73 Mbytes drive L	1110- 1114	RAID 5	ISP C Mail, 292 Mbytes, drive S	1210- 1214	RAID 5	UsersA_F, 292 Mbytes, drive W
109- 1014	RAID 5	Users, 345 Mbytes (6 disks), drive M						
100, 101, 102 – Hot spare (automatically replaces a failed disk in any server's LUN)								

Sample Unshared Direct Installation

This section shows the disks and LUNs in an unshared direct storage-system installation.

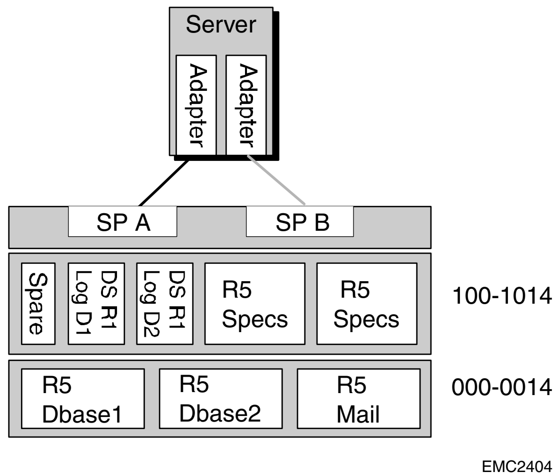


Figure 5-2 Unshared Direct Storage System

If each disk holds 73 Gbytes, then the storage system provides the server with 504 Gbytes of disk storage, all highly available. The storage-system disk IDs and LUNs are as follows.

LUNs - SP A and SP B - 1,440 Gbytes on Seven LUNs		
Disk IDs	**RAID type**	**Storage type, capacity, drive**
000- 004	RAID 5	Dbase1, 292 Mbytes, drive G
005- 009	RAID 5	Dbase2, 292 Mbytes, drive H
0010- 0014	RAID 5	Mail, 292 Mbytes, drive I
101- 102	RAID 1	Log file,Dbase1, 73 Mbytes, drive J
103- 104	RAID 1	Log file, Dbase2, 73 Mbytes, drive K
105- 109	RAID 1	Users, 292 Mbytes, drive L
1010-1014	RAID 5	Specs, 292 Mbytes, drive M
100 – Hot spare (automatically replaces a failed disk in any LUN)		

Planning Applications, LUNs, and Storage Groups

This section will help you plan your storage use — the applications to run, the LUNs that will hold them, and, for shared storage, the Storage Groups that will belong to each server.

Unshared storage systems do not use Storage Groups. For unshared storage, on the LUN and Storage Group worksheet, skip the Storage Group entry.

The worksheets include

- Application and LUN planning worksheet — lets you outline your storage needs.
- LUN and Storage Group planning worksheet — lets you decide on the disks to compose the LUNs and the LUNs to compose the Storage Groups for each server.
- LUN details worksheet — lets you plan each LUN in detail.

Make as many copies of each blank worksheet as you need. You will need this information later when you configure the storage system(s).

Sample worksheets appear later in this chapter.

Application and LUN Planning

Use the following worksheet to list the applications you will run, and the RAID type and size of LUN to hold them. For each application that will run, write the application name, file system (if any), RAID type, LUN ID (ascending integers, starting with 0), disk space required, and finally the name of the servers and operating systems that will use the LUN.

Application and LUN Planning Worksheet

Application	File System, Partition, or Drive	RAID Type of LUN	LUN ID (hex)	Disk Space Provided (Gbytes)	Server Hostname and Operating System

A sample worksheet begins as follows:

Application	File system, Partition, or Drive	RAID Type of LUN	LUN ID (hex)	Disk Space Provided (Gbytes)	Server Hostname and Operating System
Dbase1	*S*	*RAID 5*	*0*	*292 GB*	*Server1,Windows*
Dbase2	*T*	*RAID 5*	*1*	*292 GB*	*Server2,Windows*
Log file for Dbase1	*U*	*RAID 1*	*2*	*73 GB*	*Server1,Windows*
Log file for Dbase2	*V*	*RAID 1*	*3*	*73 GB*	*Server2,Windows*
Users	*W*	*RAID 5*	*4*	*365 GB*	*Server1,Windows*

Completing the Application and LUN Planning Worksheet

Application. Enter the application name or type.

File system, partition, or **drive**. Write the planned **file system**, **partition,** or **drive** name here.

With a Windows operating system, the LUNs are identified by drive letter only. The letter does not help you identify the disk configuration (such as RAID 5). We suggest that later, when you use the operating system to create a partition on a LUN, you use the disk administrator software to assign a volume label that describes the

RAID type. For example, for drive T, assign the volume ID **RAID5_T**. The volume label will then identify the drive letter.

RAID type of LUN. This is the RAID Group type you want for this file system, partition, or drive. The features of RAID types are explained in Chapter 3. For a RAID 5, RAID 1, RAID 1/0, and RAID 0 Group, you can create one or more LUNs on the RAID Group. For other RAID types, you can create only one LUN per RAID Group.

LUN ID. The LUN ID is a hexadecimal number assigned when you bind the disks into a LUN. By default, the ID of the first LUN bound is 0, the second 1, and so on. Each LUN ID must be unique within the storage system, regardless of its Storage Group or RAID Group.

The maximum number of LUNs supported on one host bus adapter depends on the operating system.

Disk space required (Gbytes). Consider the largest amount of disk space this application will need, and then add a factor for growth.

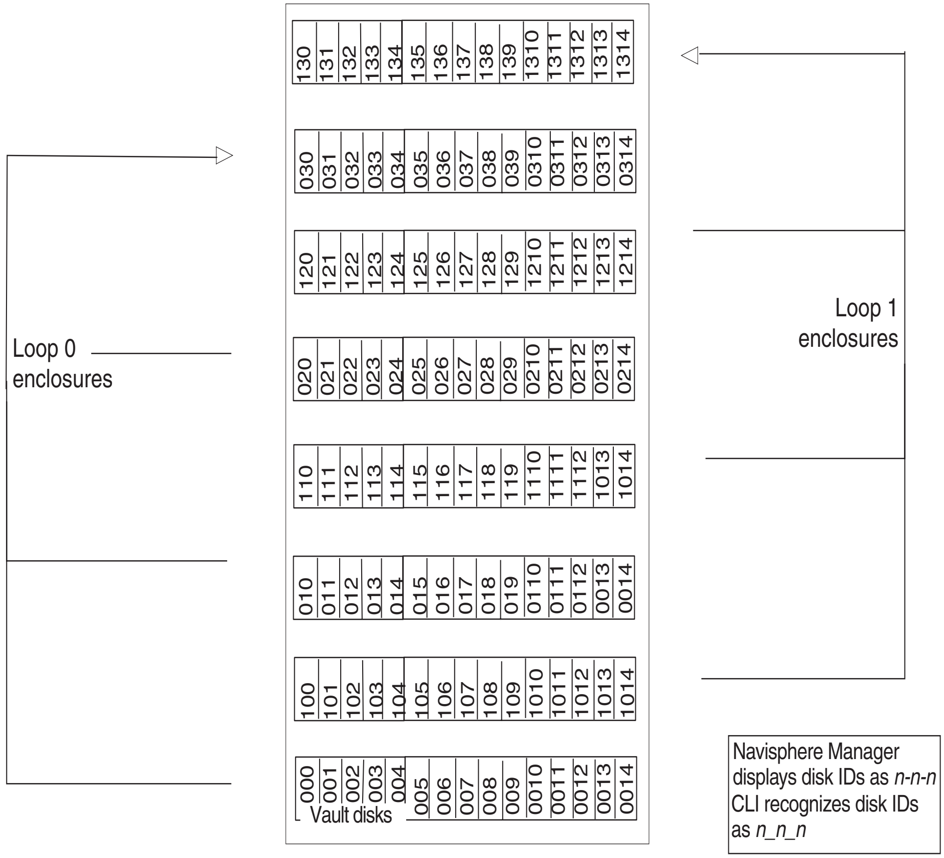
Server hostname and operating system. Enter the server hostname (or, if you don't know the name, a short description that identifies the server) and the operating system name, if you know it.

LUN and Storage Group Planning Worksheet

Use the following worksheet to select the disks that will make up the LUNs and Storage Groups in each storage system. A storage system is any group of enclosures connected to an SPE (CX600 storage system) or DPE2 (CX400 storage system). A CX600 storage system can include up to l6 DAE2 enclosures for a total of 240 disks; a CX400, can include up to three DAE2 enclosures for a total of 60 disks.

Unshared storage systems do not use Storage Groups. For unshared storage, skip the Storage Group entry.

LUN and Storage Group Worksheet for CX600 - Single Cabinet



Storage-system number or name:_______________

Storage Group ID or name:______ Server hostname:_________________ ☐Dedicated ☐Shared

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_____________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_____________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_____________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_____________________________________

Storage Group ID or name:______ Server hostname:_________________ ☐Dedicated ☐Shared

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_____________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_____________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_____________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_____________________________________

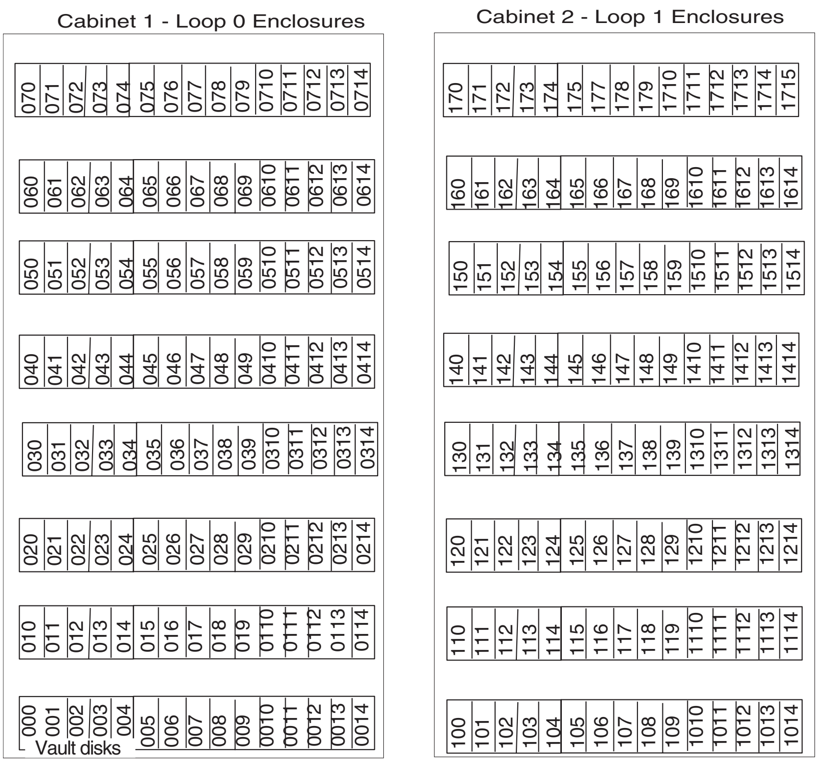
Storage Group ID or name:______ Server hostname:_________________ ☐Dedicated ☐Shared

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_____________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_____________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_____________________________________

LUN and Storage Group Worksheet for CX600 - Dual Cabinets



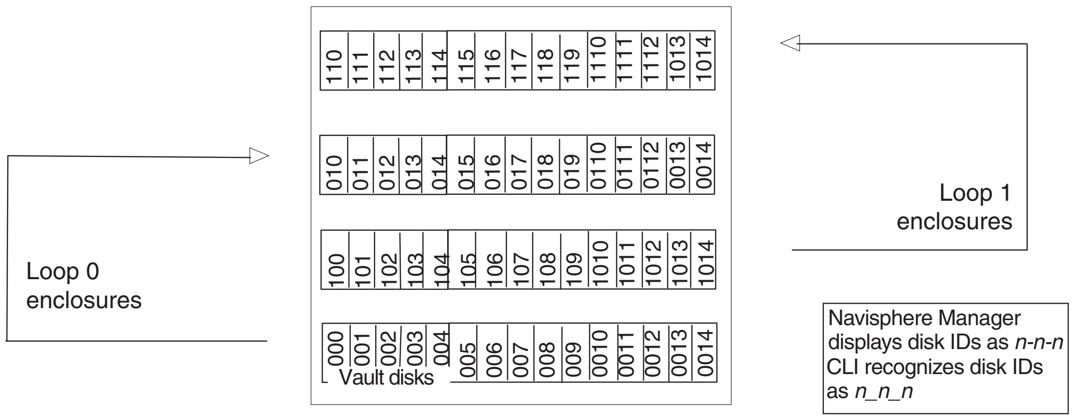
Navisphere Manager displays disk IDs as *n-n-n.* CLI recognizes disk IDs as *n_n_n*

Storage Group ID or name:______ Server hostname:__________________ ☐Dedicated ☐Shared
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs____
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs____
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs____
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs____

Storage Group ID or name:______ Server hostname:__________________ ☐Dedicated ☐Shared
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs____
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs____
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs____
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs____

Storage Group ID or name:______ Server hostname:__________________ ☐Dedicated ☐Shared
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs____
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs____
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs____
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs____

LUN and Storage Group Worksheet for CX400



Storage-system number or name:________________

Storage Group ID or name:______ Server hostname:__________________ ☐Dedicated ☐Shared

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_______________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_______________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_______________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_______________________________________

Storage Group ID or name:______ Server hostname:__________________ ☐Dedicated ☐Shared

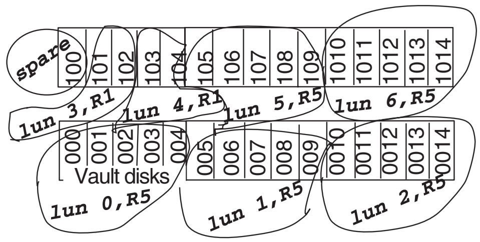
LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_______________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_______________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_______________________________________

LUN ID or name_______RAID type ___ Cap. (Gb) _____ Disk IDs_______________________________________

Part of a sample LUN and Storage Group worksheet for one cabinet follows.



Storage-system number or name:_____*SS1*___________

Storage Group ID or name:__*0*___ Server hostname:_____*Server1*_________ Dedicated ☒ Shared ☐

LUN ID or name___*0*___RAID type _*5*_ Cap. (Gb) _*292*_ Disk IDs___*000, 001, 002, 003, 004*______________

LUN ID or name___*1*___RAID type _*5*_ Cap. (Gb) _*292*_ Disk IDs___*005, 006, 007, 008, 009*______________

LUN ID or name___*2*___RAID type _*5*_ Cap. (Gb) _*292*_ Disk IDs___*0010, 0011, 0012, 0013, 0015*________

LUN ID or name___*3*___RAID type _*1*_ Cap. (Gb) __*73*_ Disk IDs___*101, 102*________________________________

Storage Group ID or name:__*0*___ Server hostname:_____*Server1*_________ Dedicated ☒ Shared ☐

LUN ID or name___*4*___RAID type _*1*_ Cap. (Gb) __*73*_ Disk IDs____*103,104*______________________________

LUN ID or name___*5*___RAID type _*5*_ Cap. (Gb) _*292*_ Disk IDs_____*105, 106, 107, 108, 109*___________

LUN ID or name___*6*___RAID type ___ Cap. (Gb) _*292*_ Disk IDs____1*010, 1011, 1012, 1013, 1014*_______

LUN ID or name_______RAID type ___ Cap. (Gb) ______ Disk IDs____

Completing the LUN and Storage Group Planning Worksheet

As shown, draw circles around the disks that will compose each LUN, and within each circle specify the RAID type (for example, RAID 5) and LUN ID. This is information you will use to bind the disks into LUNs. For disk IDs, use the form shown on page 5-10.

Disk configuration rules are as follows:

- None of the disks 000 through 008 may serve as a hot spare.
- On CX-Series systems, the hardware reserves about 7 Gbytes on each of the disks 000 through 004 for the cache vault and internal tables. To conserve disk space, you should avoid binding any other in a RAID group that includes any of the disks 000-004. Any disk you include in a RAID Group with a cache disk 000-004 will be bound to match the lower cache disk capacity, resulting in lost storage of 7 Gbytes per disk.
- Each disk in the RAID Group must have the same capacity; otherwise, all disks will be bound to match the smallest and you will waste disk space.
- If a storage system will use disks of different speeds (for example, 10K and 15K rpm), then EMC recommends that you use disks of the same speed throughout each 15-disk enclosure. For any enclosure, the hardware allows *one* juxtaposition of disks of different speeds, so if need be, you may use disks of differing speeds. However, in every case, you should use disks of the same speed and capacity in any RAID group.

Next, complete as many of the Storage System sections as needed for all the Storage Groups in the SAN (or as needed for all the LUNs with unshared storage). Copy the (blank) worksheet as needed.

For shared storage, if a Storage Group will be dedicated (not accessible by another server in a cluster), mark the Dedicated box at the end of its line; if the Storage Group will be accessible to one or more other servers in a cluster, write the hostnames of all servers and mark the Shared box.

For unshared storage, ignore the Dedicated/Shared boxes.

LUN Details Worksheet

Use the LUN Details Worksheet to plan the individual LUNs. Blank and sample completed LUN worksheets follow.

Complete as many blank worksheets as needed for all LUNs in storage systems. For unshared storage, skip the Storage Group entries.

LUN Details Worksheet

Storage system (complete this section once for each storage system)

Storage-system number or name:_______
Storage-system installation type
❏ Unshared Direct o Shared-or-Clustered Direct ❏ Shared Switched

SP information: SP A: IP address or hostname:________Port ALPA ID:______ Memory(Mbytes):___
SP B: IP address or hostname:_________Port ALPA ID:______ Memory(Mbytes):___
❏Caching Read cache:____% Write cache: ____% Cache page size (Kbytes):___
❏ RAID-3

LUN ID:_____ SP owner: ❏A ❏B SP loop (0 or 1):___

RAID Group ID:___Size, GB:___ LUN size, GB:___Disk IDs:_________________________
RAID type: ❏RAID 5 ❏RAID 3 - Memory, MB:___ ❏RAID 1 mirrored pair ❏RAID 0
❏RAID 1/0 ❏Individual disk ❏Hot spare
Caching: ❏Read and write ❏Write ❏Read ❏None

Servers that can access this LUN's Storage Group: _______________________________________

Operating system information: Device name:_______________File system, partition, or drive:__________

LUN ID:_____ SP owner: ❏A ❏B SP loop (0 or 1):___

RAID Group ID:___Size, GB:___ LUN size, GB:___Disk IDs:_________________________
RAID type: ❏RAID 5 ❏RAID 3 - Memory, MB:___ ❏RAID 1 mirrored pair ❏RAID 0
❏RAID 1/0 ❏Individual disk ❏Hot spare
Caching: ❏Read and write ❏Write ❏Read ❏None

Servers that can access this LUN's Storage Group: _____________________________________

Operating system information: Device name:_______________File system, partition, or drive:____

LUN ID:_____ SP owner: ❏A ❏B SP loop (0 or 1):___

RAID Group ID:___Size, GB:___ LUN size, GB:___Disk IDs:__________________________________
RAID type: ❏RAID 5 ❏RAID 3 - Memory, MB:___ ❏RAID 1 mirrored pair ❏RAID 0
❏RAID 1/0 ❏Individual disk ❏Hot spare
Caching: ❏Read and write ❏Write ❏Read ❏None

Servers that can access this LUN's Storage Group: ____

Operating system information: Device name:_______________File system, partition, or drive:_________

LUN Details Worksheet

Storage system (complete this section once for each storage system)

Storage-system number or name:__***SS1***_____
Storage-system installation type
☒Unshared Direct ❑ Shared-or-Clustered Direct ❑ Shared Switched

SP information: SP A: IP address or hostname:_***SS1spa***_Port ALPA ID:_***0***__ Memory(Mbytes):_***1 Gb***_
SP B: IP address or hostname:_***SS1spb***_Port ALPA ID:_***1***__ Memory(Mbytes):_***1 Gb***__
❑Caching Read cache:_***33***_% Write cache: _***67***_% Cache page size (Kbytes):_***4***_
❑ RAID-3

LUN ID:__***0***___ SP owner: ❑A ❑B SP loop (0 or 1):_***0***___

RAID Group ID:_***0***__Size,GB:_***292***__ LUN size,GB:__***292***_Disk IDs:____***000,001,002,003,004***_
RAID type: ☒RAID 5 ❑RAID 3 - Memory, MB:___ ❑RAID 1 mirrored pair ❑RAID 0
❑RAID 1/0 ❑Individual disk ❑Hot spare
Caching: ❑Read and write ❑Write ❑Read ❑None

Servers that can access this LUN's Storage Group: __***Server1*** _______

Operating system information: Device name:_______________File system, partition, or drive:__***S***______

LUN ID:__***1***____ SP owner: ❑A ❑B SP loop (0 or 1):___

RAID Group ID:_***1***__Size,GB:__***292***_ LUN size,GB:__***292***_Disk IDs:_____***005,006,007,008,009***
RAID type: ☒RAID 5 ❑RAID 3 - Memory, MB:___ ❑RAID 1 mirrored pair ❑RAID 0
❑RAID 1/0 ❑Individual disk ❑Hot spare
Caching: ❑Read and write ❑Write ❑Read ❑None

Servers that can access this LUN's Storage Group: __***Server1*** ________________________________

Operating system information: Device name:_______________File system, partition, or drive:__***T***__

LUN ID:__***2***___ SP owner: ❑A ❑B SP loop (0 or 1):__***1***_

RAID Group ID:_***2*** Size,GB:__***73***_ LUN size,GB:__***73*** _Disk IDs:______***0101,1011,1012,1013,1014***_
RAID type: ❑RAID 5 ❑RAID 3 - Memory, MB:___ ☒RAID 1 mirrored pair ❑RAID 0
❑RAID 1/0 ❑Individual disk ❑Hot spare
Caching: ❑Read and write ❑Write ❑Read ❑None

Servers that can access this LUN's Storage Group: __***Server1***_________________________________

Operating system information: Device name:_______________File system, partition, or drive:__***U***____

Competing the LUN Details Worksheet

Complete the header portion of the worksheet for each storage system as described next. Copy the blank worksheet as needed.

Storage-System Entries

Storage-system installation type: specify Unshared Direct, Shared-or-Clustered Direct, or Shared Switched.

SP information: IP address or hostname. The IP address is required for communication with the SP. You don't need to complete it now, but you will need it when the storage system is installed so that you can set up communication with the SP.

Port ALPA ID. For shared storage (switches), the SP ALPA ID is ignored for all except HP/UX servers. For unshared storage, each SP connected to a server must have a unique SP Port ALPA ID; for example, for two storage systems, use IDs 0, 1, 2, and 3. The SP Port ALPA ID, like the IP address, is generally set at installation.

Memory (Mbytes). Each CX600 SP has a minimum of 2 Gbytes and maximum of 4 Gbytes of memory. Each CX400 SP has 1 Gbyte of memory.

Caching. You can use SP memory for read/write caching or RAID 3. (We do not recommend using both caching and RAID 3 in the same storage system.) You can use different cache settings for different times of day. For example, for user I/O during the day, use more write cache; for sequential batch jobs at night, use more read cache. You enable caching for specific LUNs — allowing you to tailor your cache resources according to priority. If you choose caching, check the box and continue to the next cache item; for RAID 3, skip to entry **RAID 3**.

Read cache percentage. If you want a read cache, generally it should be about one third (33%) of the total available cache memory.

Write cache percentage. Generally, the write cache should be two thirds (67%) of the total available cache memory.

Cache page size. This applies to both read and write caches. It can be 2, 4, 8, or 16 Kbytes.

As a general guideline, we suggest 8 Kbytes. The ideal cache page size depends on the operating system and application.

RAID 3. If you want to use the SP memory for RAID 3, check the box.

RAID Group/LUN Entries

Complete a RAID Group/LUN entry for each LUN and hot spare.

LUN ID. The LUN ID is a hexadecimal number assigned when you bind the disks into a LUN. By default, the ID of the first LUN bound is 0, the second 1, and so on. Each LUN ID must be unique within the storage system, regardless of its Storage Group or RAID Group.

The maximum number of LUNs supported on one host bus adapter depends on the operating system.

SP owner. Specify the SP that will own the LUN: SP A or SP B. You can let the management program automatically select the SP to balance the workload between SPs; to do so, leave this entry blank.

SP loop (0 or 1). Each CX-Series SP has two sets of back-end loops, 0 and 1. Ideally, you will place the same amount of load on each loop. This may mean placing two or three heavily-used LUNs on one loop, and six or eight lightly used LUNs on the other loop**.** The loop designation appears in the disk ID (form *b-e-d*, where *b* is the back-end loop number, *e i*s the enclosure number, and *d* is the disk position in the enclosure. For example, 0-1-3 indicates the disk on loop 0, in enclosure 1, in the third position (fourth from left) in the storage system.

RAID Group ID. This ID is a hexadecimal number assigned when you create the RAID Group. By default, the number of the first RAID Group in a storage system is 0, the second 1, and so on, up to the maximum of 1F (31).

Size (RAID Group size). Enter the user-available capacity in gigabytes (Gbytes) of the whole RAID Group. You can determine the capacity as follows:

RAID 5 or RAID 3 Group:	disk-size * (number-of-disks - 1)
RAID 1/0 or RAID 1 Group:	(disk-size * number-of-disks) / 2
RAID 0 Group:	disk-size * number-of-disks
Individual unit:	disk-size

For example,

- A five-disk RAID 5 or RAID 3 Group of 73-Gbyte disks holds 292 Gbytes;
- An eight-disk RAID 1/0 Group of 73-Gbyte disks also holds 292 Gbytes;
- A RAID 1 mirrored pair of 73-Gbyte disks holds 73 Gbytes; and
- An individual disk of an 73-Gbyte disk also holds 73 Gbytes.

LUN size. Enter the user-available capacity in gigabytes (Gbytes) of the LUN. You can make this the same size as the RAID Group, described previously. Or, for a RAID 5, RAID 1, RAID 1/0, or RAID 0 Group, you can make the LUN smaller than the RAID Group. You might do this if you wanted a RAID 5 Group with a large capacity and wanted to place many smaller capacity LUNs on it; for example, to specify a LUN for each user. However, having multiple LUNs per RAID Group may adversely impact performance. If you want multiple LUNs per RAID Group, then use a RAID Group/LUN series of entries for each LUN.

Disk IDs. Enter the IDs of all disks that will make up the LUN or hot spare. These are the same disk IDs you specified on the previous worksheet. For example, for a RAID 5 Group in DAE2 on loop 0 (disks 10 through 14), enter **0010**, **0011**, **0012**, **0013**, and **0014**.

RAID type. Copy the RAID type from the previous worksheet. For example, RAID 5 or hot spare. For a hot spare (not strictly speaking a LUN at all), skip the rest of this LUN entry and continue to the next LUN entry (if any).

If this is a RAID 3 Group, specify the amount of SP memory for that group. To work efficiently, each RAID 3 Group needs at least 6 Mbytes of memory.

Caching. If you want to use caching (entry on page 5-18), you can specify the type of caching you want — read and write, read, or write for this LUN. Generally, write caching improves performance far more than read caching. The ability to specify caching on a LUN basis provides additional flexibility, since you can use caching for only the units that will benefit from it. Read and write caching recommendations follow.

Table 5-1 Cache Recommendations for Different RAID Types

RAID 5	RAID 3	RAID 1	RAID 1/0	RAID 0	Individual Unit
Highly Recommended	Not allowed	Recommended	Recommended	Recommended	Recommended

Servers that can access this LUN's Storage Group. For shared switched storage or shared-or-clustered direct storage, enter the name of each server (copied from the LUN and Storage Group worksheet). For unshared direct storage, this entry does not apply.

Operating system information: Device name. Enter the operating system device name, if this is important and if you know it. Depending on your operating system, you may not be able to complete this field now.

File system, partition, or drive. Write the name of the file system, partition, or drive letter you will create on this LUN. This is the same name you wrote on the application worksheet.

On the following line, write any pertinent notes; for example, the file system mount- or graft-point directory pathname (from the root directory). If any of this storage system's LUNs will be shared with another server, and the other server is the primary owner of this LUN, write **secondary**. (As mentioned earlier, if the storage system will be used by two servers, we suggest you complete one of these worksheets for each server.)

What Next?

If you have completed the worksheets to your satisfaction, you are ready to learn about the hardware needed for these systems as explained in Chapter 6.

6

Storage-System Hardware

This chapter describes the storage-system hardware components. Topics are

The storage systems attach to the server and the interconnect components as described in Chapter 1. The installation types are

- **Unshared direct**, with one server, is the simplest and least costly;
- **Shared-or-clustered direct**, with a limit of two servers, lets two servers share storage resources with high availability; and
- **Shared switched**, which has two switch fabrics, lets two or more servers share the resources of several storage systems in a storage area network (SAN).

CX-Series storage systems require at least one network connection.

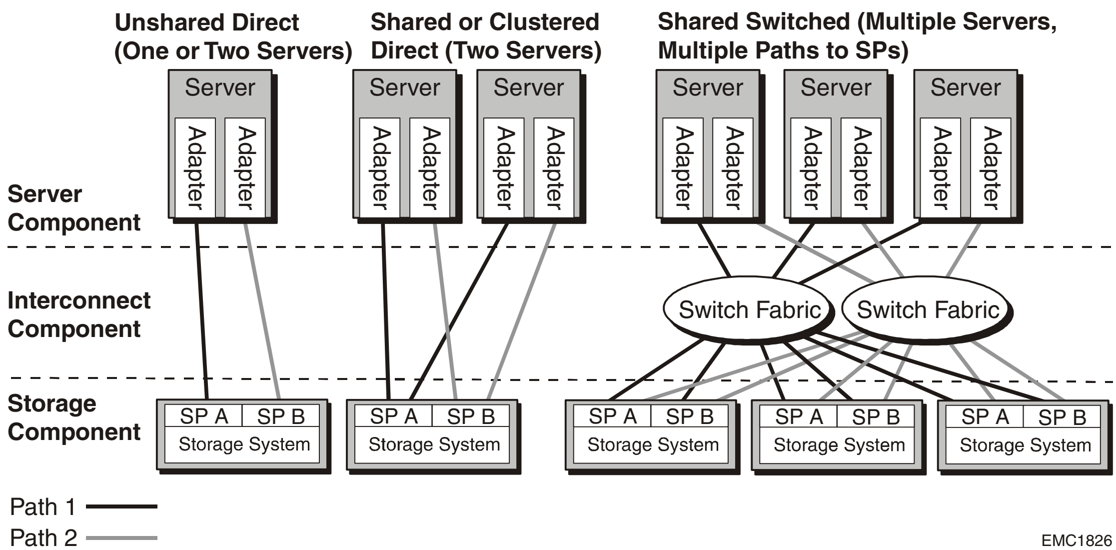


Figure 6-1 Types of Storage-System Installation

Hardware for CX-Series Storage Systems

The main hardware component for a CX600 storage system is a storage processor enclosure (SPE); for a CX400, it is a disk processor enclosure (DPE2). Each storage system has two storage processors (SPs).

An SPE SP has four front-end ports for server or switch connection and two sets of back-end loops that run the disks. An SPE can support up to 16 separate 15-disk enclosures called 2-Gbit disk array enclosures (DAE2s), for a total of 240 disks.

A DPE2 SP has two front-end ports for server or switch connection and two sets of back-end loops that run the disks. A DPE2 can support up to 3 additional DAE2s for a total of 60 disks.

Storage Hardware

A CX600 SPE or CX400 DAE2 enclosure is a sheet-metal housing with a front door, midplane, and slots for the two storage processors (SPs), two power supplies, and fan modules. A CX400 also has space for 15 disks. All components are customer replaceable units (CRUs) that can be replaced under power.

A separate standby power supply (SPS) is required to support write caching. All the storage components in a CX600 — SPE, DAE2s, SPSs, and cabinet — are shown in the following figure.

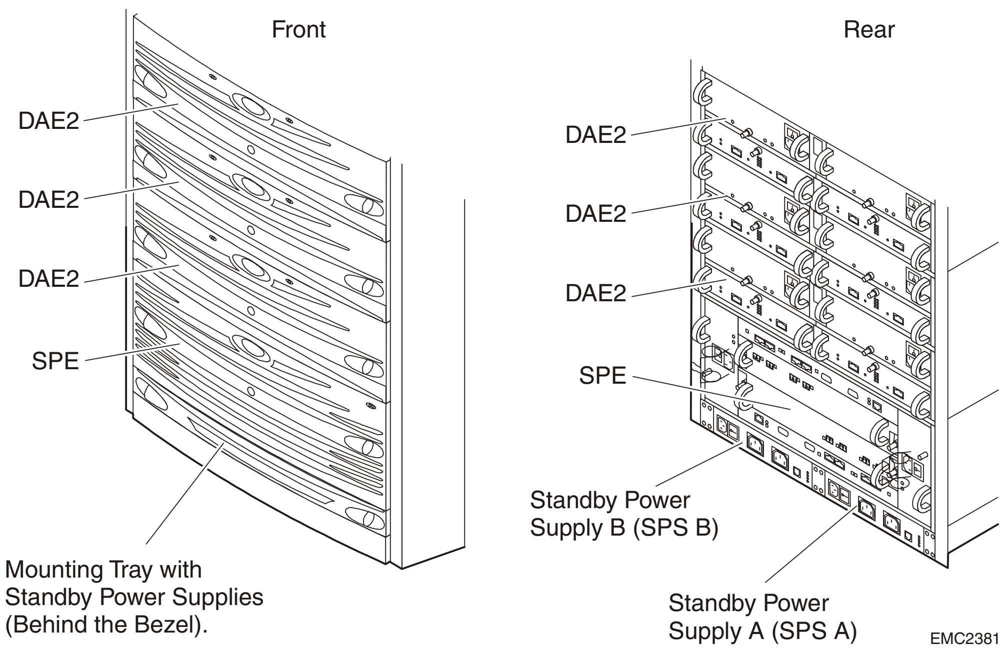


Figure 6-2 CX600 Storage System

Storage components in a CX400 are similar, except that the DPE2 is shorter than the SPE.

Disks

The disks — available in differing capacities — fit into slots in the DAE2 enclosure. Each disk has a unique ID that you use when binding or monitoring its operation. The ID is derived from the Fibre Channel loop number, enclosure address and disk position in the enclosure.

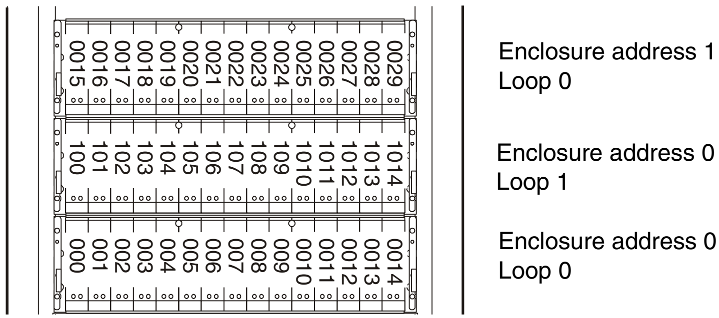


Figure 6-3 Disk Modules and Module IDs

Storage Processor (SP)

The SP provides the intelligence of the storage system. Using its own proprietary software (called Base Software), the SP processes the data written to or read from the disk modules, and monitors the modules themselves. A CX600 SP consists of a printed-circuit board with memory modules, a board that provides SAN features (called a personality board), and status lights.

Each CX600 SP has four front-end ports, and each DPE2 SP has two front-end ports for server or switch connection. Both SP types have two back-end loops that run the disks.

For high availability, a storage system comes with two SPs. The second SP provides a second route to a storage system and also lets the storage system use write caching (below) for enhanced write performance.

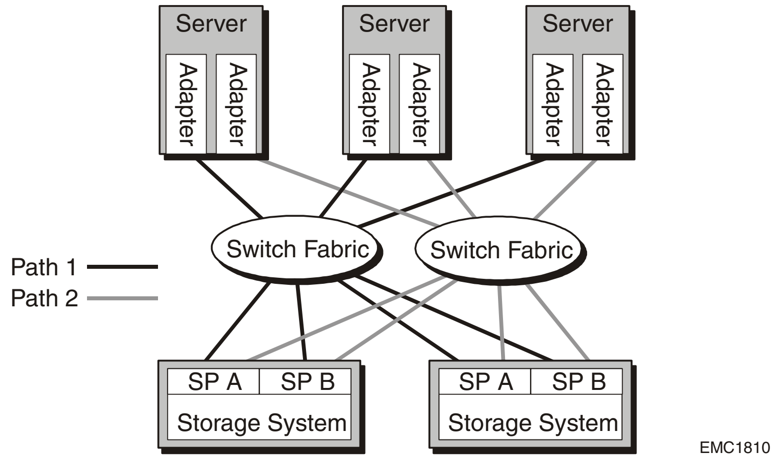


Figure 6-4 Shared Storage Systems

Chapter 5 shows more examples of storage configurations in.

Storage-System Caching

Storage-system caching improves read and write performance for several types of RAID Groups. Write caching, particularly, helps write performance — an inherent problem for RAID types that require writing to multiple disks. Read and write caching improve performance in two ways:

- For a read request — If a read request seeks information that's already in the read or write cache, the storage system can deliver it immediately, much faster than a disk access can.
- For a write request — the storage system writes updated information to SP write-cache memory instead of to disk, allowing the server to continue as if the write had actually completed. The write to disk from cache occurs later, at the most expedient time. If the request modifies information that's in the cache waiting to be written to disk, the storage system updates the information in the cache before writing it to disk; this requires just one disk access instead of two.

A standby power supply (SPS) protects data in the cache from power loss. If line power fails, the SPS provides power to let the storage system write cache contents to the vault disks. The vault disks are standard disk modules that store user data but have space reserved outside operating system control. When power returns, the storage system reads the cache information from the vault disks, and then writes it to the file systems on the disks. This design ensures that all write-cached information reaches its destination.

During normal operation, no I/O occurs with the vault; therefore, a disk's role as a vault disk has no effect on its performance.

SP Network Connection

Each SP has an Ethernet connection though which the Navisphere Manager software lets you configure and reconfigure the LUNs and Storage Groups in the storage system. Each SP connects to a network; this lets you reconfigure your system, if needed, should one SP fail.

Planning Your Hardware Components

This section helps you plan the hardware components — adapters, cables, and storage systems and site requirements — for each server in your installation.

For shared switched storage or shared-or-clustered direct storage, you must use high-availability options: two SPs per storage system and at least two HBAs per server. For shared switched storage, two switch fabrics are required.

For unshared direct storage, a server may have one or two HBAs.

Components for Shared Switched Storage

The minimum hardware configuration required for shared switched storage is two servers, each with two host bus adapters, two Fibre Channel switch fabrics with one switch per fabric, and two SPs per storage system. Two SPS units (standby power supplies) are also required. You can use more servers (up to 15 are allowed), more switches per fabric, and more storage systems (up to four are allowed).

Components for Shared-or-Clustered Direct Storage

The minimum hardware required for shared-or-clustered direct storage is two servers, each with two host bus adapters, and one storage system with two SPs. You can use more storage systems (up to four are allowed).

Components for Unshared Direct Storage

The minimum hardware required for unshared direct storage is one server with two host bus adapters and one storage system with two SPs. You can choose more storage systems (up to four are allowed).

Hardware Dimensions and Requirements

The SPE and SPS units, DAE2s, and switches require the following amounts of space and power in a standard 19-inch cabinet.

Table 6-1 CX-Series Hardware Dimensions and Requirements

Device	Dimensions	Vertical Size	Weight	Power	VA Rating
SPE	17.8 cm (7.0 in) x 45.1 cm (17.75 in) x 60.3 cm (23.75 in)	4 U (4 NEMA units at 1.75 in)	33 kg (72.6 lb) with rails	100 V-240 V ac. At 200 V, 2.6 A max	520
DPE2	13.3.4 cm (5.25 in) x 45.1 cm (17.75 in) x 60.3 cm (23.75 in)	3 U	44.7 kg (98.3 lb) with SPs, 15 disks, and mounting rails.	100 V-240 V ac. At 200 V, 2.9 A max	580
DAE2	17.8 cm (7.0 in) x 45.1 cm (17.75 in) x 60.3 cm (23.75 in)	3 U	43.3 kg (95.3 lb) with LCCs, 15 disks, and mounting rails	100 V-240 V ac. At 200 V, 2.0 A max	400
SPS Tray	4.44 cm(1.75 in) x 45.1 cm (17.75 in) x 60.3 cm (23.75 in)	1 U	26 kg (57.15 lb) with tray (two SPS units at 25 lb each)	100 V-240 V ac. At 200 V, 2.0 A max for two SPS units	40
Switch, 32-port	8.88 cm (3.5 in)	2 U	19.5 kg (43 lb)	100 V-240 V ac. At 200 V, 2.0 A max.	400
Switch, 16-port	4.44 cm (1.75 in)	1 U	11.8 kg (26 lb)	100 V-240 V ac. At 200 V, 1.25 A max.	250

The total of the device VA ratings may not exceed 4800 VA per cabinet.

Cabinets for Enclosures

Pre-wired 19-inch-wide cabinets, ready for installation, are available in the following dimensions to accept storage-system enclosures.

Vertical Space	Exterior Dimensions	Comments
179 cm or 70 in (40 NEMA units or U; one U is 1.75 in)	Height: 190 cm (75 in) Width: 60 cm (24 in) Depth: 91 cm (36 in) plus service clearances, which are 76 cm (30 in) at front and 90 cm (36 in) at rear. Weight (empty) with ac cables: 152.3 kg (335 lb)	Both cabinets accept combinations of: SPEs at 4 U, DPE2s at 3 U DAE2s at 3 U, SPS units at 1 U, Switches at 1 or 2 U Both require 200–240 volts ac. at 50/60 Hz. There are two power strips, each with 21 IEC-320 C14 outlets (40 U cabinet) or 12 outlets (39 U cabinet). Filler panels of various sizes are available. Plug options are L6–30P and IEC 309 30 A.

As an example, a shared switched storage system with eight DAE2s (120 disks) has the following requirements.

Category	Requirement
Vertical cabinet space	31 U, including one SPS (1 U), one SPE (4 U), eight DAE2s (8*3 U equals 24U), and two 16-port switches (2U) for a total of 31 U.
Weight	603 kg (1,326 lb) including the cabinet (173.6 kg), SPE (33 kg), SPS units (26 kg), eight DAE2s (346.5 kg(, and two switches (23.6 kg)
Power	4,260 VA max, including the SPE (520 VA), SPS units (40 VA), eight DAE2s (8 * 400 VA equals 3200 VA)., and two switches (500 VA).

Data Cable and Configuration Guidelines

CX-Series storage systems require optical cable between servers, switches, and SPs. The cabling between SPE or DPE2 and DAE2 enclosures (1 m, 3.3 feet) is copper.

The standard for cable is 50 microns. EMC does not recommend mixing 62.5 micron and 50 micron optical cable in the same link. In certain situations, however, you can add a 50 micron adapter cable to the end of an already installed 62.5 micron cable plant. Contact your customer service representative for details.

Generally, you should minimize the number of cable connections, since each connection degrades the signal slightly and shortens the maximum distance the signal can carry.

Table 6-2 Data Cable Types and Sizes

Length	Typical Use
1 m (3.3 ft) copper	In a single cabinet, connects the SPE or DPE2 to a DAE2. In the same cabinet, connects a DAE2 to a DAE2.
5 m (16.5 ft) equalized copper 10 m (33 ft) equalized copper	Cabinet to cabinet, connects the SPE or DPE2 to a DAE2. Cabinet to cabinet, connects a DAE2 to a DAE2.
3 m (10 ft) optical, LC-LC	In a single cabinet, connects the SPE or DPE2 to switches.
10 m (33 ft) optical, LC-LC	Cabinet to cabinet, connects the SPE or DPE2 to a DAE2 or a DAE2 to a DAE2.
10 m (33 ft) optical, LC-LC 30 m (100 ft) optical LC-LC 50 m (164 ft) optical, LC-LC	In a room, connects the SPE or DPE2 to switches or connects the SPE to an HBA. (Adapter must support optical cable.)
100 m (328 ft) optical, LC-LC	In a room, connects the SPE to switches or connects the SPE to an HBA. (Adapter must support optical cable.)
300 m (1000 ft, 0.2 mi) optical, LC-LC	In a room, connects SPE to switches or connects the SPE to an HBA. (Adapter must support optical cable.)
10 m (33 ft) optical, LC-SC 30 m (100 ft) optical LC-SC 50 m (164 ft) optical, LC-SC 100 m (328 ft) optical, LC-SC 300 m (1000 ft,.2 mi) optical, LC-SC	LC-SC cables are used on the EMC 1-Gbit Fibre Channel products. You can use LC-SC cables to connect an HBA to a switch, a switch to the SPE, or an HBA to the SPE.
The minimum bend radius for optical cables is 3 cm/1.2 in. For the 50-micron cable operating at 2.125Gbit/s the recommended maximum distance is 300 meters and at 1.0625 Gbit/s, the recommended, maximum distance is 500 meters. For the 62.5-micron cable operating at 2.125 Gbit/s the recommended maximum distance is 160 meters and at 1.0625 Gb/s, the recommended, maximum distance is 300 meters. All of these maximum lengths assume no more than three connections/splices in the link.	

Component planning diagrams and worksheets follow.

Hardware Planning Worksheets

Following are worksheets to note hardware components you want. Some installation types do not have switches or multiple servers.

Cable Planning Template

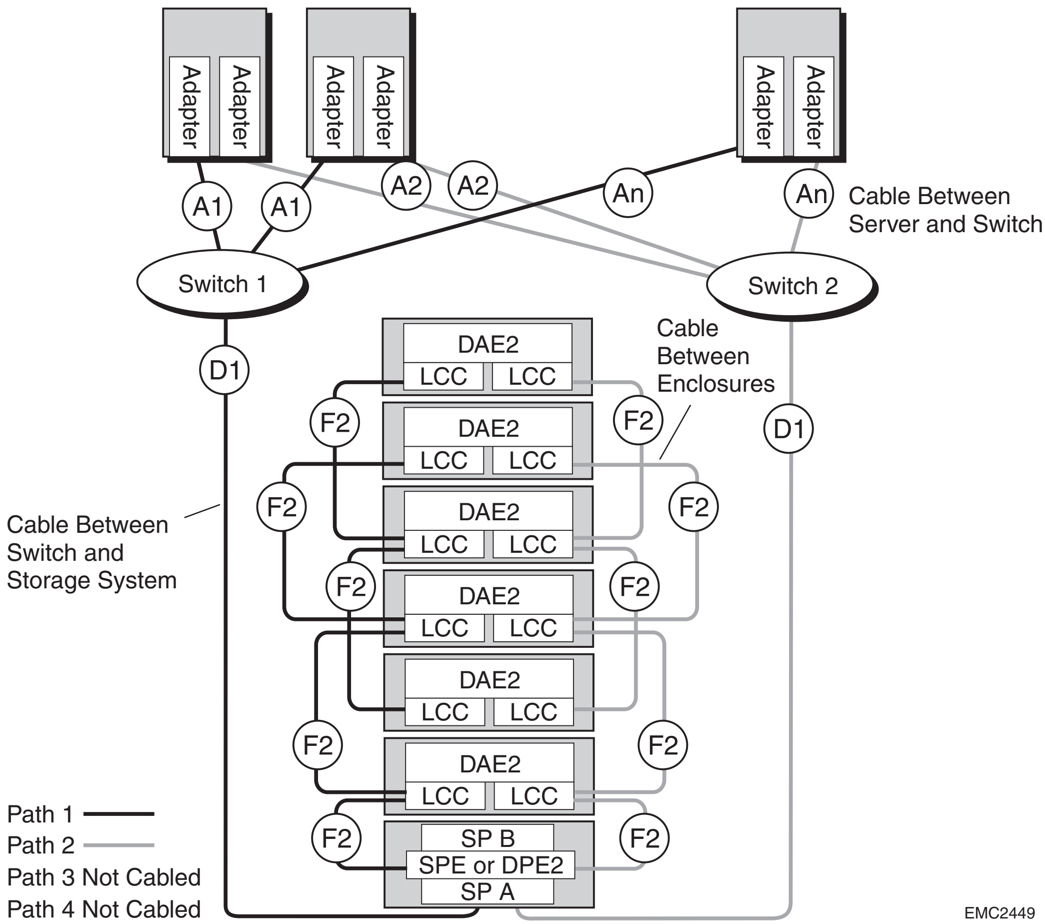


Figure 6-5 Cable Planning Template — Shared Storage System

The cable identifiers apply to all storage systems.

Blank Hardware Component Worksheet

Hardware Component Worksheet

Number of servers:____ Adapters in servers:____ Switches: 32-port:____16-port:____

SPEs or DPE2s:_____Rackmount cabinets:___

DAE2s:_____ ....LCCs:_____

Cables between server and switch port- Cable A
Cable $A_{1,}$ Optical: Number:____ .... Length________m or ft
Cable $A_{2,}$ Optical: Number:____ .... Length________m or ft
Cable A_n,Optical: Number:____ .... Length________m or ft

Cables between switch ports and storage-system SP ports - Cable D
Cable $D_{1,}$ CX600, up to 4/SP; CX400, up to 2/SP, Optical:Number:_____ .... Length________m or ft
Cable D_m CX600, up to 4/SP; CX400, up to 2/SP, Optical:Number:_____ .... Length________m or ft

Cables between enclosures - Cable F
Cable F_2 (Copper, 1.0 m): Number (2 per DAE2): ______

Sample Cable Diagram

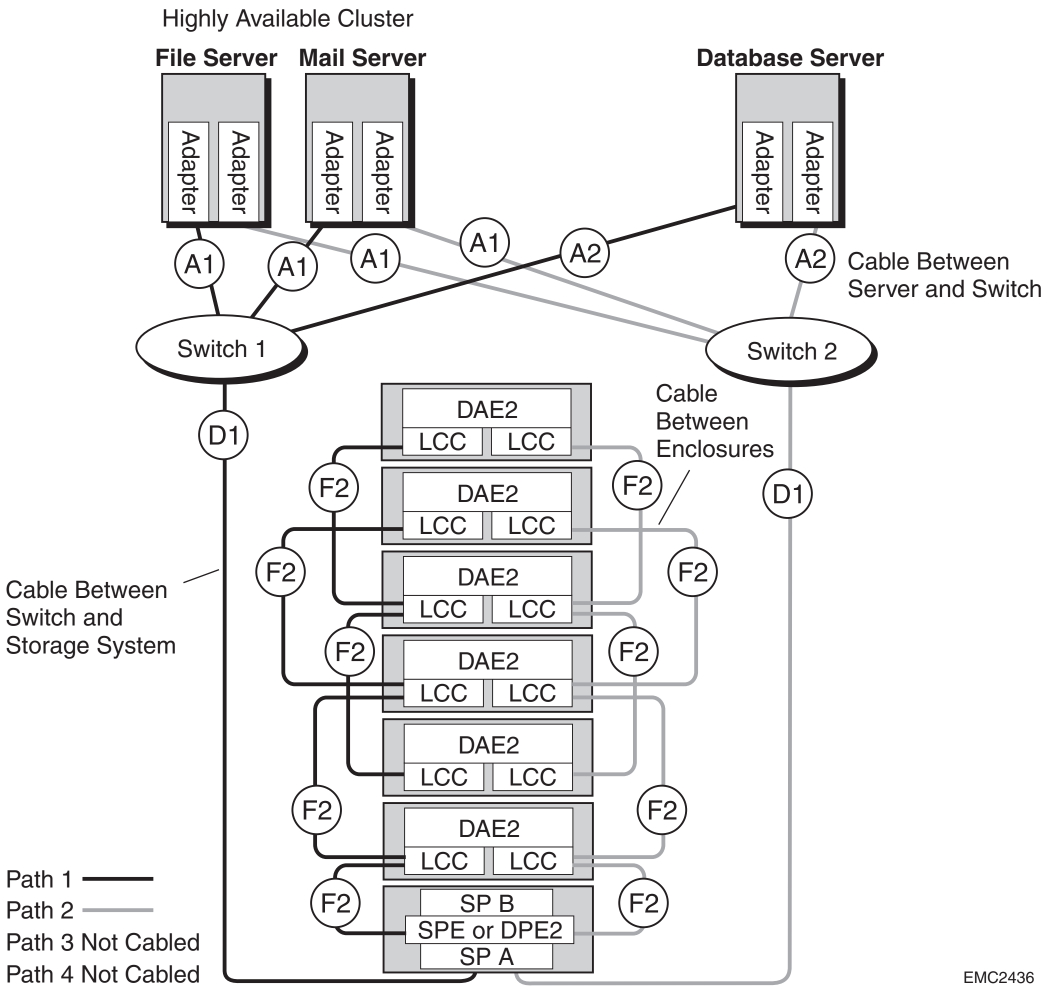


Figure 6-6 Sample Shared Storage Installation

Sample Hardware Component Worksheet

Hardware Component Worksheet

Servers:__*3*_ Adapters in servers:__*6*_ Switches: 32-port:____16-port:__*2*__

SPEs or DPE2s:__ *1*___Rackmount cabinets:_*1*__

DAE2s:__*6*___ .... LCCs:___*12*__

Cables between server and switch - Cable A

Cable A_1, Optical: Number:__*4*__ .... Length____*10*_m or ft

Cable A_2, Optical: Number:__*2*__ .... Length___*33*_m or ft

Cable A_n, Optical: Number:_____ .... Length________m or ft

Cables between switch ports and storage-system SP ports - Cable D

Cable D_1, CX600, up to 4/SP; CX400, up to 2/SP, Optical:Number:__*2*___ .. Length__*33*____m or ft

Cable D_m CX600, up to 4/SP; CX400, up to 2/SP, Optical:Number:_____ ..... Length________m or ft

Cables between enclosures

Cable F_2 (Copper, 1.0 m): Number (2 per DAE): ___*12*_

What Next? If you have completed the worksheets to your satisfaction, then you are ready to order some of this equipment.

7

Storage-System Management

This chapter explains the applications you can use to manage storage systems from servers. Topics are

Introducing Navisphere Management Software

EMC ControlCenter Navisphere management software lets you bind and unbind disks, manipulate caches, examine storage-system status and logged events, transfer control from one SP to another, and examine events recorded in storage-system event logs. It features a graphical user interface (GUI) and extensive on-line help.

Navisphere management software is web-based and allows any computer connected to the internet to manage storage systems.

Navisphere software products run entirely in storage-system SPs. The SPs are grouped in storage-system domains. You can manage each storage system in a domain though any SP in the domain that has Manager installed. Using a web browser on any connected host, you point at such an SP, log in, and, depending on your privilege level, can then administrate, manage, or monitor storage systems in the domain. You create and modify domains and log-in accounts using Manager. The web-based nature of Navisphere lets it work with any operating system that can support a web browser.

Navisphere Products

The Navisphere products are

- Navisphere Manager (called Manager), which lets you manage multiple storage systems on multiple servers simultaneously. Manager includes an Event Monitor that checks storage systems for fault conditions and can notify you and/or customer service if any fault condition occurs.
- Navisphere Analyzer, which lets you measure, compare, and chart the performance of SPs, LUNs, and disks.
- Navisphere Integrator, which provides an interface between Navisphere products and HP OpenView®, CA Unicenter®, and Tivoli® interfaces.
- Navisphere CLI (Command Line Interface), which runs on each host. The CLI lets you bypass the Manager GUI and type management commands directly to storage systems. A major benefit offered by the CLI is the ability to write command scripts to automate management operations.

Using Navisphere Software

Each storage system that runs web-based Manager software can be assigned to a storage domain — a group of storage systems on an internet or intranet you define using Manager.

A server connected to storage systems can run Windows, a UNIX operating system such as Linux®, Sun Solaris®, or NetWare. Servers connected to the SAN (Storage Area Network) can run different operating systems.

You can create one or more domains for any installation, provided that any connected storage system can be defined in only one domain. Each storage domain must have at least one member node that has Manager installed.

Each storage system in the domain is accessible from any other in the domain. Using an internet browser, you point at a storage system that has Manager installed. The security software then prompts you to log in. After logging in, depending on the privileges of your account, you can monitor, manage and define user accounts for any system in the domain. Storage systems outside the domain are not viewable from the domain.

You can run the internet browser on any supported PC or laptop with a network controller and browser software. At least one storage system — ideally at least two for higher availability — in a domain must have Manager installed.

The privilege of monitoring, managing, and creating accounts on systems in a domain depends on the type of account used when you log in.

The following figure shows an intranet that connects nine storage systems. It shows two domains, a US Division domain with five storage systems (four systems on SANs) and a European Division with four storage systems. The 13 servers that use the storage may be connected to the same or a different network, but the intranet shown is the one used to manage the storage systems.

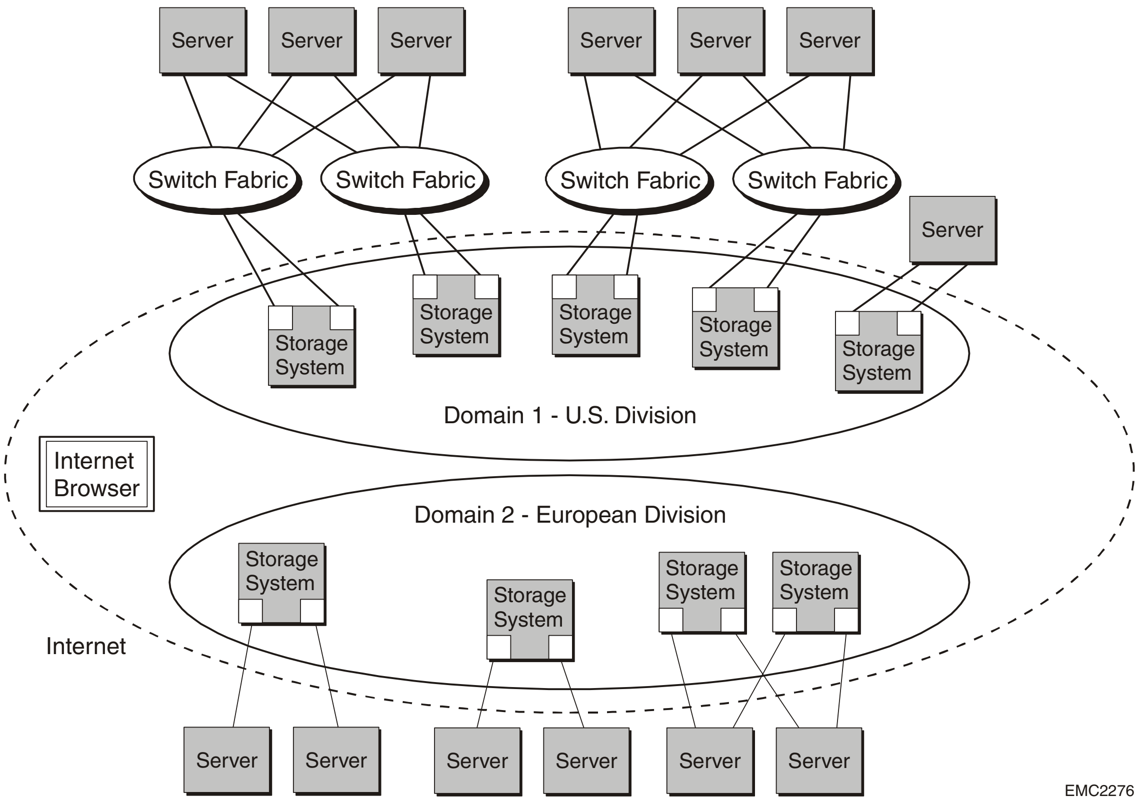


Figure 7-1 Storage Domains on the Internet

Navisphere Management Worksheet

The following worksheet will help you plan your Navisphere storage management arrangement.

A sample completed worksheet follows the blank worksheet.

Blank Storage Management Worksheet for Navisphere 6.X

Navisphere Storage Management Worksheet

Write the hostnames, operating systems, and connected storage system numbers of all servers that will use the storage.

Server:___________Op sys: _________Stor sys number/name______

Server:___________Op sys: _________Stor sys number/name______

Server:___________Op sys: _________Stor sys number/name______

Server:___________Op sys: _________Stor sys number/name______

Server:___________Op sys: _________Stor sys number/name______

Server:___________Op sys: _________Stor sys number/name______

Server:___________Op sys: _________Stor sys number/name______

Server:___________Op sys: _________Stor sys number/name______

Write the domain number/name, storage system number/name (number corresponding to list above), and mark the appropriate product box. One domain can have many storage systems and needs only one copy of a Navisphere product.

Domain number/name:________Stor sys number/name________ Navisphere products: Manager❑ Analyzer❑

Domain number/name:________Stor sys number/name________ Navisphere products: Manager❑ Analyzer❑

Domain number/name:________Stor sys number/name________ Navisphere products: Manager❑ Analyzer❑

Domain number/name:________Stor sys number/name________ Navisphere products: Manager❑ Analyzer❑

Domain number/name:________Stor sys number/name________ Navisphere products: Manager❑ Analyzer❑

Domain number/name:________Stor sys number/name________ Navisphere products: Manager❑ Analyzer❑

Domain number/name:________Stor sys number/name________ Navisphere products: Manager❑ Analyzer❑

Domain number/name:________Stor sys number/name________ Navisphere products: Manager❑ Analyzer❑

Sample Storage Management Worksheet for Navisphere 6.X

Navisphere Storage Management Worksheet

Write the hostnames, operating systems, and connected storage system numbers of all servers that will use the storage.

Server: ***cpc1033*** Op sys: ***Windows 2000*** Stor sys number/name ***1***

Server: ***cpc1034*** Op sys: ***Windows 2000*** Stor sys number/name ***1***

Server: ***cpc1035*** Op sys: ***Windows 2000*** Stor sys number/name ***1***

Server: ***cpc1036*** Op sys: ***Windows 2000*** Stor sys number/name ***2 3***

Server: ***cpc1040*** Op sys: ***Solaris*** Stor sys number/name ***4***

Server: ***cpc1041*** Op sys: ***Solaris*** Stor sys number/name ***4***

Server: ***cpc1042*** Op sys: ***Solaris*** Stor sys number/name ***5***

Server: ***cpc1043*** Op sys: ***Linux*** Stor sys number/name ***5***

Write the domain number/name, storage system number/name (number corresponding to list above), and mark the appropriate product box. One domain can have many storage systems and needs at least one copy of a Navisphere product.

Domain number/name: ***1*** Stor sys number/name ***1*** Navisphere products: Manager☒ Analyzer☒

Domain number/name: ***1*** Stor sys number/name ***2*** Navisphere products: Manager❑ Analyzer❑

Domain number/name: ***1*** Stor sys number/name ***3*** Navisphere products: Manager❑ Analyzer❑

Domain number/name: ***2*** Stor sys number/name ***4*** Navisphere products:: Manager☒ Analyzer☒

Domain number/name: ***2*** Stor sys number/name ***5*** Navisphere products:: Manager❑ Analyzer❑

Domain number/name:________ Stor sys number/name________ Navisphere products: Manager❑ Analyzer❑

Domain number/name:________ Stor sys number/name________ Navisphere products:: Manager❑ Analyzer❑

Domain number/name:________ Stor sys number/name________ Navisphere products: Manager❑ Analyzer❑

Index