# *SPARCstorage Array Configuration Guide*

**Sun Microsystems Computer Company**
A Sun Microsystems, Inc. Business
2550 Garcia Avenue
Mountain View, CA 94043 U.S.A.
415 960-1300 FAX 415 969-9131

Part No.: 802-2041-10
Revision A, March 1995

Please
Recycle

Adobe PostScript

# *Contents*

# *Figures*

# *Tables*

# Product Description 1

## 1.1 SPARCstorage Array Model 100 System Features

The SPARCstorage™ Array Model 100 Series provides SCSI disk expansion for SPARCstation™, SPARCserver™ and SPARCcenter™ systems. It features three drive trays—each tray configured with up to ten half-height, single-connector, 3.5-inch disk drives.

The SPARCstorage Array Model 100 Series, shown in Figure 1-1, features front and rear component accessibility.

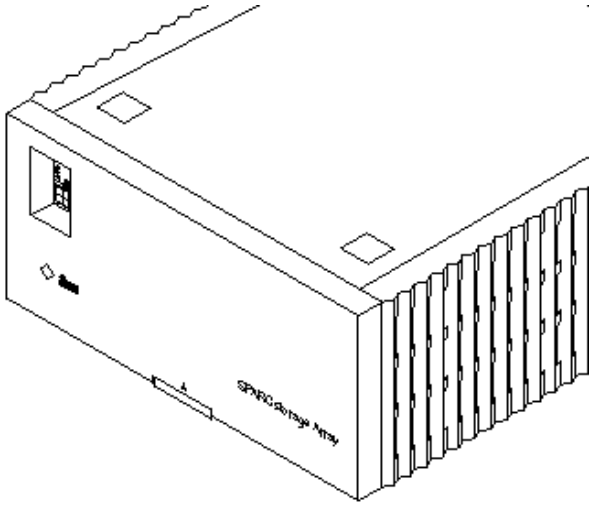


*Figure 1-1* SPARCstorage Array Model 100 Series

Internal system components (see Figure 1-2) are:

- Power supply
- Backplane
- Array controller, with
  - Fibre Channel Optical Module (FC/OM)
  - Battery module
- Fan tray
- Drive trays (3)

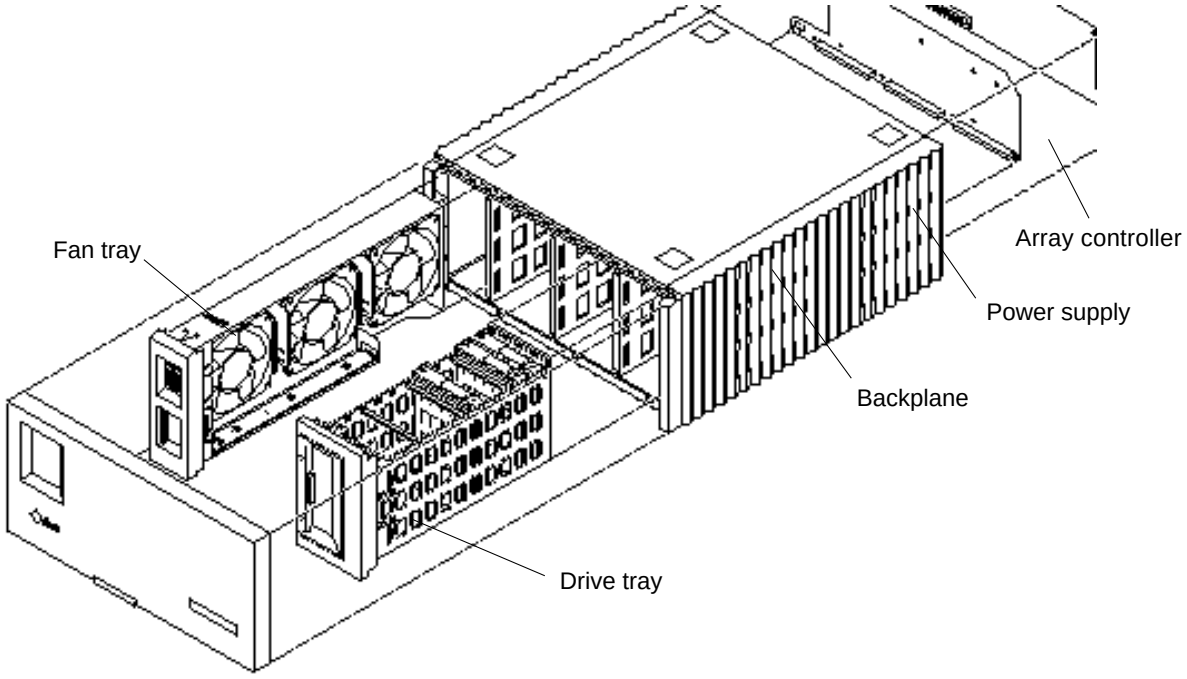



*Figure 1-2* Internal Components

## 1.2 SPARCstorage Array Model 200 System Features

The SPARCstorage Array Model 200 Series is a rackmount disk array controller. Up to six Differential SCSI disk trays may be connected to it. See Figure 1-3.

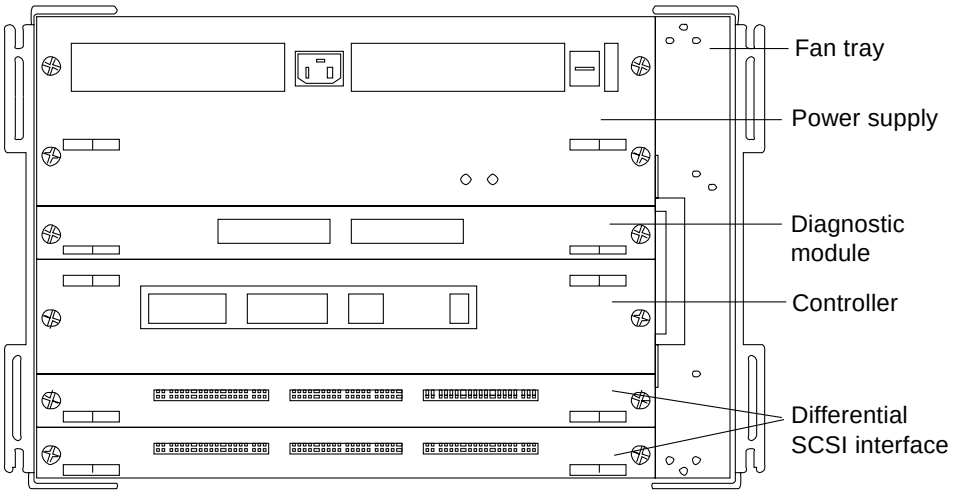



*Figure 1-3* SPARCstorage Array Model 200 Series Subsystem

Functional parts of the SPARCstorage Array Model 200 Series hardware include:

- Removable power supply
- Removable LCD-display diagnostic module
- Removable disk array controller module
- Removable Differential SCSI interface modules (two) with three differential SCSI ports per module
- Removable fan tray
- Internal backplane with attached power and signal buses
- Rack-mountable chassis

## 1.3 General Information for SPARCstorage Array Model 100 and 200 Systems

The SPARCstorage Array connects to a host system using fibre optic cables. These cables connect to a Fibre Channel Optical Module (FC/OM) mounted on a Fibre Channel SBus (FC/S) card on the host system side, and to an FC/OM mounted on the array controller on the SPARCstorage Array side.

You can have a maximum of two FC/OMs per FC/S card on the host system side and two FC/OMs on the array controller on the SPARCstorage Array side. Because of the number of FC/OMs available on both the host system and the SPARCstorage Array sides, you have several hardware configuration options when connecting a SPARCstorage Array to a server. Chapter 2, “Hardware Configurations,” contains all the information for you to choose the hardware configuration that best suits your needs.

The SPARCstorage Array also has storage management software which offers several software options. Chapter 3, “Software Configurations,” contains all the information for you to choose the software configuration that best suits your needs.

# *Hardware Configurations* 2

You have several options when it comes to connecting a SPARCstorage Array to a host system. The two factors that determine which options you should select are:

- The number of SPARCstorage Arrays in the configuration
- The number of host systems in the configuration

When reviewing the following options, keep in mind that every SPARCstorage Array comes standard with one FC/OM and one fibre optic cable and that every FC/S card comes standard with one FC/OM.

# 2

## 2.1 Option 1

*Table 2-1* Option 1

| Option | Hardware Needed | Notes |
|---|---|---|
| One SPARCstorage Array on a single host SBus slot | Order:<br>- One SPARCstorage Array<br>- One FC/S card | Simplest, most common configuration |

Figure 2-1 shows how you would connect the SPARCstorage Array to a host system using this option.

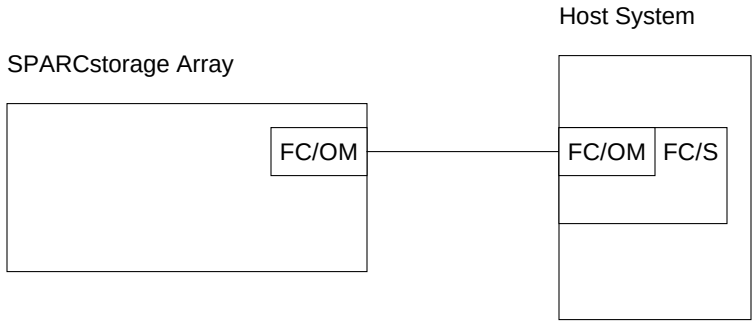



*Figure 2-1* Option 1 Connection

## 2.2 Option 2

*Table 2-2* Option 2

| Option | Hardware Needed | Notes |
|---|---|---|
| Two SPARCstorage Arrays on a single host SBus slot | Order:<br>- Two SPARCstorage Arrays<br>- One FC/S card<br>- One FC/OM | - Highest capacity per SBus slot<br><br>- Connecting 2 arrays on one SBus slot could possibly cause the FC/S to be a performance bottleneck on extremely demanding sequential I/O applications<br><br>- Lower cost than using 2 FC/S cards for two arrays |

Figure 2-2 shows how you would connect the SPARCstorage Arrays to a host system using this option.

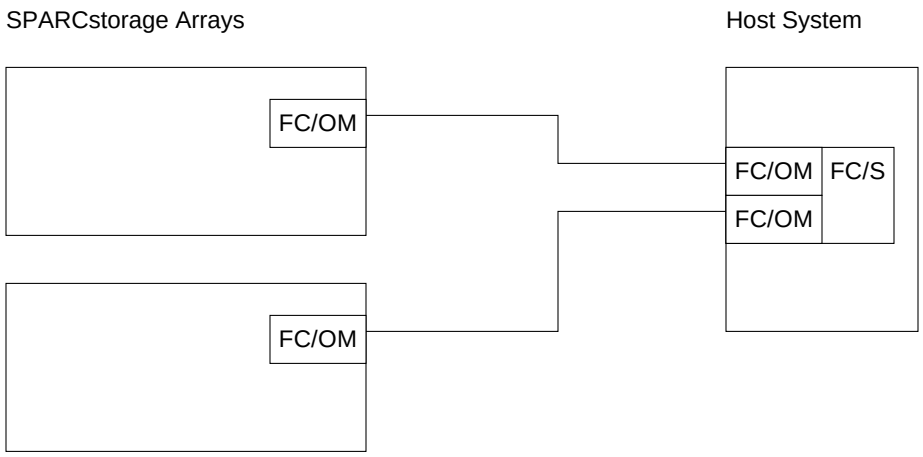



*Figure 2-2* Option 2 Connection

## 2.3 Option 3

*Table 2-3* Option 3

| Option | Hardware Needed | Notes |
|---|---|---|
| Dual-host connection to a single SPARCstorage Array | Order:<br>- One SPARCstorage Array<br>- Two FC/S cards<br>- One FC/OM<br>- One fibre optic cable | Two possible reasons for this configuration:<br>- Static dual-porting. Two systems connect to an array, but each disk in the array is "owned" by only one system at a time. One host system functions as a backup for the other host. * †<br><br>- Sharing an array between two systems. Each system uses the reserve command to reserve a subset of the disks so that the other system is not allowed to access them. † |

*. Software to control and manage this type of high availability is not supplied with the SPARCstorage Array and must be purchased separately.

†. Dual host configurations are potentially dangerous and can result in data corruption if both systems access the same disk at the same time.

Figure 2-3 shows how you would connect the SPARCstorage Array to the host systems using this option.

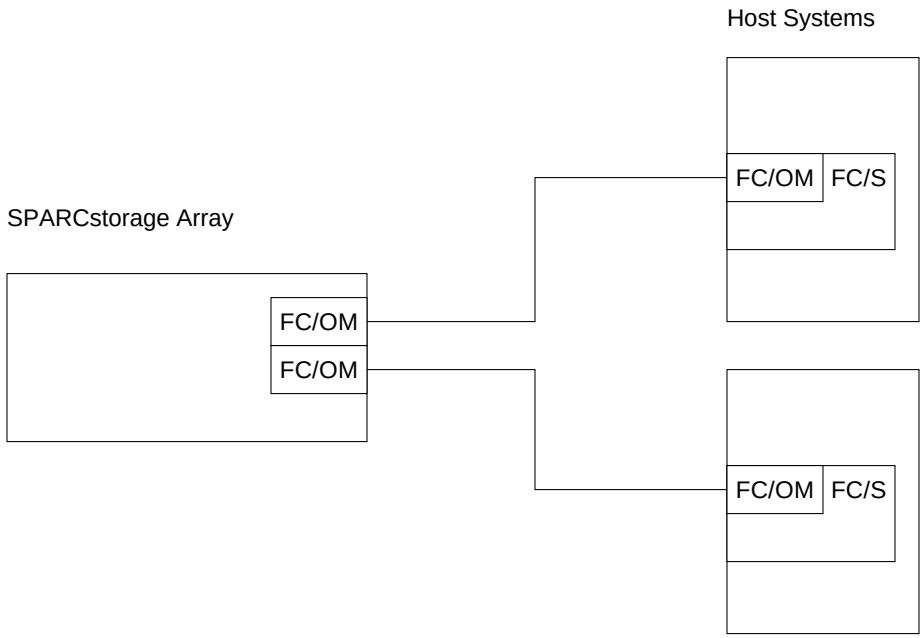



*Figure 2-3* Option 3 Connection

## 2.4 Option 4

*Table 2-4* Option 4

| Option | Hardware Needed | Notes |
|---|---|---|
| Fully redundant, dual-hosted SPARCstorage Array | Order:<br>- Two SPARCstorage Arrays<br>- Four FC/S cards<br>- Two FC/OMs<br>- Two fibre optic cables | - This configuration is for the highest levels of availability where all hardware components are redundant.<br>- When each array is set up so that it holds a complete copy of the data for both host system A and host system B, then any component in either array can fail with no loss of data service to the host systems * † |

*. Software to control and manage this type of high availability is not supplied with the SPARCstorage Array and must be purchased separately.

†. Dual host configurations are potentially dangerous and can result in data corruption if both systems access the same disk at the same time.

Figure 2-4 shows how you would connect the SPARCstorage Arrays to the host systems using this option.

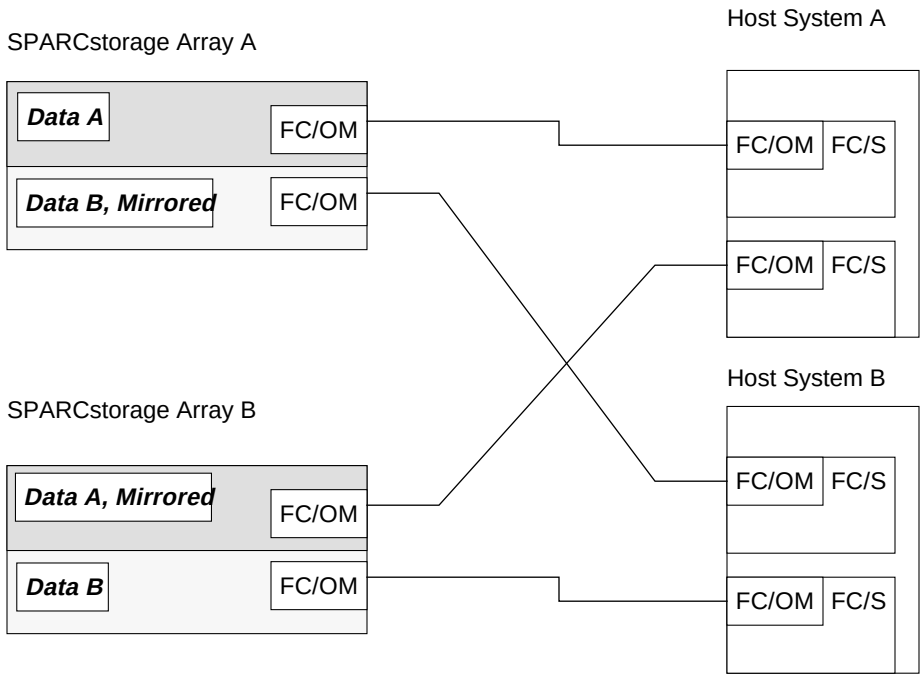



*Figure 2-4* Option 4 Connection

# 2

# *Software Configurations* 3

The SPARCstorage Volume Manager software can be configured for different levels of RAID, which stands for Redundant Array of Independent Disks. RAID is defined as a disk array in which part of the storage capacity is used to hold redundant information about user data that is stored on the remainder of the disk space capacity. The redundant information enables the user data to be regenerated if one of the array's member disks or the access path to it fails.

Following are the data layouts offered by the SPARCstorage Volume Manager storage management software:

- Independent disks
- RAID Level 0 (also known as *striping*)
- RAID Level 1 (also known as *mirroring*)
- RAID Level 0 + 1 (also known as *mirroring striped disks*)
- RAID Level 5

Note that the SPARCstorage Volume Manager storage management software allows all these data layouts to be mixed at the same time within the same array, with no restrictions on how the layouts are mixed and matched. These data layouts may be spread across multiple SPARCstorage Arrays. This flexibility allows you to configure your SPARCstorage Arrays to meet your specific requirements.

These data layouts are explained in greater detail in the following sections.

# 3

## 3.1 Independent Disks

A data layout using independent disks has the following characteristics:

- Best for most applications
- No redundant data protection
- Lowest cost per usable Mbyte

## 3.2 Striping (RAID 0)

Striping data spreads the data out over more than one physical disk so that the data is laid out evenly across the disks.

If you were striping data over three physical disks, the first block of data would go on the first disk, the second block would go on the second disk, and the third block would go on the third disk. The fourth block of data would then go back on the first disk, the fifth block would go on the second disk, and so on. Figure 3-1 shows how data is spread out over three disks by striping the data.

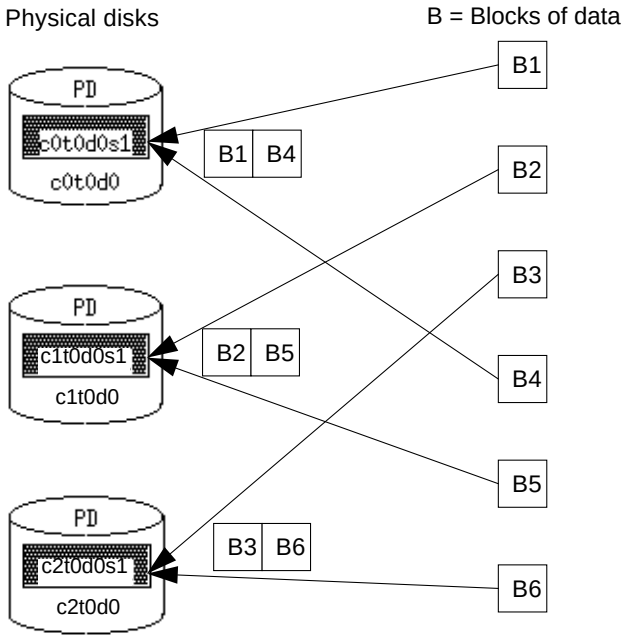



*Figure 3-1* Example of Striped Data

A data layout using striping has the following characteristics:

- Spreads data across multiple disk spindles for better performance
- Can be tuned to optimize either random or sequential I/O performance
- No redundant data protection, lower reliability than independent disks
- Same low cost per usable Mbyte as independent disks

## 3.3 *Mirroring (RAID 1)*

Mirroring data copies the same data onto two or more separate physical disk drives. This is useful if you want to make sure data is available even if one physical disk fails; the data can be retrieved from the other physical disk that has the copy of the original data. In addition, it can sometimes improve read performance on busy data sets.

If you were mirroring data to two physical disks, you would send the same data to both disks. Figure 3-2 shows how data would be mirrored to two physical disks.

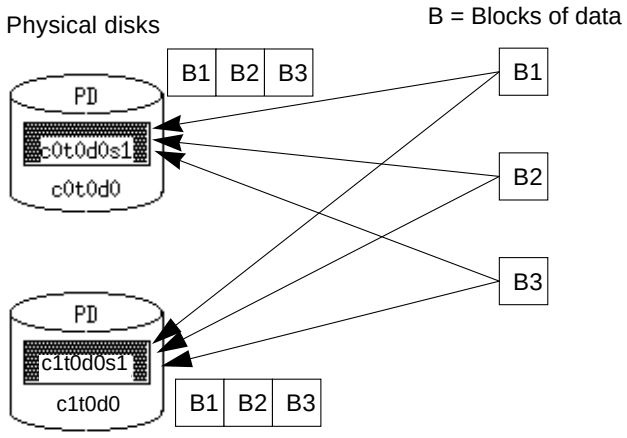



*Figure 3-2* Example of Mirrored Data

A data layout using mirroring has the following characteristics:

- Duplicate copies of data, so if a disk fails, data is still available and applications keep running
- Performance roughly the same as independent disks
- Highest cost per usable Mbyte

## 3.4 Mirroring Striped Disks (RAID 0 + 1)

Mirroring striped disks consists of two separate operations:

1. Data is striped across several physical disks.
2. The data from the striped disks is mirrored on separate physical disks.

This is useful if you want to get data written to or read from physical disks quickly and also want to make sure that data is available even if a disk fails.

Figure 3-3 shows how you would stripe and mirror data over four physical disks.

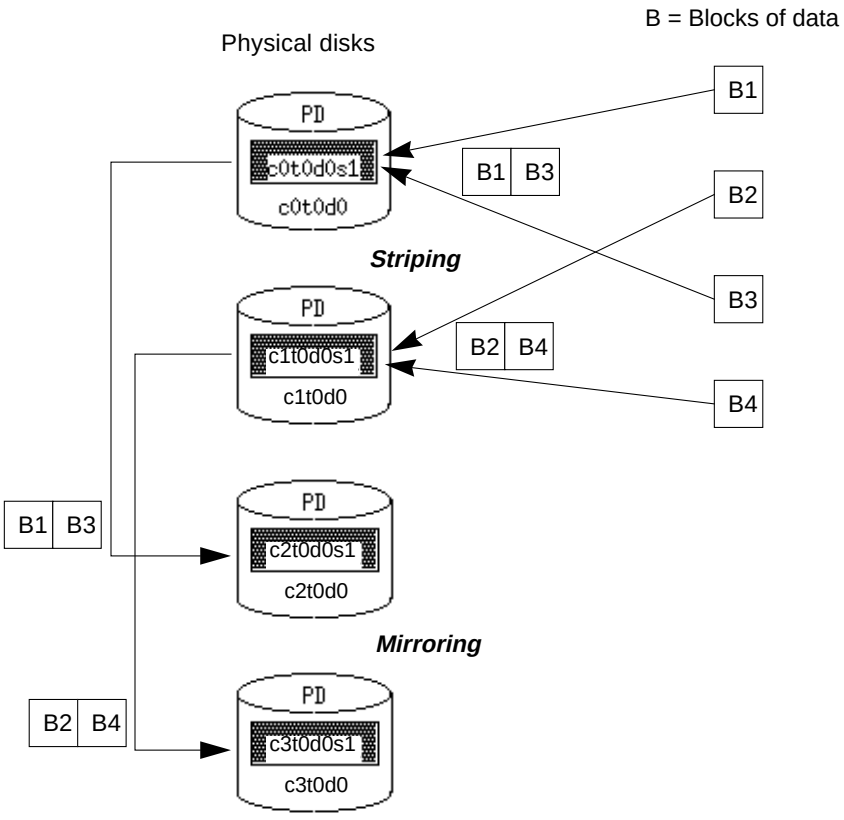



*Figure 3-3* Example of Mirroring Striped Data

A data layout using mirroring striped disks has the following characteristics:

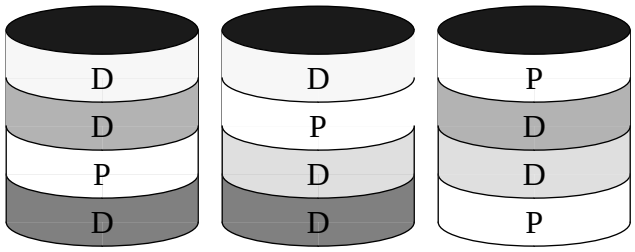
- Duplicated copies of striped data remain available even if a disk fails
- Combines performance of striping with data protection of mirroring
- Has cost per usable Mbyte disadvantages of mirroring

## 3.5 RAID 5

A RAID 5 configuration is similar to a striped configuration in that data is spread out evenly across several disks in a system. However, a RAID 5 configuration adds a *parity* block to the stripe to provide redundancy. This parity block contains the result of an exclusive OR (XOR) procedure done on the data in the data stripes. If the data on one of the disks in the RAID 5 configuration becomes inaccessible due to a hardware or software failure, data can be restored by XORing the contents of the remaining data disks with the parity disk. The data on the failed disk can be rebuilt from the output of the XOR process.

When a write is performed on disks set up in a RAID 5 configuration, the data is written to all but one of the disks; the remaining disk then gets the parity information written to it. The next write performed on the disks in a RAID 5 configuration would work the same way, except that the parity information would be written to a different disk than that used in the first write. That way, the parity information gets spread across all the disks in the configuration, so that if one disk fails, the data from that disks can be restored using the parity information written on the remaining disks in the configuration.

For example, if three disks were used in a RAID 5 configuration, data written to those disks in this manner: for the first set of writes, data would be written to the first two disks in the configuration and the parity information would be written to the third disk. On the second set of writes to the disks, the data would be written to the first and third disks, and the parity information would be written to the second. On the third set of writes, the data would be written to the second and third disks, and the parity information would be written to the first. The fourth write would be performed as the first write was, with data on the first and second disks and parity information on the third. Figure 3-4 shows how RAID 5 information would be written to disks using this setup.




D = Data Stripe Unit
P = Parity Stripe Unit

*Figure 3-4* Graphical Representation of RAID 5 Writes

A data layout using RAID 5 has the following characteristics:

- Spreads data across multiple disk spindles for better performance
- Data remains available even if a disk fails because of parity
- Slightly higher cost per usable Mbyte as independent disks or striping

## 3.6 Determining Which RAID Level to Use

The different RAID levels have varying degrees of benefits and liabilities in each of the following areas:

- Cost — number of disks required
- Data reliability — ability to get data even if disk fails
- Data transfer capacity — ability to move large amounts of sequential access data rapidly
- I/O rate — ability to satisfy large numbers of random I/O requests per unit time

Refer to Table 3-1 for to determine which RAID level best suits your needs.

*Table 3-1* Comparisons of RAID Levels

❶ — much better than single disk ❸ — equal to single disk ❹ — slightly worse than single disk
❷ — slightly better than single disk ❺ — much worse than single disk

| RAID Levels | Cost | Data Reliability | Data transfer capacity | | I/O Rate | |
|---|---|---|---|---|---|---|
| | | | Writing Data | Reading Data | Write | Read |
| 0 | ❸ | ❺ | ❶ | ❶ | ❶ | ❶ |
| 1 | ❺ | ❶ | ❹ | ❷ | ❹ | ❷ |
| 0 + 1 | ❺ | ❶ | ❶ | ❶ | ❶ | ❶ |
| 5 | ❹ | ❶ | ❶ | ❶ | ❶ | ❶ |

## 3.7 Other Facts on RAID and the SPARCstorage Array

- You can use any combination of independent disks or disks using RAID levels 0, 1, 0 + 1, or 5 in the same array.
- RAID groups may even span multiple arrays.
- Optional hot spares are automatically swapped in to replace any failed disk in a RAID 0 + 1 or 1 group without human intervention. Hot spares are extra disk drives in the array that are powered up and ready to use. If a drive in a RAID 0 + 1 or 1 group fails, the Volume Manager detects the failure and automatically rebuilds the data from the failed drive onto a hot spare drive.
- There can be one or more hot spares per array. A hot spare drive can be tied to one or more particular drives, one or more particular RAID groups, or may be designated as a spare for any drive in any RAID group.
- The combination of hot spares with warm pluggability provides new levels of high availability and administrative flexibility. Hot spares allow maintenance to be deferred for long periods of time and warm pluggability means that the array downtime is very short when a failed drive is eventually changed.
- Since RAID is not ideal for many situations, the SPARCstorage Array allows you to configure disk drives as regular, independent disks, rather than forcing you to use RAID.

- RAID stripe sizes are adjustable to optimize for either random or sequential I/O patterns.
- Striped data organizations (RAID 0, 0 + 1 and 5) can be tuned to optimize for either random or sequential I/O performance.
- To optimize for random performance, striping is used to evenly balance the I/O load across disk spindles in the RAID. This is done by setting the stripe width as large or larger than the typical application I/O request. For example, if the typical I/O request is 8 Kbytes, setting the stripe width to 64 Kbytes might be appropriate. This tends to evenly distribute I/O request across all the disk drives in the RAID, and each I/O request is serviced by a single disk drive.
- Sequential performance is optimized when data is spread out so that each application I/O request spans all the drives in the RAID group. This requires setting the stripe width so that it is small relative to the size of the typical I/O request. For example, in a RAID group four data disks wide, if typical application I/O size is 8 to 16 Kbytes, a stripe width of 2 Kbytes may be best.

# *Index*

## Reader Comments

We welcome your comments and suggestions to help improve this manual. Please let us know what you think about the *SPARCstorage Array Configuration Guide*, part number *802-2041-10*.

■ The procedures were well documented.

| Strongly Agree | Agree | Disagree | Strongly Disagree | Not Applicable |
|---|---|---|---|---|
| ❑ | ❑ | ❑ | ❑ | ❑ |

Comments _____

■ The tasks were easy to follow.

| Strongly Agree | Agree | Disagree | Strongly Disagree | Not Applicable |
|---|---|---|---|---|
| ❑ | ❑ | ❑ | ❑ | ❑ |

Comments _____

■ The illustrations were clear.

| Strongly Agree | Agree | Disagree | Strongly Disagree | Not Applicable |
|---|---|---|---|---|
| ❑ | ❑ | ❑ | ❑ | ❑ |

Comments _____

■ The information was complete and easy to find.

| Strongly Agree | Agree | Disagree | Strongly Disagree | Not Applicable |
|---|---|---|---|---|
| ❑ | ❑ | ❑ | ❑ | ❑ |

Comments _____

■ Do you have additional comments about the *SPARCstorage Array Configuration Guide*?

_____

_____

_____

_____

Name: _____

Title: _____

Company: _____

Address: _____

_____

Telephone: _____

Email address: _____

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

# BUSINESS REPLY MAIL

FIRST CLASS MAIL PERMIT NO. 1 MOUNTAIN VIEW, CA

POSTAGE WILL BE PAID BY ADDRESSEE

SUN MICROSYSTEMS, INC.
Attn: Manager, Hardware Publications
MS MPK 14-101
2550 Garcia Avenue
Mt. View, CA 94043-9850