Fibre Channel Storage Manager

Installation and User's Guide for AIX Operating System

SC26-7290-00

Fibre Channel Storage Manager

Installation and User's Guide for AIX Operating System

SC26-7290-00

Take Note!

Before using this information and the product it supports, be sure to read the general information under “Notices” on page ix.

First Edition (April 1999)

Publications are not stocked at the address given below. If you want additional IBM* publications, ask your IBM representative or write to the IBM branch office serving your locality.

A form for reader’s comments is provided at the back of this publication. If the form has been removed, address your comments to:

International Business Machines Corporation
RCF Processing Department
Dept G26/Bldg. 050-3
5600 Cottle Road
SAN JOSE, CA 95193-0001
U.S.A.

FAX: 1-800-426-6209

You can also send your comments electronically to:

starpubs@vnet.ibm.com

Contents

Figures

Notices

References in this publication to IBM products, programs, or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Subject to IBM's valid intellectual property or other legally protected rights, any functionally equivalent product, program, or service, may be used instead of the IBM product, program, or service. The evaluation and verification of operation in conjunction with other products, except those expressly designated by IBM, are the responsibility of the user.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

Read Before Using

IMPORTANT

YOU ACCEPT THE TERMS OF THIS IBM LICENSE AGREEMENT FOR MACHINE CODE BY YOUR USE OF THE HARDWARE PRODUCT OR MACHINE CODE. PLEASE READ THE AGREEMENT CONTAINED IN THIS BOOK BEFORE USING THE HARDWARE PRODUCT. SEE "IBM License Agreement for Machine Code" on page xi.

Trademarks

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

- AIX
- AIXwindows
- IBM
- RS/6000

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

IBM License Agreement for Machine Code

Regardless of how you acquire (electronically, preloaded, on media or otherwise) BIOS, Utilities, Diagnostics, Device Drivers or Microcode (collectively called “Machine Code”), you accept the terms of this Agreement by your initial use of a Machine or Machine Code. The term “Machine” means an IBM machine, its features, conversions, upgrades, elements or accessories, or any combination of them. Acceptance of these license terms authorizes you to use Machine Code with the specific product for which it is provided.

International Business Machines Corporation or one of its subsidiaries (“IBM”), or an IBM supplier, owns copyrights in Machine Code.

IBM grants you a nonexclusive license to use Machine Code only in conjunction with a Machine. As the rightful possessor of a Machine, you may make a reasonable number of copies of Machine Code as necessary for backup, configuration, and restoration of the Machine. You must reproduce the copyright notice and any other legend of ownership on each copy of Machine Code you make.

You may transfer possession of Machine Code and its media to another party only with the transfer of the Machine on which the Machine Code is used. If you do so, you must give the other party a copy of these terms and provide all user documentation to that party. When you do so, you must destroy all your copies of Machine Code.

Your license for Machine Code terminates when you no longer rightfully possess the Machine.

No other rights under this license are granted.

You may not, for example, do any of the following:

1. otherwise copy, display, transfer, adapt, modify, or distribute in any form, Machine Code, except as IBM may authorize in a Machine’s user documentation.
2. reverse assemble, reverse compile, or otherwise translate the Machine Code, unless expressly permitted by applicable law without the possibility of contractual waiver;
3. sublicense or assign the license for the Machine Code; or
4. lease the Machine Code or any copy of it.

The terms of IBM’s Machine warranty, which is incorporated into this Agreement by reference, apply to Machine Code. Please refer to that warranty for any questions or claims regarding performance or liability for Machine Code.

About This Book

This book provides information on the IBM 2102 Fibre Channel RAID Storage Server and the IBM Fibre Channel Storage Manager for AIX. Topics covered include definitions of the manager and array programs, their installation and configuration, and how to interpret their error messages. In addition, this book contains information about the programming interface to the **fcparray** device driver.

Before you begin the tasks described in this book, ensure that your system meets the following hardware and software requirements.

You must have a 2102 Storage Server and the appropriate host adapters to run the Fibre Channel Storage Manager software. Also, ensure that you have the latest level of hardware documentation.

SMIT Examples: The examples that rely on the System Management Interface Tool (SMIT) assume that you are using AIX from an ASCII display. SMIT is also available within the AIXwindows environment.

Who Should Use This Book

This book is written for the system administrator and system programmers of an AIX system. These are persons responsible for installing and maintaining the system. To use this book, you must already be familiar with AIX commands, utilities, and procedures.

Related Publications

Additional information that is related to the subsystem is available in the following publications:

- *AIX: Installation Guide and Reference*
- *AIX Version 4 Getting Started*, SC23-2527
- *AIX Version 4 Messages Guide and Reference*, SC23-2641
- *AIX and Related Products Documentation Overview*, SC23-2456
- *Adapters, Devices, and Cable Information for Multiple Bus Systems*, SA38-0516
- *AIX Version 4 Topic Index and Glossary*, SC23-2513
- *AIX Version 4 Kernel Extensions and Device Support Programming Concepts*, SC23-2611
- *AIX 4.3 System Management Guide: Operating System and Devices,* SC23-4126
- *Diagnostic Information for Multiple Bus Systems,* SA38-0509
- *Small Computer System Interface (SCSI) Specification* from American National Standards Institute (ANSI)
- *PCI Local Bus Specification* from PCI Special Interest Group.
- *Translated Safety Notices for Open Attachment,* GC26-7246
- *Fibre Channel RAID Storage Server, Introduction and Planning Guide, IBM 2102,* GC26-7281
- *Fibre Channel Storage Manager, Installation Guide for Microsoft Windows NT,* SC26-7283
- *StorWatch Fibre Channel RAID Specialist, Installation Guide for Microsoft Windows NT and Windows 95,* SC26-7284

Related Publications

- *Fibre Channel Storage Manager and StorWatch Fibre Channel RAID Specialist, User's Guide,* SC26-7285
- *Fibre Channel Storage Manager, Installation Guide for Sun Solaris Operating System*, SC26-7286
- *Fibre Channel Storage Manager, Installation Guide for Hewlett Packard HP-UX Operating System*, SC26-7287
- *Fibre Channel RAID Storage Server and Expandable Storage Unit, User's Guide, IBM 2102 Model F10 and Model D00,* SC26-7288
- *Fibre Channel Storage Hub, Installation, Service, and User's Guide, IBM 2103 Model H07,* SC26-7289
- *Fibre Channel RAID Storage Server and Expandable Storage Unit, Service Guide, IBM 2102 Model F10 and Model D00*, SY27-7604
- *Seascape Solutions Rack, Installation and Service Guide, IBM 2101 Model 100*, SY27-7606

Ordering Publications

You can order publications from your sales representative or from your point of sale.

To order additional copies of this book, use order number SC26-7290.

Use *AIX and Related Products Documentation Overview*, for information on related publications and how to obtain them. The *AIX and Related Products Documentation Overview* is available:

- On the World Wide Web. Using your Web browser, go to the IBM Home Page by opening the following URL:

  `http://www.ibm.com/`

  On the IBM Home Page, search on ″AIX and Related Products Documentation Overview″ to access the publication.
- In the *AIX Hypertext Information Base Library*. This online documentation is designed for use with the InfoExplorer hypertext retrieval system.
- As a hardcopy publication. Use order number SC23-2456.

Additional publications are available for purchase from IBM. For a list of publications available in your country:

- In the U.S. and Puerto Rico, call **1–800–426–7282**.
- In the United Kingdom, call **01705–565000** or **0161–9056001**.
- In Canada, call **1–800–465–1234**.
- In other countries, contact the IBM support organization that services your area, your IBM marketing representative, or your IBM reseller.

Web Site

For more fibre channel storage information, go to the following Web site:

http://www.ibm.com/storage/fcss

Chapter 1. Introducing RAID Subsystems

Data storage and backup were once achieved through disk partitioning and tape backup. But disk partitioning degraded performance, because only one device was available for file access. Online data requiring a secure environment, also referred to as redundant information, absorbed critical amounts of memory, further degrading response time.

Redundant Array of Independent Disks (RAID) technology offers the user these primary features:

- Larger disk capacity
- Immediate availability and recovery of data
- Redundancy of data at a user-selected level

Using RAID technology, data is stored across a series of compact disk drives known as a disk array. Depending on the RAID level selected, this storage technique provides the data redundancy required for a secure system without depleting memory, and may have the further advantage of faster retrieval through multiple channel access. Also, in the event of hardware failure, a single disk drive can usually be replaced without interrupting normal system operation.

This chapter contains the following topics:

- Disk arrays
- Write cache
- RAID levels
- 2102 Fibre Channel RAID Storage Server features
- Configuring the 2102 Fibre Channel RAID Storage Server overview
- Overview of the Fibre Channel Storage Manager for AIX

Disk Arrays

Disk arrays are groups of disk drives that work together with a specialized array controller to achieve higher data-transfer and input and output (I/O) rates than those provided by single large drives. The array controller keeps track of how the data is distributed across the drives.

Arrays can also provide data redundancy, so that no data is lost if a single drive in the array fails. Two methods of writing data to the disk drives are used in a RAID subsystem:

striping Data for a given file may be written in segments to different drives in the array, rather than being written to a single drive. By using multiple drives, the array can provide higher data-transfer rates and higher I/O rates when compared to a single large drive.

mirroring Data that is simultaneously written to two separate disks within the same array.

The method used for writing the data to an array is determined by the RAID level defined for that array. “RAID Levels” on page 5 discusses in detail the various RAID levels and their function.

Arrays are contained within an array subsystem. Depending on how you configure it, an array subsystem can contain one or more arrays, also referred to as *logical units* (LUNs). Each LUN has its own characteristics (for example: RAID level, logical block size, and logical unit size). The AIX operating system views each array as a disk with its own logical name.

Logical Units

Figure 1 illustrates the terms used in describing disk arrays.

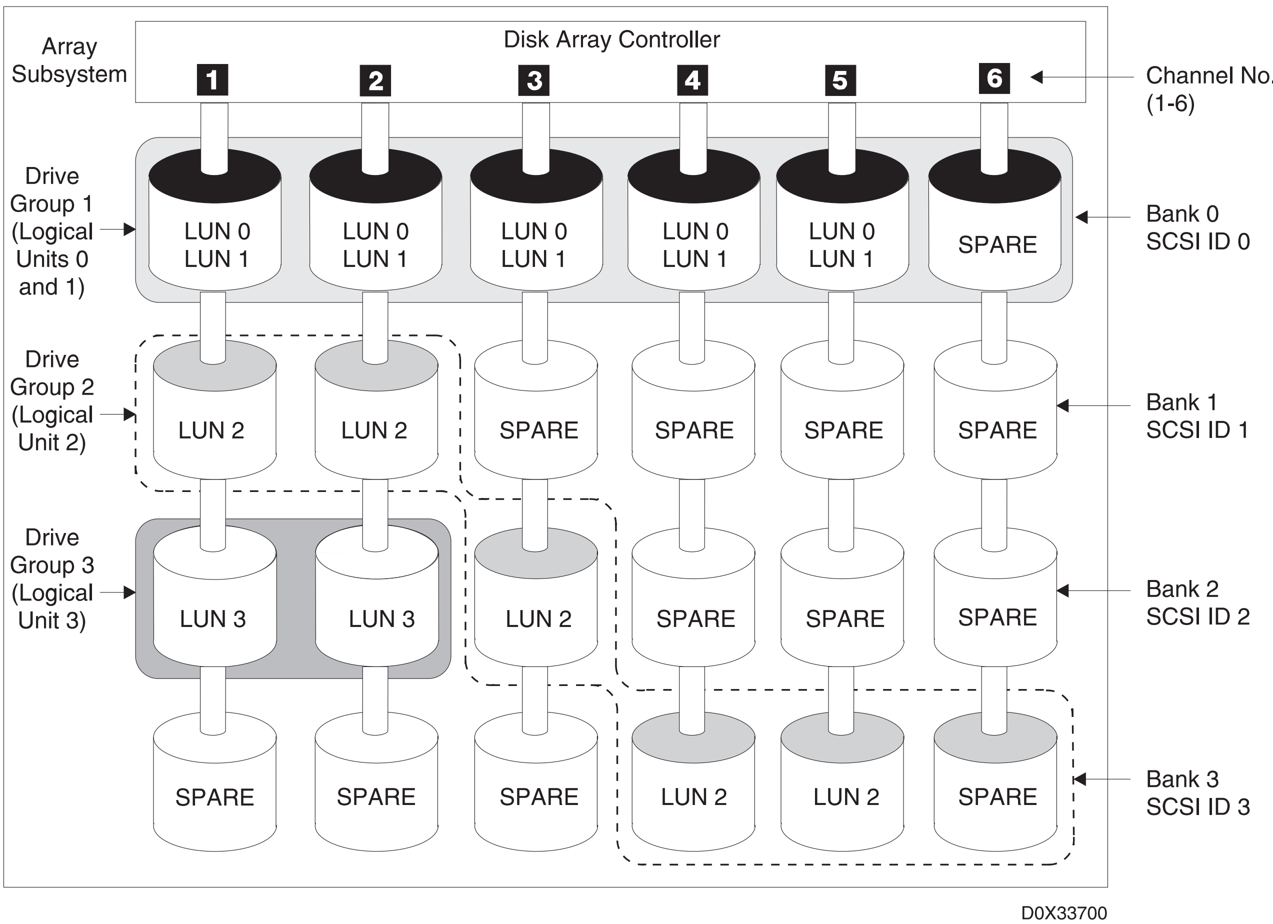

Figure 1. Logical Units, Groups, and Banks

Each array is a single LUN and a unique AIX device with its own hard disk name and location code. Each LUN has its own array parameters (for example, RAID level and segment size). For most purposes, a LUN is equivalent to an array.

A LUN is treated as a single disk drive by the operating system. There are no special requirements for using it. You can use AIX commands and utilities just as you would with a single drive.

For example, after you configure a LUN with the Fibre Channel Storage Manager, you can use AIX commands to make the LUN available to the system by adding it to a volume group, creating a file system on it, and so on. When adding a configured LUN to your system, treat it in the same way you would treat an external hard disk.

Figure 1 on page 2 shows four LUNs. LUNs 0 and 1 are in drive group 1, LUN 2 is in drive group 2, and LUN 3 is in drive group 3.

Drive Groups

A *drive group* is a set of 1 to 20 drives that have been configured into one or more LUNs. Most of the time, you should create only one LUN in a drive group, so in this case, the terms *logical unit* and *drive group* are synonymous. A LUN cannot span drive groups and the LUNs in a drive group must have the same RAID level.

Drive grouping is used only by the Fibre Channel Storage Manager. The operating system treats the individual LUN, not the drive group, as a single disk drive, sending I/O to it and retrieving I/O from it.

Attention: In a drive group, two or more LUNs may reside on the same physical drives. This has implications for recovery after the failure of a single drive. If that drive is part of a drive group with more than one LUN (hdisk) on it, you will have to perform recovery procedure on all the hdisks affected by the drive failure. For example, in Figure 1 on page 2, LUNs 0 and 1 are in the same drive group (Group 1). These two LUNs are completely separate hdisks, even though they are in the same drive group. If the drive shown as Channel 3, SCSI ID 0 fails, the data on the logical volumes and file systems on both LUNs may be affected.

Figure 1 on page 2 shows three drive groups. Drive group 1 is made up of the five drives of bank 0, and contains LUNs 0 and 1. Drive group 2 is made up of drives from banks 1, 2, and 3, and contains LUN 2. Drive group 3 is made up of drives from bank 2 and contains LUN 3.

Drive Banks

A *bank* is a physical grouping of drives that cannot be changed (unlike a drive group, which is an arbitrary grouping of drives that can be changed by reconfiguring). One bank of drives contains up to six drives in the same horizontal plane. Drives in an array are identified by channel number and SCSI ID. Channel numbers go from 1 to 6 within a bank. All drives in a bank have the same SCSI ID.

Write Cache

Write cache improves array performance by keeping recently written information in memory located in the array controllers. Data is written to the array drives when there is free time or when the cache memory has filled up.

Here are the steps of the write cache function:

1. The host issues a write operation to the controller.
2. The controller stores the data from this operation in its cache.
3. The controller notifies the host that the write operation is complete.
4. The controller flushes the data out to the drives.

Write cache offers two options: fast write and fast load.

Attention: The 2102 Storage Server, with the fast-write cache feature enabled, provides a high degree of protection of cached data. However, when both of the following conditions occur, cached data may be lost:

- There is data in the cache when power is lost to the array
- The power is not restored until after the cache batteries have drained below the level needed to maintain the data.

If a normal shutdown is done to the system attached to the 2102 Storage Server, and the power to the Storage Server is turned off a full 90 seconds after the systems have been shut down, the data in the cache of the Storage Server will be flushed to disk and the integrity of the file system is maintained.

This danger also exists if a controller failure occurs and the fast-load cache method is used. The fast-load option should not be used except when restoring data from backup or archive. This method of write cache does not offer protection in the event of a controller failure.

Use of the fast-write option offers additional protection of cached data by mirroring the data in the cache of the second controller. If a controller failure occurs during a write operation, the cached data will be retrieved from the alternate controller and written to the drives in the array. If the 2102 Storage Server finds that the charge left in the cache batteries of either of the two controllers is below the level where data integrity cannot be guaranteed should a power failure occur, it turns off write caching to the array until the battery power is restored. When this occurs, a FCP_ARRAY_ERR16 is logged. These error messages will continue to be logged until the battery charge is restored.

If there was unwritten data in the cache and the power in the cache batteries drained before power was restored, an FCP_ARRAY_ERR15 is logged against the hdisk. A message about this error is also displayed on the system console. This is a potentially irrecoverable error. If this error is received, the data on the hdisk should be checked, at a minimum, or restored from backup.

The hardware battery status can be queried by using the raidmgr command. To query the cache status of the DAC or the hdisk, use the following command:

```
fraidmgr -P cache -l hdisk# (or dac#)
```

Fast-Write Option

The fast-write option offers advantages over the fast-load option. With the fast-write option, writes are cached only when there is sufficient charge in the batteries on both controllers to hold the data in the event of a power or controller failure. In addition, the cache data is mirrored in the cache of the second controller, providing added protection in the event of a power or controller failure.

Fast-Load Option

The fast-load option is typically used instead of the fast-write option when loading the data from archive storage.

Attention: If the controller fails before flushing the data out to the drives, all data from the write operation is lost.

When you use the fast-load function, you need to identify a method of retrieving lost data should the controller fail. The following scenarios provide examples of using the fast-load function and a method for retrieving lost data in the event of a controller failure.

- When restoring data to your 2102 Storage Servers from a tape drive or mainframe, you may use the fast-load function to install data more quickly than with a traditional method. Since you have a backup of your data, you need not be as concerned about a controller failure. If the controller does fail and your data is lost, you can obtain another copy of the data from your tape drive or mainframe.
- If you have a single controller, you may use the fast-load function to enhance performance. Since drives are the single largest failing component in a storage system, the risk of controller failure is low. However, if the controller does fail and your data is lost, you will need a way to regenerate the data (such as re-entering it).

RAID Levels

The *RAID level* determines the degree of error recovery available when a part of the system fails. Depending on the RAID level chosen (with the exception of RAID level 0), if a hard drive fails within an array, the data from that disk can be reconstructed from the data stored on other hard drives within the array. It is also sometimes possible that this data can be reconstructed or ″put back into service″ with little or no impact to the users and programs that are using the system at the time of the failure.

The valid RAID levels are: 0, 1, 2, 3, 4, 5. The levels of RAID supported varies between different hardware and software platforms, but the most common levels supported are 0, 1, 3, and 5.

Note: The 2102 Storage Server and Fibre Channel Storage Manager support levels 0, 1, and 5.

Each of the levels supported by the Fibre Channel Storage Manager uses a different method of writing data to provide specific benefits with each method. For example, RAID levels create *array parity* information in order to reconstruct data lost because of a drive failure in the array. RAID level 0 does not provide array parity and has no data redundancy, but it provides a high I/O rate. The following information provides details about RAID levels 0, 1, and 5. Recommendations about which level is best for performance or reliability is provided in “Selecting RAID Levels” on page 29.

RAID Level 0

Attention: It is critical that you regularly back up data on a RAID level 0 array. This is the only way to recover data in the event of a disk drive failure. The ability to recover data on the array depends on the backup procedures you have established.

RAID level 0 writes data across the drives in the array, one segment at a time. In this example, a segment is defined as 2 blocks of 512 bytes each, as shown in Figure 2 on page 6. Blocks 0 and 1 are written to drive 1, blocks 2 and 3 are written do drive 2, and so on until each drive contains a single segment. Then blocks 10 and 11 are written to drive 1, blocks 12 and 13 to drive 2, and so on.

Note: Segment 0 is zero blocks in this figure and therefore not shown.

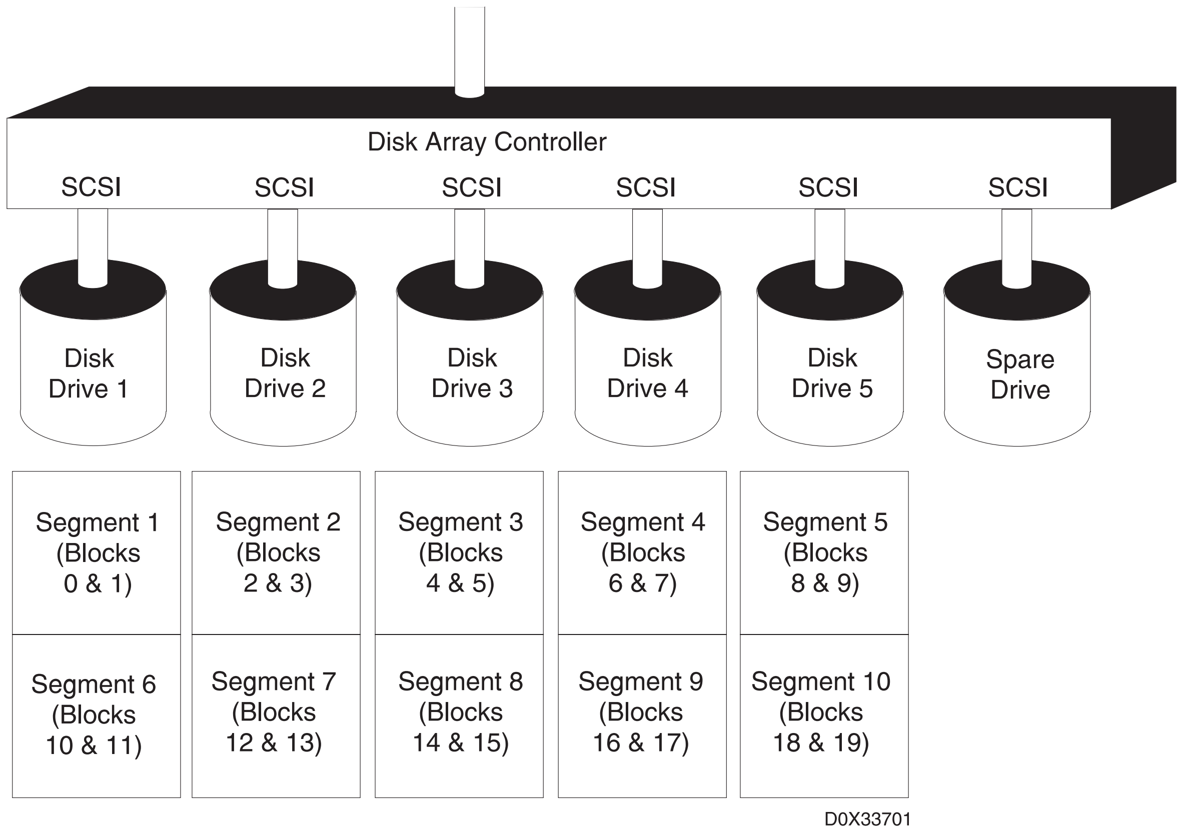

Figure 2. RAID Level 0

The host system treats a RAID level 0 array like a standard hard drive. RAID level 0 errors are reported in the same way as normal drive errors, and the recovery procedures are the same as those used on a standard hard drive. For example, if a drive fails, the array controller returns the same errors that occur during a read or write retry failure. All data on the array may be lost. However, unlike what it does for other RAID levels, the array controller never marks a RAID level 0 array as degraded or dead as the result of a drive failure.

RAID level 0 offers a high I/O rate, but is a non-redundant configuration. That is, there is no array parity information generated for the purpose of reconstructing data in the event of a drive failure. Therefore, there is no error recovery beyond what is normally provided on a single drive. All data in the array must be backed up regularly (using the file-by-file backup method) to protect against data loss.

RAID Level 1

Attention: Even though a RAID level 1 array has data redundancy, you should regularly back up data on the array. This is the only way to recover data in an event such as accidental file deletion or disaster recovery. You can continue to operate the array in degraded mode until you replace the drive. However, you should replace the drive as soon as possible. If you cannot replace the drive immediately, you should back up your data file-by-file to prevent potential data loss.

RAID level 1 transparently mirrors data by striping segments across data drives and mirrored data drives. Any time data is written to a drive, it is also written to a mirrored drive without the system directly controlling the location of the mirrored data.

Traditionally, RAID level 1 has been used for critical fault-tolerant transaction processing. Mirrored data provides high reliability. When a small block size is used, mirrored data also provides a high I/O rate. However, RAID level 1 is a more costly RAID solution because it requires a mirrored data drive for every data drive in the array.

In this example, a *segment* is defined as two blocks of 512 bytes each, as shown in Figure 3 on page 8. Blocks 0 and 1 are written to data drive 1, while the mirrored data blocks 0 and 1 are written to the mirrored data drive 1. Then data blocks 2 and 3 are written to data drive 2, while the mirrored data blocks 2 and 3 are written to the mirrored data drive 2, and so on.

Note: Segment 0 is zero blocks in this figure and is therefore not shown.

If a drive fails in a RAID level 1 array, you can continue to use the array normally. A RAID level 1 array operating with a single failed drive is said to be operating in degraded mode. Whenever you read or write to a LUN in degraded mode, the array controller retrieves the failed drive's data from its mirrored drive. Although you can continue to operate the RAID level 1 array with a failed drive, you should replace the drive and restore the array as soon as possible.

Although RAID level 1 does not have parity, using Parity Scan allows the array controller to compare the data drive and mirrored data drive. If the array subsystem experiences an abnormal system shutdown, you should run array parity check/repair on the affected LUNs. See "Checking and Repairing Array Parity" on page 58 for more information.

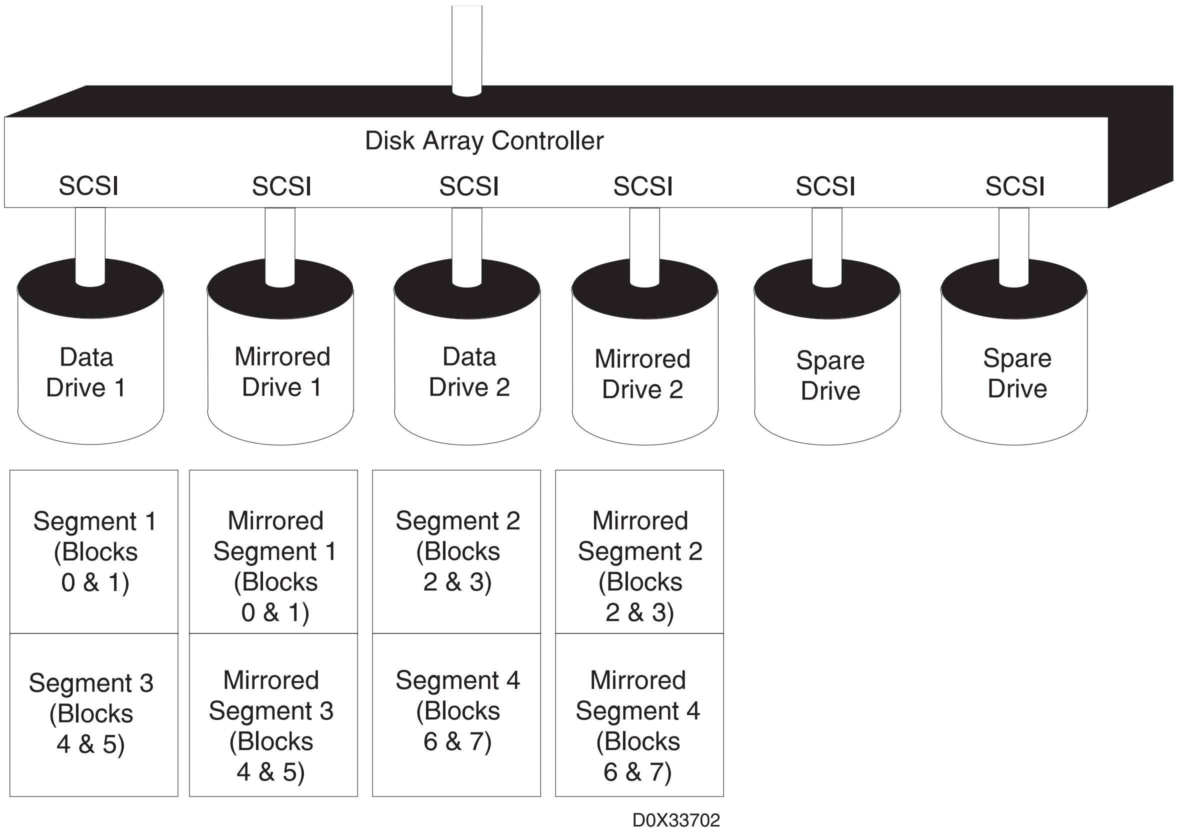

Figure 3. RAID Level 1

RAID Level 5

Attention: Even though a RAID level 5 array maintains parity information, it is still important that you regularly back up data on the array. This is the only way to recover data, for example, in the event of accidental file deletion or disaster recovery. You can continue to operate the array in a degraded mode until you can replace the drive. You should replace the drive as soon as possible. If you cannot replace the drive almost immediately, you should back up your data file-by-file to prevent potential data loss.

RAID level 5 stripes data across all drives in the array one segment at a time (a segment can contain multiple blocks). RAID level 5 also writes array parity data, but the parity data is spread across all the drives. The result is a transfer rate equal to that of a single drive, but with a high overall I/O rate.

In Figure 4 on page 9, as an example, a segment is defined as two blocks of 512 bytes each. Blocks 0 and 1 are written in the first position to drive 2, blocks 2 and 3 are written to drive 3, blocks 4 and 5 are written to drive 4, blocks 6 and 7 are written to drive 5, and the parity data for the blocks in data segments 1, 2, 3, and 4 is written to drive 1, and so on.

If a drive fails in a RAID level 5 array, you can continue to use the array normally. A RAID level 5 array operating with a single failed drive is said to be operating in degraded mode. Whenever you access a LUN in degraded mode, the array controller recalculates the data on the failed drive by using data and parity blocks on the operational drives.

For example, to recalculate data in data segment 4 in Figure 4 (the first position on drive 5), the array controller would use the parity information from drive 1 and the data from drives 2, 3, and 4 (data segments 1, 2, and 3) to reconstruct the data. This process is repeated to reconstruct each block of the failed drive, as needed, so you can continue to operate the RAID level 5 array.

Parity Scan allows the array controller to check the integrity of the data and array parity. If the array subsystem experiences an abnormal system shutdown, you should run array parity check/repair on the affected LUNs. See “Checking and Repairing Array Parity” on page 58 for more information.

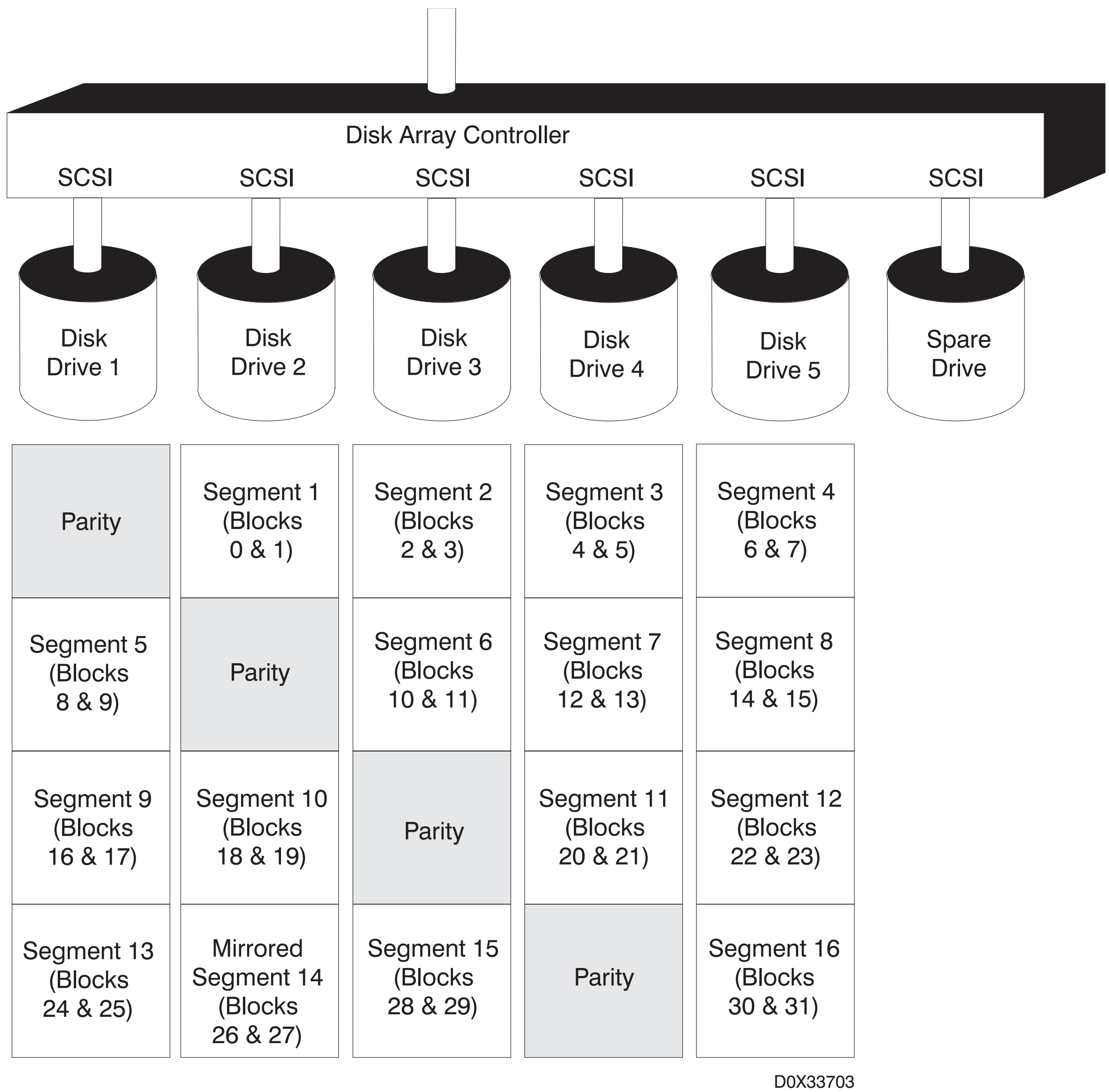

Figure 4. RAID Level 5

2102 Fibre Channel RAID Storage Server Features

The three major components of the Fibre Channel Storage Manager are:

- 2102 Storage Server controllers
- RAIDiant Disk Array Managers (SMIT interfaces)
- **fcparray** device driver

The controllers, interfaces, and device driver described in this documentation combine to provide an advanced storage technique designed to secure online information while speeding file access through multiple, independent transactions.

The 2102 Storage Server controllers are supported by the FC SCSI RAIDiant Disk Array Manager (FCRDAM) and the **fcparray** device driver. The **fcparray** device driver allows for the definition of multiple paths between a host and a disk array and manages the switch-over between these paths in the event of a component failure. This ensures a secure data environment in the event of a component failure.

The 2102 Storage Server consists of two disk array controllers with a fibre channel host interface and multiple SCSI busses that have attached disk drives. The array can be configured to maximize either performance or reliability.

Configuration is performed using the Fibre Channel Storage Manager. Depending on how the array subsystem is configured, it may contain one or more arrays in which each array has a different RAID level. Advantages and disadvantages of each RAID level are discussed in “Selecting RAID Levels” on page 29.

On a system with two controllers, the 2102 Storage Server is normally configured in a dual-active mode in which both controllers are in an active state. In the *dual-active* mode, each controller may communicate with a different subset of arrays.

A controller may also be in either a passive state or in a failed state. When the controller is in a *passive* state, it can respond to queries from the host, but it does not control any of the arrays. The *failed* state is also referred to as a held-in-reset state. Controllers enter the passive or failed states due to failures in the controller hardware or software or the communications path between the host and controller. When a problem is detected, the controller affected (whether reporting trouble itself or suffering communication loss) is placed into the passive or failed state, and the remaining controller takes over control of all disk arrays.

A passive state indicates that the failure is more likely to be recoverable without service intervention (for instance, in the case of a communication failure). A failed state indicates that the failure is likely to be long-term and require hardware service (for instance, in the case of a failed controller component).

The Fibre Channel Storage Manager automatically attempts to restore passive controllers to active status at regular intervals. The Fibre Channel Storage Manager first verifies that the problem that caused the controllers to be placed in the passive state has been resolved. See “Verifying the 2102 Fibre Channel RAID Storage Server Software Installation” on page 35 for more information.

The 2102 Storage Server supports global hot-spare drives (hereafter referred to as hot-spare drives). A *hot-spare drive* is a drive, which is not part of a drive group, that is set to the hot-spare status. A hot-spare drive will take over for any other drive that fails within the 2102 Storage Server, assuming the hot-spare drive is at least as large as the failed drive and the failed drive is part of a redundant array configuration (RAID 1 or 5). The hot-spare drive is temporarily marked as part of the LUN while the original drive is failed. The 2102 Storage Server will immediately begin rebuilding array data onto the hot-spare drive once it is activated. When the original drive is replaced and rebuilt, the hot-spare drive will return to hot-spare status.

Also, note that hot-spare drives will interact with robustness. A hot-spare drive located on the same channel as a non-failed data drive within the LUN for which it is sparing will create a lack of robustness within an otherwise robust LUN. Thus, to maintain a robust configuration, always place the hot-spare drives on a channel distinct from all data drives. This will leave LUNs robust through a single drive failure. However, a second drive failure on the same LUN will still create a robustness issue.

Note: The 2102 Storage Server is not supported as a boot device.

Configuring the 2102 Fibre Channel RAID Storage Server Overview

As with any device connected to an AIX system, the 2102 Storage Server must be defined and configured before it can be accessed.

- A device is defined when the system creates a record for it in the customized database portion of the Object Data Manager (ODM). The ODM database contains information about all the devices that can be connected to the system. In the ODM database, device information is stored in an object relationship. The ODM database stores generic descriptions of the various supported devices in its predefined database. For each instance of a particular device type, the system creates a separate copy of the predefined record.
- A device is configured when all necessary device drivers are loaded into the kernel and all other components necessary for communication with the device are configured. When configuration is completed, the device is marked as available.

If the 2102 Storage Server is configured to provide high availability, its configuration also includes the definition of multiple paths between the host and the disk array that provide redundant access to the data on the disk array.

At boot time, or whenever you invoke the **cfgmgr** command, the AIX operating system configures the logical units on a disk array automatically, including defining the multiple paths. You can also configure the logical units on a disk array using the SMIT interface.

To be configured, a 2102 Storage Server must have logical units (LUNs) defined on it. It comes with several LUNs preconfigured. However, you can change the characteristics of the LUNs by using the Fibre Channel Storage Manager. If you change the composition of a LUN, you must reconfigure the 2102 Storage Server.

Automatic Configuration at Boot Time

At boot time, configuration of the 2102 Storage Server is triggered when the system detects the fibre channel adapter (or one of the fibre channel adapters) to which the disk array is connected. The system identifies the adapter and then executes the configuration routines (configuration methods) associated with the device.

Figure 5 provides a brief overview of the configuration methods for defining an array subsystem.

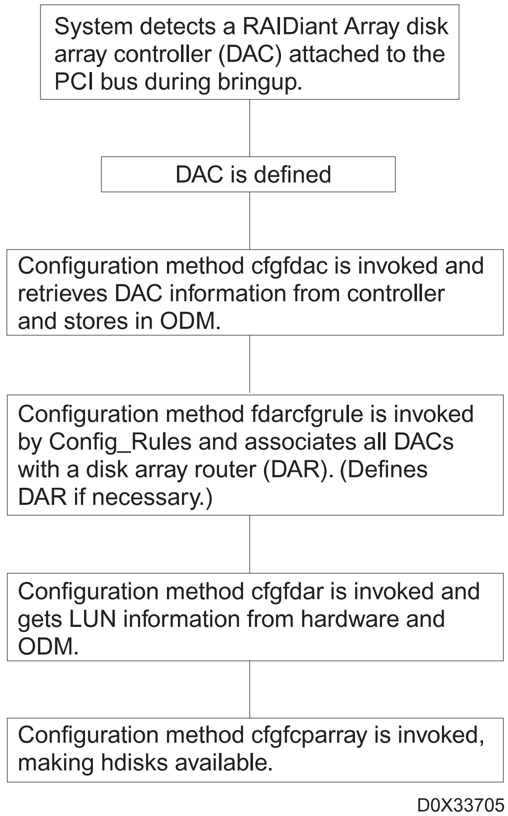

Figure 5. Configuration Methods

Manual Unconfiguration and Reconfiguration of 2102 Devices

After the boot sequence completes, you can change the configuration of 2102 Storage Server devices such as the DACs, the dars, and the hdisks.

Unconfiguring Devices

To unconfigure all 2102 devices, you must first unconfigure all the hdisks, then all the dars, then finally all the DACs.

Note: The -l option puts the device in a defined state, but does not remove the device definition from the Object Data Manager (ODM). The -d option deletes or removes the device definition from the ODM.

1. To unconfigure a LUN, enter:
   ```
   rmdev -l hdisk# -d
   ```
2. To recursively unconfigure the LUNs and the dar, enter:
   ```
   rmdev -l dar# -d -R
   ```
3. To unconfigure the DAC, enter:
   ```
   rmdev -l dac# -d
   ```

Reconfiguring Devices

The 2102 devices in the defined state can be reconfigured, or brought back to the available state, by running **cfgmgr** as described in “Configuring the 2102 Fibre Channel RAID Storage Server Overview” on page 11. However, all the devices will be reconfigured, not just the 2102 devices. To reconfigure *only* the 2102 devices, enter the following sequence of commands:

1. Enter either of the following commands for each 2102 DAC:
   ```
   cfgmgr -l dac#
   ```
   ```
   mkdev -l dac#
   ```
2. After all DACs are configured in step 1, enter:
   ```
   /usr/lib/methods/fdarcfgrule
   ```
3. Enter the following command for each 2102 dar:
   ```
   cfgmgr -l dar#
   ```

Configuration Summary

Figure 6 on page 14 illustrates the configuration of a 2102 Storage Server connected to a host system through two adapters, two controllers, and separate fibre channel networks. This configuration results in making LUN 1 available as hdisk5. The figure includes all the logical names assigned by the system during configuration. The solid line indicates the primary path to the LUN. The dotted line indicates the secondary path.

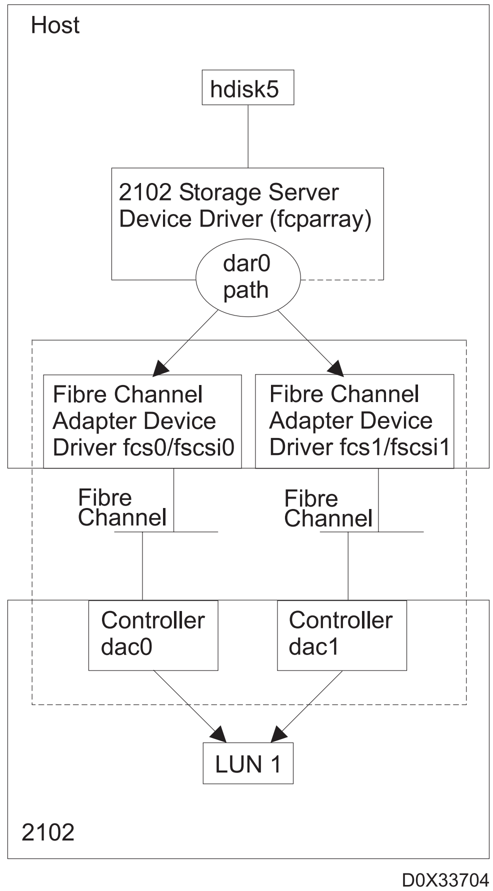

Figure 6. Configuration Results

2102 Storage Server ODM Information

The operating system stores information about each device that can be connected to it in a database maintained by the Object Data Manager (ODM). The information stored about each device includes the following:

- Device class, subclass, and type
- Device parent
- Device location code
- Connection point
- Attributes

The following sections describe the values stored for each of these categories for the 2102 Storage Server.

Device Class, Subclass, and Type

A device's class, subclass, and type unambiguously identify it to the operating system. With this information, the operating system can locate the device's record in the Object Data Manager (ODM) and obtain other information stored in the database about the device.

The device class describes the general category of devices to which the device belongs. For example, all printers are in the printer class. The 2102 Storage Server belongs to the disk class.

The device subclass describes how the device is connected to the host, typically the type of adapter through which the system communicates with the device. For example, for standard SCSI disks, the value would be SCSI. Because the system does not access a 2102 Storage Server directly through an adapter, the subclass of a 2102 Storage Server is the disk array router pseudo-device, identified by the name dar.

The device type describes the characteristics of a particular device within a class. For example, for a standard SCSI disk, the type can be a description of its storage capacity, such as 2 GB. Because the 2102 Storage Server can be configured to provide varying amounts of storage, array is used as its type designation.

Class	Type	Subclass	
array	ibm-dac	fcp	controller (dac)
disk	array	fdar	device (hdisk)
driver	fdar	node	router(dar)

Device Parent

A device’s parent is the device that must be configured or defined before the device itself may be configured or defined. A device can only have a single parent. For example, the parent device of a standard SCSI disk is the SCSI adapter, because the system cannot communicate with the disk drive without going through the adapter. The parent device is specified by its system-assigned name, such as scsi0, scsi1, fscsi0, or fscsi1. Because the adapter through which the system accesses the 2102 Storage Server may vary, the parent of a 2102 Storage Server is the disk array router (DAR) pseudo-device, such as dar0.

Device Attributes

The 2102 Storage Server supports the same attributes as other SCSI disks. Additional attributes supported by the 2102 Storage Server are:

- RAID level
- Size
- SCSI address
- Cache method
- Location
- Worldwide name
- IEEE volume name

Identifying Hard Disks and Logical Units

There are times when you need to determine the hard disk name of a logical unit, which volume groups your logical units are in, or which logical volumes are on a given logical unit. For example, to back up data after a drive in an array fails, you will need to know the hard disks and logical volumes or file systems affected by that failure.

Table 1 on page 16 explains how to make some of the array logical unit/AIX connections.

Table 1. Disk and Unit Identification

Task	Tools and Procedures
Find the hard disk names of logical units, or vice versa, or determine which disks are on a physical drive.	The FC SCSI RAIDiant Disk Array Manager (FCRDAM): The **List All SCSI RAID Arrays** option in the 2102 Storage Server identifies a single logical unit and gives its hard disk name. See "Logical Unit Device Names and Location Codes".
Determine the logical volumes and file systems on a logical unit.	SMIT or the lspv command: In general, it is easier to use the lspv command. From the console, enter lspv -l hdisk, where hdisk is the disk name of the logical unit you want to check. The resulting display will show you all the logical volumes and files systems on the logical unit.
Determine which logical units are in which volume groups.	Either SMIT or the lspv command: In general, it is easier to use the lspv command. From the console, enter lspv. The resulting display shows you the disk contents of all the volume groups on your system.

Logical Unit Device Names and Location Codes

Logical units are assigned disk names using the hdisk form, the same as any other disk storage unit in AIX. These names are automatically assigned whenever you create a logical unit with the FCRDAM. The names are deleted when you delete the logical unit. You can display these names and the location codes associated with them by using the **List All SCSI RAID Arrays** option in the FC SCSI RAIDiant Disk Array Manager menu.

Some Fibre Channel Storage Manager commands operate at the controller level, rather than the logical unit level. For example, when you run parity check/repair, you perform it on an individual logical unit. A controller is always accessed using the dac device name.

Understanding the Use of Logical Units in a SCSI RAID Array

Logical units are managed using the Fibre Channel Storage Manager. Logical units can be added, deleted, or combined into logical volumes through the use of naming conventions and location codes. Refer to the following information:

- "Configuring the 2102 Fibre Channel RAID Storage Server Overview" on page 11, which describes the procedures used in creating and configuring a logical unit
- "Adding Logical Units to the Operating System" on page 46
- "Deleting Logical Units from the Operating System" on page 47
- "Extending Logical Volumes and Logical Units" on page 22, which explains the restrictions that apply to extending or combining logical volumes

Creating and Configuring Logical Units

Logical units are configured by using the Fibre Channel Storage Manager. These methods allow you to change the factory configuration of your logical units as well as reconfigure any logical units that you have already modified.

Refer to the following information for these procedures:

- "Fibre Channel RAID Array Configuration" on page 38
- "Changing Reconstruction Rate Parameters" on page 57

Logical Unit Parameters

Each logical unit has a set of parameters that determine how data is stored on it. Each logical unit can have different parameters, with the following exception: all logical units in the same drive group are automatically assigned the same RAID level and drive map by the Fibre Channel Storage Manager. That is, they contain the same drives, with no overlap.

You can define up to 32 logical units per array (logical unit 0 to logical unit 31) with AIX Version 4.3.2 (3/99 update) (or higher) on the 2102 Storage Server with Version 3.0 (or higher) of the controller microcode.

Table 2 summarizes the meanings of the logical unit parameters. More detailed descriptions appear following the figure.

Table 2. Logical Unit Parameters. (See the Create a SCSI RAID Array and Change/Show SCSI RAID Array menus in SMIT.)

Parameter	Meaning
RAID level	Determines how data is stored on the logical unit and whether there is data redundancy.
Block size (bytes)	Displays the block size used by the array controller to send data to the host.
Size of array (block)*	Sets the size of the logical unit.
Segment size	Determines the amount of data written to a single drive in the logical unit before the controller writes data on the next drive.
Segment 0 size	Displays the size of the first segment in the logical unit.
Reconstruction delay interval	Sets the amount of time between reconstruction operations.
Blocks to reconstruct per delay interval	Sets the number of blocks reconstructed in one reconstruction operation.
Select Drives*	Determines what drives make up the logical unit.
Reservation lock	Determines if other hosts may utilize a logical unit.
Write caching enabled	Determines if write cache is to be used.
Write cache method	If write cache is enabled, this is the write cache method used.
Size of read prefetch	Determines the multiplier used for prefetching blocks into cache for reads.
Command Queue Depth	Sets the command queue to be used for a logical unit.

Note: Changing the RAID level or drive map forces you to change those parameters for all the logical units in that drive group.

*Parameters can be set or changed only from the Create a SCSI RAID Array SMIT menu. The defined values are displayed for information only in the Change/Show SCSI RAID Array SMIT menu.

RAID Level: The RAID level parameter determines how data is stored on the logical unit. Data is either striped or mirrored. It also determines whether there is data redundancy. RAID Levels 1 and 5 offer data redundancy; RAID level 0 does not. Which levels you select depends on your storage and performance needs. See “RAID Levels” on page 5 for a discussion of all the RAID levels.

Note that the AIX Logical Volume Manager provides mirroring capability without using RAID 1 on the 2102 Storage Server. However, mirroring with the 2102 Storage Server off-loads the CPU and AIX from doing the mirroring.

The RAID level also determines the number of drives that can be included in the logical unit and the maximum size of that unit.

If you want to create a new logical unit from space in an existing drive group, the Fibre Channel Storage Manager automatically assigns the new unit the same RAID level as the other units in the drive group.

To change the RAID level of an existing drive group, there must be sufficient space within the drive group to accommodate the needs of the new RAID level. For instance, when switching from RAID 0 to RAID 1, no more than half of the space in the drive group must be in use by LUNs. When switching from RAID 0 to RAID 5, there must be at least one drive's worth of free space. RAID 1 can be freely transitioned to any other level, and RAID 5 can be freely transitioned to RAID 0.

Block Size: The Block Size parameter displays the block size used by the array controller to send data to the host system. The default logical block size of RAID levels 0, 1, and 5 is 512 bytes.

Size of Array: The Size of Array parameter sets the size of the logical unit. Size is determined by the RAID level, number of drives, and the amount of space allocated on those drives. Unallocated space on a set of drives (a drive group) is displayed below the lines showing the configured logical units in the drive group. This space can be used to create other logical units. By default, all available space is allocated for a logical unit.

To determine the size of a logical unit you want to create on your array subsystem, you need to know the capacity of a single drive. To determine this value, when you first receive your multi-bank array subsystem, select the **List All SCSI RAID Array** option from the FC SCSI RAIDiant Disk Array Manager menu in SMIT to determine the space of all preconfigured five-drive RAID level 5 logical units.

To determine the size of one individual drive in a subsystem having one logical unit per drive group (bank), divide the logical unit size of a five-drive logical unit by 4. Four drives are used for data in a RAID level 5 logical unit, and one drive is used for parity. The result yields the approximate amount of space on each of your drives.

Each time you create a new logical unit from spare drives, use the following formulas to determine the maximum size of the logical unit you can create. This calculation depends on both the RAID level of the unit you want to create and the size of the drives:

RAID level 0
: Multiply the number of drives by the drive capacity.

RAID level 1
: Multiply the number of drives by the drive capacity and divide by 2.

RAID level 5
: Multiply one less than the number of drives by the drive capacity.

The resulting figure for each of the preceding calculations yields the approximate size of the logical unit, if you were to use the total capacity of the selected drives.

Segment Size: The Segment Size parameter specifies the amount of data written on a single drive in the logical unit before the controller continues writing the data on the next drive in the logical unit. For example, if the segment size of a RAID level 0 logical unit is 512 blocks, the controller will write 512 blocks of data on drive 1, the next 512 blocks of data on drive 2, the next on drive 3, and so on.

The recommended segment size is 512 blocks for RAID levels 0 and 1. The Fibre Channel Storage Manager accepts a segment size as large as 65,535 blocks (although you should not use an odd number of blocks if you want to enhance array performance).

The recommended segment size for RAID 5 is the largest I/O operation commonly performed to the LUN. If this is not known, 128 or 256 blocks are good values. A smaller value will cause single I/O operations to be split into multiple internal operations, while a larger value may cause unnecessary I/O to fill out the segment. If most I/O is sequential, pick the higher value; for database or other random I/O, pick the lower. Also see the performance tuning selection.

The segment size for an existing LUN can be changed to the next larger or smaller size, to a minimum size of 32 blocks and a maximum size of 512 blocks in multiples of two. For instance, a LUN with a segment size of 128 blocks may be changed to 64 or 256 blocks. If a larger change is desired, it must be made in multiple operations.

Segment 0 Size: The first segment on a drive unit is segment zero. The default segment 0 size is zero (0) blocks. If you change the value of segment zero to anything but zero, the array segments may not be properly aligned on the drives and array performance may decrease.

Attention: Changing the segment zero size deletes any data on the logical unit. Use this parameter only after backing up all data in all logical volumes and file systems on the affected logical units.

Reconstruction Delay Interval and Blocks to Reconstruct Per Delay Interval: These parameters control the speed of data reconstruction on RAID level 1 and 5 logical units. Data on a RAID level 0 array cannot be reconstructed. Normally, you do not need to set these parameters until you are actually reconstructing data on a drive. For more information on these parameters, as well as advice on parameter settings, refer to “The Reconstruction Rate” on page 57.

Select Drives: Channel SCSI ID...: The Select Drives: Channel SCSI ID... parameter defines the drives included in the logical unit. Individual drives are identified by channel number and SCSI ID, which is determined by the physical location of the drive in the array subsystem (or integrated array). Each drive bank in an array has the same SCSI ID, and each bank contains channels 1 through 6.

Attention: You can change the drive map only after you delete the logical unit. If there are other logical units in the same drive group as the logical unit you want to change, you must delete and recreate all these logical units. Use this parameter only after backing up all data in all logical volumes and file systems on all affected logical units.

Figure 7 on page 20 shows channel numbers and SCSI IDs of a 2102 Storage Server.

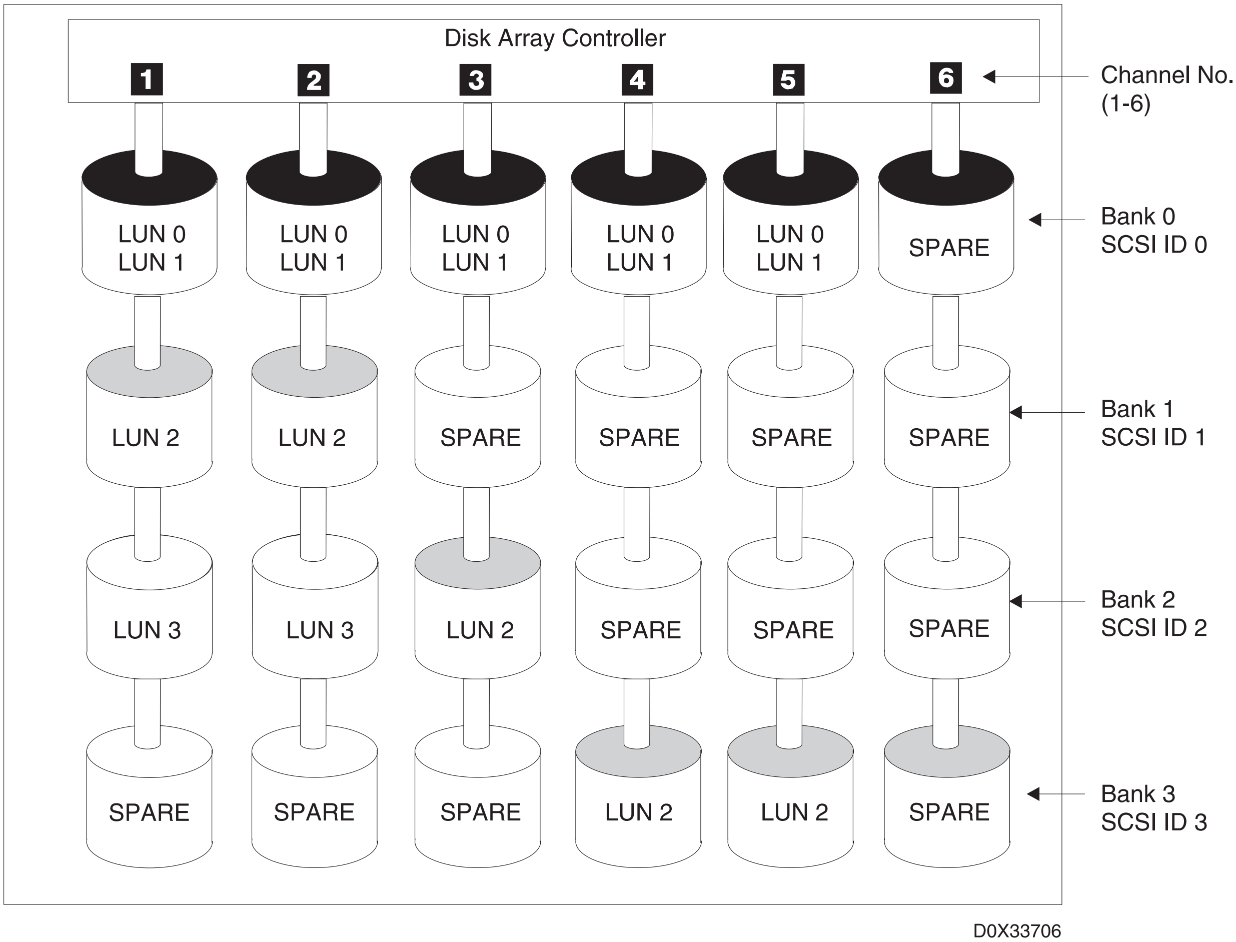

Figure 7. Channel Numbers and SCSI IDs

The RAID level of the logical unit sets some restrictions on drive selection:

0 Each logical unit is allowed 1 to 20 drives.

1 Each logical unit is allowed 2 to 30 drives. You must specify an even number of drives.

The mirrored pairs are created by grouping entries, for example, the first and second drives entered, or the third and fourth entered.

5 Each logical unit is allowed 3 to 20 drives.

Note: The drives you select for a logical unit depend on your storage and performance needs. In general, always use the maximum number of drives in a logical unit.

Defining a RAID level 5 logical unit as containing drives (1, 3), (2, 3), (3, 3), (4, 3), and (4, 0) is not robust, because the last two drives are on the same channel (channel 4). However, defining a RAID level 5 logical unit as containing drives (1, 3), (2, 3), (3, 3), (4, 3), and (5, 3) is robust; because, although the drives have the same SCSI ID, they are on different channels. Similarly, defining a RAID level 1 logical unit as (1, 3), (1, 0), (2, 3), and (2, 0) is not robust, since the first and second drives entered must be a mirrored pair. The third and fourth drives have the same problem.

If you want to change the drive selections for an existing logical unit, you must first delete the unit and then recreate it with the new drives. In addition, if there are other logical units in the same drive group, you must delete all of them and then recreate them. If you want to create a new logical unit from space in an existing drive group, that logical unit must include the same drives as the other logical units in the drive group.

Note: If you want to change cabling, refer to the *FCRDAM and Expandable Storage Unit, Service Guide, IBM 2102 Model F10 and Model D00.*

Reservation Lock: The Reservation Lock parameter determines if other hosts may utilize a logical unit. If this option is selected, the logical unit holds a lock on reservations attempted by other hosts. Otherwise, other hosts are able to utilize this logical unit.

Note: When creating or modifying a logical unit (LUN), IBM does not recommend setting the Reservation Lock field to NO in a multiple host environment. Setting this field to NO in a multiple host environment can expose the system to file corruption unless an application exists to coordinate access to the shared disk.

Write Caching Enable and Write Cache Method: If the Write Cache parameter is enabled, you can select either the fast-write or fast-load option. The fast-write option offers advantages over fast-load. With the fast-write option, writes are cached only when there is sufficient charge in the batteries on both controllers to hold the data in the event of a power or controller failure. In addition, the cache data is mirrored in the cache of the second controller, providing added protection in the event of a power or controller failure. The fast-load option is typically used instead of fast-write when loading the data from an archive.

Note: The write-cache method chosen must also be supported by the 2102 Storage Server hardware.

Size of Read Prefetch: The Size of the Read Prefetch field is the multiplier for the amount of disk blocks to prefetch on a read operation. Valid values are 0 to 100. For example, select 1 to have the same number of blocks prefetched as you request in the read; select 0 to disable the read-ahead function.

Command Queue Depth: The Command Queue Depth parameter allows the user to customize the depth of the adapter command queue utilized for a particular logical unit. A typical default value for this parameter is 30; however, customization for your particular system may improve performance.

Adding Logical Units to the Operating System

After you configure a logical unit using the Fibre Channel Storage Manager, you must add that logical unit to a volume group and create logical volumes and file systems on it before you can use it. Use standard AIX procedures to do this; treat the array in the same way you would treat a standard disk drive.

See “Adding Logical Units to the Operating System” on page 46 for the procedure required to create a file system on a configured logical unit so that you can use it. This procedure makes the logical unit its own volume group and creates a single logical volume or file system on it. If you want to add the logical unit to an existing volume group, extend a logical volume or file system onto it, or change other file system options, see the procedures covered in your AIX system documentation.

Attention: The AIX Logical Volume Manager has a limit of 1016 physical partitions in a physical volume. When creating AIX volume groups, you need to select a physical partition size sufficiently large so that you do not have more than 1016 physical partitions per physical volume. For example, if you are creating a volume group composed of a RAID 5 LUN composed of 5 physical disks of 2 GB each, you will need to use a physical partition size of 8 MB minimum; the default size of 4 MB will exceed the limit.

You will know you have exceeded the 1016 physical partitions per physical volume (hdisk) if you receive this message while trying to create a volume group on a 2102 Storage Server hdisk:

```
0516-862 mkvg:  Unable to create volume group
```

In general, you should not mix hdisks with different RAID characteristics in the same volume group. If you mix RAID 0 and RAID 5 hdisks, and a file system has partitions on both hdisks, a drive failure in the RAID 0 LUN will cause you to lose all your data in the file system.

If performance is important, you should plan your volume groups and logical volume placement before you create them because recreating them and restoring your data can be time-consuming. You can get performance-planning assistance from the AIX Support Family's Consult Line service.

Deleting Logical Units from the AIX Operating System

It is necessary to delete a logical unit whenever you want to change or modify most logical unit parameters. The procedures required are basically the reverse of those used to add logical units. The procedures are described in "Deleting Logical Units from the Operating System" on page 47. See Table 2 on page 17 for a list of parameters that may be changed.

Extending Logical Volumes and Logical Units

You can extend an existing logical volume onto a logical unit just as you can with any other disk storage device. However, you should remember that the data reliability of a combined logical volume is only as great as the reliability of its weakest part. For example, if you combine a RAID level 5 logical unit and a RAID level 0 logical unit into the same logical volume, the resulting volume has the data redundancy protection of a RAID level 0 logical unit (that is, none). Even though the RAID level 5 portion of the logical volume has data redundancy, if the RAID level 0 portion fails, the whole volume may become unusable, and all data on the volume will be lost. However, a RAID level 1 and a RAID level 5 logical volume would offer data redundancy.

In general you should not combine redundant and non-redundant disk storage systems in the same logical volume.

Performance Tuning

Array performance depends on many different factors, including RAID level and I/O block size. Here are a few suggestions that may improve performance on your array.

- Make your segment size at least the same size as your average I/O size and not larger than your maximum I/O size, especially for larger maximum I/O sizes. For example, if your average I/O size is 16 KB (32 blocks), change your segment size to 32 blocks as well.
- If data redundancy is not important, try using RAID level 0 instead of RAID level 1 or 5. In some circumstances, RAID level 0 provides high performance. However, all data will be lost if a single drive fails.

Load Balancing

In a dual active controller environment, options available through the FC SCSI RAIDiant Disk Array Manager Router menu allow you to enable dynamic load balancing. With load balancing enabled, I/O requests or data transfer rates are monitored. LUN ownership is switched between the controllers to maintain a balance of array controller ownership, which may result in some performance gains.

See “Using Dynamic Load Balancing” on page 48 for more information.

Do not enable load balancing on a multi-initiator configuration.

Overview of the Fibre Channel Storage Manager for AIX

The Fibre Channel Storage Manager software operates in an AIX Version 4.3.2 (3/99 update) (or higher) environment. The Storage Manager allows you to configure, change, and restore logical units (LUNs), as well as switch logical unit ownership between the two disk array controllers. It is also used to monitor and repair disk arrays in the AIX system environment. Some of the tasks performed using the Fibre Channel Storage Manager are:

- Check device status for the array devices on your system.
- Create and delete SCSI RAID arrays (LUNs) or subarrays.
- Configure and reconfigure SCSI RAID arrays (LUNs).
- Change the Fibre Channel Storage Manager LUN configuration attributes.
- Check and repair array parity on SCSI RAID arrays (LUNs).
- Change or show SCSI RAID array (LUN) ownership.
- Display disk array router configuration attributes.
- Restore a RAID level 1 or 5 LUN after a single drive failure.
- Restore a RAID level 0 LUN after a drive failure, or a RAID level 1 or 5 LUN after multiple drive failures.

The SMIT FC SCSI RAIDiant Disk Array Manager menus can be started from the SMIT System Management menu. Refer to the following topics for information about starting and using the FC SCSI RAIDiant Disk Array Manager menus.

- “Accessing the FC RAIDiant Disk Array Manager Menu” on page 35
- “Fibre Channel Storage Manager Tasks and Procedures” on page 24

FC SCSI RAIDiant Disk Array Manager Menu

After the SMIT FC SCSI RAIDiant Disk Array Manager is invoked, the following options are available to the user.

- **Disk Array Router Configuration**

Use this option to:

– Display the disk array routers (DARs) configured on the system and the contents and status of each DAR, such as disk array controllers and disks configured for each DAR.
– Change the notification frequency value for health checks of the controllers, automatic error notification, and, if activated, load balancing.

- **Disk Array Controller Configuration**

  Use this option to:

  – List, add, or configure a disk array controller
  – Change or show disk array ownership
  – Remove a disk array subsystem

- **Parity Check/Repair**

  Use this option to manually run array parity check and repair. You will need to run parity check/repair after an abnormal system shutdown and whenever there is an error reported from Parity Scan. Checking parity ensures the integrity of the data on a RAID level 1 or 5 LUN.

  See “Checking and Repairing Array Parity” on page 58 for more information.

- **FC SCSI RAIDiant Disk Array Manager**

  This option uses the FC SCSI RAIDiant Disk Array Manager to perform the following functions:

  – List, create, change, or delete SCSI RAID arrays
  – Change or show disk drive status. This function is primarily used to reconstruct a failed disk drive.

Fibre Channel Storage Manager Tasks and Procedures

Table 3 presents some array-specific tasks and the conditions that require the user to perform a specific task.

Note: *Do not* use this table for the initial installation. This table identifies the major tasks associated with setting up and configuring the 2102 Storage Server. Go to “Installing a 2102 Storage Server for the First Time on AIX Version 4” on page 31 for complete installation instructions.

Table 3. Fibre Channel Storage Manager Tasks

Task	When to Perform	Refer to:
Configure the logical units	When you install the first array subsystem on your system	“Chapter 4. Configuring the 2102 Fibre Channel RAID Storage Server” on page 37
Create a logical unit	When you create a disk array for the first time	“Chapter 4. Configuring the 2102 Fibre Channel RAID Storage Server” on page 37
Delete a logical unit	When you want to remove your logical unit	“Chapter 4. Configuring the 2102 Fibre Channel RAID Storage Server” on page 37

Table 3. Fibre Channel Storage Manager Tasks (continued)

Task	When to Perform	Refer to:
Change the reconstruction rate of a logical unit	When you want to change your logical unit configuration after initial installation	"Changing Reconstruction Rate Parameters" on page 57
Restore the logical unit	Whenever you get an error message indicating that one or more drives have failed or developed performance problems. Restoring a logical unit involves replacing the failed drives and then reconstructing the data on the array (for RAID level 1 and 5) or reformatting the array and recopying backed-up data to it (RAID level 0)	"Restoring and Recovering Logical Units" on page 59
Check and repair parity on a logical unit	After an abnormal system shutdown.	"Checking and Repairing Array Parity" on page 58

Chapter 2. Planning the 2102 Fibre Channel RAID Storage Server Installation

The IBM 2102 Fibre Channel RAID Storage Server is ideally suited for RS/6000 applications requiring large storage capacities and high data availability. The 2102 Storage Server offers high-reliability disk drives, redundant power supplies, redundant cooling, and concurrent maintenance on power supplies, cooling, drives and controllers.

The 2102 Storage Server can be configured to implement high-availability RAID-1 or RAID-5 architectures. With its disk drives configured in a RAID array, the 2102 Storage Server offers continued data availability in the event of a single disk drive failure, and performs automatic data reconstruction when the disk drive is replaced. Individual disk drives are hot-pluggable and can be replaced by the customer in case of a disk drive failure.

System throughput and I/O response time can be further enhanced by configuring the two controllers in a dual active controller configuration. In this configuration, both controllers are dedicated to different groups of drives. If a failure is detected in either controller, control of all disk drives is passed to the alternate.

The 2102 Storage Server is designed to be mounted in a standard 19-inch rack.

This chapter contains the following topics:

- Host adapters
- Configurations and cabling
- Selecting RAID levels

Host Adapters

The 2102 Storage Server requires the use of one or more fibre channel SCSI adapters.

Configurations and Cabling

If you configure the 2102 Storage Server with a single controller accessed through a single host adapter, planning its connection to your system is similar to adding any standard SCSI disk. However, single-controller/single-adapter configurations are not recommended for installations that require protection from single-points-of-failure.

If the 2102 Storage Server is configured with a second controller connected to the host through a second adapter and a separate SCSI bus, single-points-of-failure are eliminated. Such a configuration takes advantage of the high-availability features offered by the 2102 Storage Server.

In the dual-controller/dual-adapter configurations, each adapter and controller combination defines a path between the host and the disk array. The **fcparray** device driver software monitors the status of the components in the paths and manages the switching between the paths in the event a component failure makes a path unavailable.

The following diagrams describe various supported configurations.

Note: The following diagrams are provided to visually show the possible system, array, and controller configurations. The cabling depicted in each diagram may not show all of the possible cabling choices available for a particular configuration. For detailed information on cables and cable part numbers, see *Adapters, Devices, and Cable Information for Multiple Bus Systems*.

For all configurations, 50-micrometer multimode shortwave fibre optic cable is required. Media interface adapters (MIAs) are required on each 2102 Storage Server controller connection, and are not shown.

Single Host with One Adapter – Single 2102 Storage Server

This configuration is an example of a minimum configuration and is not recommended as a high-availability configuration.

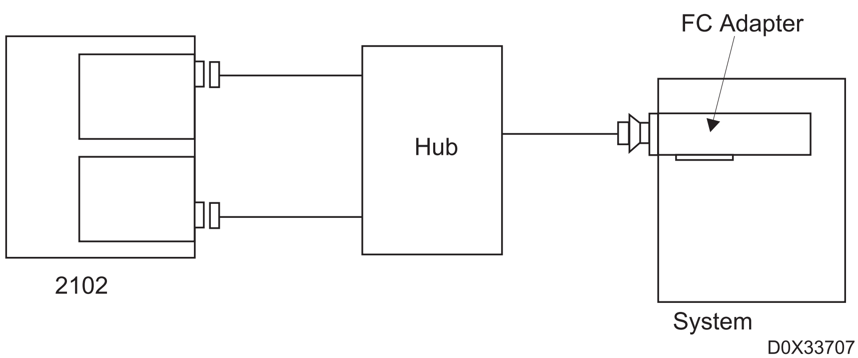

Figure 8. Single or Dual-Array Controllers/Single 2102 – Single Adapter/Single Host

Single Host with Dual Adapters – Single 2102 Storage Server

This configuration is the minimum recommended configuration. It consists of one host with two SCSI adapters and one 2102 Storage Server with two controllers.

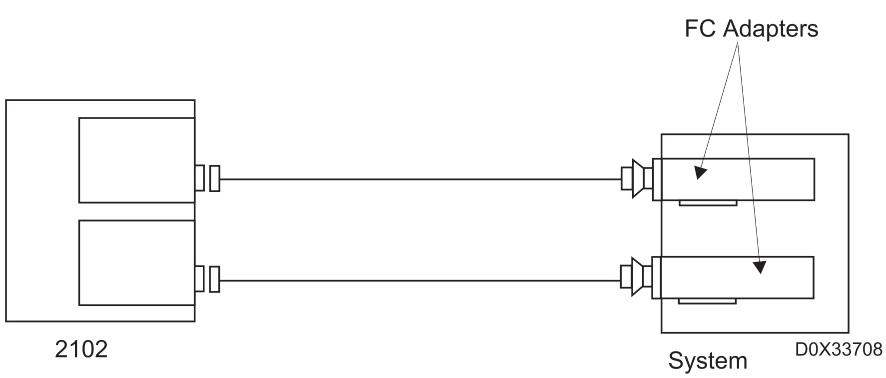

Figure 9. Dual-Array Controllers/Single 2102 – Dual Adapters/Single Host

Single Host with Dual Adapters – Dual 2102 Storage Server

This configuration consists of one host with two SCSI adapters and two 2102 Storage Server with two controllers each.

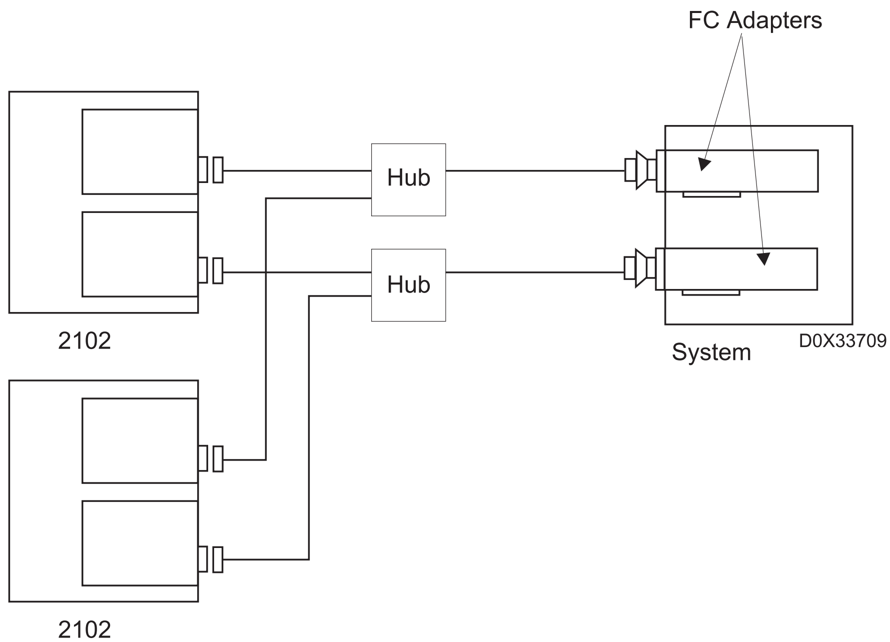

Figure 10. Dual-Array Controllers/Dual 2102s – Dual Adapters/Single Host

Selecting RAID Levels

Each of the RAID levels supported by the 2102 Storage Server offers differing degrees of performance and reliability. When choosing which RAID level to use, you should consider which RAID level best supports the characteristics of the applications accessing the data. For example, to support mission-critical applications, select a RAID level that provides data redundancy. Likewise, choice of RAID level can depend on the number of write operations an application performs. Characteristics of each RAID level indicate which types of applications they best support:

Level 0

In RAID level 0, data is striped, that is, spread across the drives in the array one segment at a time. A segment is a selectable number of blocks. Note that only a single copy of the data is stored to the disks and no parity information about the data is maintained. While RAID level 0 offers higher performance than other RAID levels, it does not offer any data redundancy.

Recommendation: Select RAID level 0 for applications that would benefit from the increased performance capabilities of this RAID level but, because no data redundancy is provided, do not use RAID level 0 for mission-critical applications that require high availability.

Level 1

In RAID level 1, data is mirrored, that is, two disks store identical information. This RAID level offers data redundancy and high availability but no significant performance gain. Note, however, that this RAID level has the highest overhead of all the RAID levels: only half the total storage capacity is usable.

Recommendation: Select RAID level 1 for applications with high levels of write operations, such as transaction files, and where cost is not a major concern.

Level 5

In RAID level 5, data is striped across the drives in the array and, in addition, parity information is maintained. However, instead of using a dedicated disk to store the parity information, RAID level 5 dedicates the equivalent of one entire disk for storing check data but distributes the parity information over all the drives in the group. Only two disks are involved in a single segment write operation: the target data disk and the corresponding disk that holds the check data for that sector.

The primary benefit of the RAID level 5 distributed check-data approach is that it permits multiple write operations to take place simultaneously. It also allows multiple reads to take place simultaneously and is efficient in handling small amounts of information.

Recommendation: Select RAID level 5 for applications that manipulate small amounts of data, such as transaction processing applications, or if the high drive space requirement of RAID 1 cannot be tolerated.

Chapter 3. Installing and Verifying the 2102 Fibre Channel RAID Storage Server

This chapter contains the following topics:

- Installing your 2102 Fibre Channel RAID Storage Server
- Installing a 2102 Fibre Channel RAID Storage Server software update
- Accessing the FC SCSI RAIDiant Disk Array Manager menu
- Verifying the 2102 Storage Server software installation

Installing Your 2102 Fibre Channel RAID Storage Server

To correctly install the 2102 Storage Server and the associated software, you must perform the following steps.

Attention: Failure to follow these steps may result in the LUNs being incorrectly identified to the operating system and will not allow further configuration of the array subsystem.

You will also need to refer to the *FCRDAM and Expandable Storage Unit, Service Guide, IBM 2102 Model F10 and Model D00*, which explains the 2102 Fibre Channel RAID Storage Server service aids.

The installation instructions for the 2102 Storage Server vary depending on whether or not you are installing the Storage Server for the first time.

- If you are installing a 2102 Storage Server for the first time, go to "Installing a 2102 Storage Server for the First Time on AIX Version 4".
- If you are you adding another 2102 Storage Server, go to "Installing an Additional 2102 Storage Server on an Existing AIX Version 4 System" on page 33.

Installing a 2102 Storage Server for the First Time on AIX Version 4

This section describes the steps for installing the 2102 Storage Server hardware and software for the first time on AIX Version 4.

Notes:

1. Do not run the diagnostics on the 2102 Storage Server at any stage during the installation of the Storage Server software. Wait until installation is complete before running diagnostics.
2. Array controller microcode and EEPROM files are distributed on the 2102 Storage Server controller microcode diskette included with the Storage Server and with additional controllers.

   Downloading of array controller microcode and EEPROM files is not normally required because those files are already installed on the controllers. Read and follow any instructions shipped with the 2102 Storage Server controller microcode diskette for situations in which download is required.

   If required, physical disk drive microcode will be shipped on a separate diskette with the disk drive modules. Follow the instructions included with that diskette.

Prerequisites

The prerequisites for installing the 2102 Storage Server hardware and software for the first time on AIX Version 4 are:

1. Back up the system.
2. Ensure that there is at least 8 MB of free space in the **/tmp** directory.
3. Ensure that the 2102 Storage Server is not connected to the host system at any time before installing the Storage Server software. If you do attach it and boot the system, you will have configured Other FC Disk Drives into the system. If you do this, remove these disk drives from the configuration (provided they are not being used) with the command:

   ```
   rmdev -l <hdisk#> -d
   ```

   Turn the 2102 Storage Server's power off after removing the FC disk drive definitions from the system, and leave it off until you have installed the 2102 Storage Server software.
4. Ensure that the 2102 Storage Server software is not already installed by using the command:

   ```
   lslpp -l devices.fcp.disk.array.rte
   lslpp -l devices.fcp.disk.array.diag
   ```

Actions

Attention: The install process requires that all activity on the 2102 Storage Server be halted. Make sure that all file systems are unmounted and all processes which access the 2102 Storage Server are halted on all machines attached to the Storage Server before proceeding.

The task can be accomplished through SMIT.

1. Insert the AIX Version 4 install media into the appropriate media drive.

   Note: The 2102 Storage Server requires AIX Version 4.3.2 (3/99 update) (or higher).

   To continue this task using SMIT, go to step 2.
2. Enter the following SMIT fast path:

   ```
   smit install_all
   ```

   a. Select an **Input Device** from the list.
   b. Select **devices.fcp.disk.array.ALL** from the list of **SOFTWARE to Install**.
   c. Press Enter to begin the installation process.
   d. When the installation process is finished, go to step 3.
3. During the installation process, the screen will display messages while the 2102 Storage Server software is loaded onto the system. If the messages indicate that the installation was not successful, repeat the installation procedure or contact your IBM service representative.
4. Shut down the host system.
5. Attach the 2102 Storage Server hardware to the host system and turn the power on to the Storage Server.
6. Restart the host system.
7. Verify the 2102 Storage Server software and controller connections. (See "Verifying the 2102 Fibre Channel RAID Storage Server Software Installation" on page 35.)

8. Run the command:

```
chdev -l dar#
```

where dar# is the device name of each disk array router (dar) on the system. The dar device names were displayed in step 8. This step puts the array controllers into dual-active mode.

9. Ensure that all disk array controllers are now active.

Note: If the instructions shipped with the 2102 Storage Server controller microcode diskette require downloading of microcode and EEPROM files, those actions would begin at this point.

The installation of the 2102 Storage Server hardware and software is not complete.

10. Run diagnostics to each DAC on the system to ensure all paths and hardware are functioning correctly.

Installing an Additional 2102 Storage Server on an Existing AIX Version 4 System

This checklist describes the required steps for attaching an additional 2102 Storage Server to a host system that already has the 2102 Storage Server software loaded. This includes attaching a 2102 Storage Server to a new host on which AIX and the 2102 Storage Server software have been preinstalled.

Notes:

1. Do not run diagnostics on the 2102 Storage Server at any stage during the installation of the Storage Server software; wait to run the diagnostics until the installation is complete.
2. Array controller microcode and EEPROM files are distributed on the 2102 Storage Server controller microcode diskettes included with the Storage Server and with additional controllers.

Prerequisites

- Use the *FCRDAM and Expandable Storage Unit, Service Guide, IBM 2102 Model F10 and Model D00* to ensure that the new Storage Server is correctly attached to the host.
- No users should be using the system, since installation requires that the system will be power-cycled.

Actions

1. Verify the 2102 Storage Server software and controller connections. (See “Verifying the 2102 Fibre Channel RAID Storage Server Software Installation” on page 35.)
2. Run the **cfgmgr** command.
3. Run the following command to check whether the disk array controllers (DACs) and the disk array router (dar) for the new 2102 Storage Server are configured. Enter:

```
lsdev -C | grep 2102
```

If the DACs and dar are not available, return to step 1 and begin again.

4. Run the command:

```
chdev -l dar#
```

where dar# is the device name of each disk array router (dar) on the system. The dar device names were displayed in step 1. This step puts the array controllers into dual-active mode.

5. Ensure that all disk array controllers are now active.

   Note: If the instructions shipped with the 2102 Storage Server controller microcode diskette require you to download the microcode and EEPROM files, perform those actions now.

   The installation of the new 2102 Storage Server is now complete.

6. Run diagnostics to each DAC on the system to ensure all paths and hardware are functioning correctly.

Installing a 2102 Fibre Channel RAID Storage Server Software Update

This checklist describes the steps required to update the 2102 Storage Server software on a host system that is already running the 2102 Storage Server software. These instructions are valid for any supported level of the AIX operating system.

Prerequisites

- Back up the system
- Ensure there is at least 8 MB of free space in the **/tmp** directory

Action

Attention: The download process requires that all activity on the 2102 Storage Server be halted. Make sure that all file systems are unmounted and all processes which access the 2102 Storage Server are halted on all machines attached to the Storage Server before proceeding.

1. Using SMIT, remove all devices associated with the 2102 Storage Server by entering the following SMIT fast path:
   ```
   smit fraidmgr
   ```
2. Insert the new 2102 Storage Server software media.
3. Enter the following SMIT fast path:
   ```
   smit update_all
   ```
4. Select the step appropriate for your level of the AIX operating system:

   For AIX Version 4, select an **Input Device** from the list (default tape drive).

   a. Select **devices.fcp.disk.fcparray.ALL** from the list of SOFTWARE to Install.

   b. Toggle **COMMIT** software to **no** and **SAVE** replaced files to **yes**.
5. Press Enter to begin the installation process.

   During the installation process, the screen will display messages while the **fcparray** 2102 Storage Server software is loaded onto the system. If the messages indicate that the installation was not successful, repeat the installation procedure or contact your IBM service representative.
6. Shut down the host system.
7. Restart the host system.
8. Verify the 2102 Storage Server software and controller connections. (See “Verifying the 2102 Fibre Channel RAID Storage Server Software Installation” on page 35.)

Accessing the FC RAIDiant Disk Array Manager Menu

To access the SMIT FC RAIDiant Disk Array menu, enter the following SMIT fast path:

```
smit fraidmgr
```

The screen then displays the FC RAIDiant Disk Array Manager menu. See "FC SCSI RAIDiant Disk Array Manager Menu" on page 23 for a complete description of the items in this menu.

Verifying the 2102 Fibre Channel RAID Storage Server Software Installation

Perform the following steps to verify the 2102 Storage Server software installation.

1. Select the step appropriate for your level of the AIX operating system:

   For AIX Version 4, verify the 2102 Storage Server software installation by entering the following command:

   ```
   lslpp -l devices.fcp.disk.array.rte
   lslpp -l devices.fcp.disk.array.diag
   ```

   The displayed text should indicate that the 2102 Storage Server software is in the applied state.
2. Select the **Disk Array Router Configuration** option from the FC RAIDiant Disk Array Manager menu in SMIT

   Note: For complete instructions on accessing the FC RAIDiant Disk Array Manager menu through SMIT, see "Accessing the FC RAIDiant Disk Array Manager Menu".

   The FC RAIDiant Disk Array Router Configuration menu is displayed.
3. Select the **List Disk Array Router Configuration** option. This selection results in a screen display similar to the following:

   ```
   dar0 Available   2102-F10 Disk Array Router
     dac0 Available   00-02-01-0  2102 Disk Array Device ACTIVE
       hdisk4 Available  00-02-01-0 2102 Disk Array Device
       hdisk3 Available  00-02-01-0 2102 Disk Array Device
       hdisk2 Available  00-02-01-0 2102 Disk Array Device
       hdisk1 Available  00-02-01-0 2102 Disk Array Device
     dac1 Available    00-02-01-0  2102 Disk Array Device ACTIVE
   ```
4. Check the list displayed. You should see from one to eight new disks of the type `2102 Disk Array Device` listed for each Storage Server you have installed. The number displayed depends on how many banks are in your Storage Server. Also, each 2102 Storage Server controller should display `ACTIVE` at the end of the display line. See the `dac0` and `dac1` examples in step 3.
5. If you are in the middle of installing a 2102 Storage Server or the dual-active Storage Server software support, and you have not yet downloaded microcode, then return to the installation instructions. Otherwise, continue with these instructions.
6. If the display does not show from one to eight new, available disks:

- Check the hardware installation to make sure the subsystem is installed correctly, then display the devices again. If the new disks are still not displayed, call your service representative.
- If the controller hardware was connected to the system before the 2102 Storage Server software was installed, the disks will be identified as `Other FC SCSI Disk`. If this condition exists, see “Removing Incorrectly Configured LUNs at Installation” on page 51 for recovery procedures.

If the disk array controllers are not in an active state, go to “Adding a 2102 Fibre Channel RAID Storage Server Controller to the Operating System” on page 52 to reconfigure the controller.

7. After you have verified the installation, go to “Chapter 4. Configuring the 2102 Fibre Channel RAID Storage Server” on page 37.

Chapter 4. Configuring the 2102 Fibre Channel RAID Storage Server

Configuring the 2102 Storage Server involves creating and configuring logical units and, in some cases, redefining various controller or router attributes. All of these tasks can be accomplished using SMIT or, for some tasks, the AIX command line. In most instances, logical units are managed using the Fibre Channel Storage Manager.

Logical units can be added, deleted, or combined into logical volumes through the use of naming conventions and location codes. Configuration of logical units is possible using the Fibre Channel Storage Manager and defaults and is described in “Fibre Channel RAID Array Configuration” on page 38.

Note: If you are upgrading your 2102 Storage Server software, go to “Fibre Channel RAID Array Configuration” on page 38 for instructions.

Refer to the following information for other configuration concepts and procedures:

- “Understanding the Use of Logical Units in a SCSI RAID Array” on page 16
- “Overview of the Fibre Channel Storage Manager for AIX” on page 23
- “Restoring and Recovering Logical Units” on page 59
- “Checking and Repairing Array Parity” on page 58

This chapter contains the following topics:

- Starting the Fibre Channel Storage Manager for AIX
- SCSI RAID Array configuration
- Creating a subarray from an existing drive group
- Modifying and displaying drive status
- Adding logical units to the operating system
- Deleting logical units from the operating system
- Changing disk array ownership
- Using dynamic load balancing
- Changing the controller health-check frequency
- Changing the automatic error notification frequency
- Disabling or enabling write cache
- Changing 2102 Fibre Channel RAID Storage Server settings

Starting the Fibre Channel Storage Manager for AIX

You can access the Fibre Channel Storage Manager for AIX using the System Management Interface Tool (SMIT). SMIT is a menu-based user interface that constructs commands from the options you choose and executes them.

To start the Fibre Channel Storage Manager using SMIT, perform the following steps:

1. Enter the following SMIT fast path:

   ```
   smit fraidmgr
   ```

2. From the RAIDiant Disk Array menu, select the FC SCSI RAIDiant Disk Array Manager option.

   The screen then displays the RAIDiant Disk Array Manager menu.

Fibre Channel RAID Array Configuration

The procedures to configure or reconfigure the logical units in your 2102 Fibre Channel RAID Storage Server using advanced configuration depend on whether you are installing the 2102 Storage Server for the first time or changing configuration parameters after installation.

Note: Before you can perform this procedure, all appropriate hardware and software must be installed. See "Host Adapters" on page 27 and "Chapter 3. Installing and Verifying the 2102 Fibre Channel RAID Storage Server" on page 31 for additional information on this hardware and software.

If you are installing the 2102 Storage Server for the first time, and:

- You want to change the factory-set configuration, use the Fibre Channel Storage Manager to reconfigure the logical unit using the procedures in "Changing the RAID Level, Drive Selection, or Size of Array Parameters".
- You want to add the logical unit to a volume group and create logical volumes and file systems, refer to "Adding Logical Units to the Operating System" on page 46.

If you are changing configuration parameters after installation or for a software upgrade:

1. Back up the data if there is any data in the logical volumes or file systems on the logical unit.
2. Unmount all file systems on the logical unit you want to reconfigure.
3. Delete all logical volumes and file systems from the logical unit you want to reconfigure, and remove the logical unit from its volume group. Refer to "Deleting Logical Units from the Operating System" on page 47.
4. Reconfigure the logical unit using the Fibre Channel Storage Manager by following the procedures in "Changing the RAID Level, Drive Selection, or Size of Array Parameters".
5. Add the reconfigured logical unit back to the volume group, and recreate the logical volumes and file systems on it. Refer to "Adding Logical Units to the Operating System" on page 46.
6. Restore the data from backup to the logical volumes and file systems on the logical unit, if necessary.

Note: Perform the appropriate set of tasks for each logical unit you want to configure or reconfigure.

Changing the RAID Level, Drive Selection, or Size of Array Parameters

During the initial installation of the SCSI RAID array, you can use this procedure to change the drive selection or the size of the array. To change the values defined for these logical unit parameters, you must:

- Delete the logical unit
- Recreate the logical unit

You can also use this procedure after installation if you want to change any of the logical unit parameters and no data has been written to the logical unit.

Attention: Failure to back up data before deleting the logical unit will result in a loss of data.

If you want to change any of the default values of a logical unit that has data on it, you must first back up the data in all logical volumes and file systems on the logical unit, delete the logical volumes and file systems from the logical unit, and then remove the logical unit from its volume group before continuing.

Also, if a subsystem has multiple LUNs per drive group, each logical unit in the drive group must be deleted before these logical unit's RAID levels can be changed. If data exists on this logical unit, a backup of the logical volumes and file systems on this second logical unit must be performed as well.

If you do not know whether there are other logical units in the drive group, or even what a drive group is, continue with “Deleting a SCSI RAID Array”. It contains steps that will enable you to determine this information.

Complete the following prerequisites; then use the “Deleting a SCSI RAID Array” procedure to change the RAID level, drive selection, or size of the array of an existing logical unit.

REQTEXT

Prerequisites:

1. Back up the data if there is any data in the logical volumes of file systems on the logical unit.
2. Unmount all file systems on the logical unit you want to reconfigure.
3. Delete all logical volumes and file systems from the logical unit you want to reconfigure, and remove the logical unit from its volume group.

Deleting a SCSI RAID Array

To complete any of the following tasks, you must first delete the logical unit:

- Change the RAID level of a logical unit
- Change the drive map of a logical unit
- Change the logical unit, block, or segment size of a logical unit
- Remove a drive from the array subsystem

If the logical unit is part of a volume group, you must first back up all files in the logical volumes and file systems on the logical unit and remove the logical unit from its volume group. See “Deleting Logical Units from the Operating System” on page 47.

Use the following SMIT procedure to delete a SCSI RAID array:

1. Select the FC SCSI RAIDiant Disk Array Manager option from the RAIDiant Disk Array menu.
2. Select the Delete a SCSI RAID Array option. A display similar to the following appears:

```
Move cursor to desired item and press Enter.

hdisk4 Raid 0 00-02-00-1,4    1 MB Status Optimal
```

```
       hdisk4  32 Channel 3   ID 2  Optimal
hdisk0 Raid 5 00-02-00-0,0 3835 MB Status Optimal
       hdisk0  11 Channel 1   ID 1  Optimal
       hdisk0  21 Channel 2   ID 1  Optimal
       hdisk0  31 Channel 3   ID 1  Optimal
hdisk1 RAID 1 00-02-00-0,1 1917 MB Status Optimal
       hdisk1  20 Channel 2   ID 0  Optimal
       hdisk1  50 Channel 5   ID 0  Optimal
hdisk2 RAID 1 00-02-00-0,2 1917 MB Status Optimal
       hdisk2  30 Channel 3   ID 0  Optimal
       hdisk2  40 Channel 4   ID 0  Optimal
```

3. Select the RAID array (LUN) you want to change or delete.

   Attention: If there is any data on the other logical units in the drive group, do not proceed further unless you have backed up all data on these other logical units.
4. After receiving notification that the logical unit has been deleted, return to the RAIDiant Disk Array Manager menu.

Creating a SCSI RAID Array

Use the following SMIT procedure to create or recreate a SCSI RAID array:

1. Select the Create a SCSI RAID Array option from the FC SCSI RAIDiant Disk Array Manager menu.

   Note: You may create four arrays per controller. If you exceed this number, the results are unpredictable.
2. Select the Create a SCSI RAID Array option from the Create a SCSI RAID Array menu.
3. When prompted, select a disk array controller.
4. When prompted, select the new RAID level.
5. When prompted, select each drive you want in the logical unit.

   The following considerations apply when assigning drives:

   0 Each logical unit is allowed 1 to 20 drives.

   1 Each logical unit is allowed 2 to 30 drives. You must specify and even number of drives. The mirrored pairs are created by grouping entries, for example, the first and second drives entered, or the third and fourth entered.

   5 Each logical unit is allowed 3 to 20 drives.

   Note: The drives you select for a logical unit depend on your storage and performance needs. In general, always use the maximum number of drives in a logical unit.
6. Enter new values for the Create a SCSI RAID Array menu fields.

   Note: You cannot change the RAID Level parameter or drive list without restarting the creation procedure.

   Refer to the following information to determine the values you may use for the logical unit parameters:

   - “RAID Level” on page 17
   - “Size of Array” on page 18
   - “Reconstruction Delay Interval and Blocks to Reconstruct Per Delay Interval” on page 19

- “Select Drives: Channel SCSI ID...” on page 19
- “Reservation Lock” on page 21
- “Write Caching Enable and Write Cache Method” on page 21
- “Segment Size” on page 19
- “Size of Read Prefetch” on page 21
- “Command Queue Depth” on page 21

Note: Depending on the RAID level and the number and type of selected physical drives, it is possible to create a logical unit (LUN) up to 45 GB in size. However, the Logical Volume Manager has a limit of 1016 physical partitions per physical volume. For the default physical partition size of 4 MB, this translates to a maximum physical volume size of approximately 4 GB. Larger physical volumes can be configured by increasing the physical partition size.

You will know you have exceeded the 1016 physical partitions per physical volume (hdisk) if you receive this message while trying to create a volume group on a 2102 Storage Server hdisk:

```
0516-862 mkvg:  Unable to create volume group
```

7. Repeat step 6 on page 40 until you have modified all the values you want to modify.
8. After you have set all the parameters, press Enter. The array controller automatically formats the new logical unit. Wait for the format operation to finish.
9. If you need to configure more logical units, repeat the steps in this procedure. If you do not need to configure more logical units, exit the FC SCSI RAIDiant Disk Array Manager.
10. Go to “Adding Logical Units to the Operating System” on page 46 to add the configured logical units to your system.

Creating a Subarray from an Existing Drive Group

Use the following SMIT procedure to create a SCSI RAID subarray (logical unit) from space in an existing drive group. The new logical unit will use the same drives and the same RAID level as the other drives in the drive group.

1. Select the FC SCSI RAIDiant Disk Array Manager option from the RAIDiant Disk Array menu.
2. Select the Create a SCSI RAID Array option.
3. Select the Create a SCSI RAID Subarray option. A display similar to the following appears:

```
Move cursor to desired item and press Enter.

hdisk4 Raid 0 00-02-00-1,4    1 MB Status Optimal
       hdisk4  32 Channel 3   ID 2  Optimal
hdisk0 Raid 5 00-02-00-0,0 3835 MB Status Optimal
       hdisk0  11 Channel 1   ID 1  Optimal
       hdisk0  21 Channel 2   ID 1  Optimal
       hdisk0  31 Channel 3   ID 1  Optimal
hdisk1 RAID 1 00-02-00-0,1 1917 MB Status Optimal
       hdisk1  20 Channel 2   ID 0  Optimal
       hdisk1  50 Channel 5   ID 0  Optimal
```

```
hdisk2 RAID 1 00-02-00-0,2 1917 MB Status Optimal
       hdisk2  30 Channel 3   ID 0  Optimal
       hdisk2  40 Channel 4   ID 0  Optimal
```

4. Select the RAID array (LUN) on which you want to create the subarray.
5. Enter values for the Create a SCSI RAID Subarray menu fields.

   Refer to the following information to determine the values you may use for the logical unit parameters:
   - “Size of Array” on page 18
   - “Reconstruction Delay Interval and Blocks to Reconstruct Per Delay Interval” on page 19
   - “Reservation Lock” on page 21
   - “Write Caching Enable and Write Cache Method” on page 21
   - “Size of Read Prefetch” on page 21
   - “Command Queue Depth” on page 21
6. Repeat step 5 until you have modified all the values you want to modify.
7. After you have set all the parameters, press Enter.
8. The array controller automatically formats the new subarray. Wait for the format operation to finish.
9. If you need to configure more subarrays, repeat the steps in this procedure. If you do not need to configure more subarrays, exit the RAIDiant Disk Array Manager.
10. Go to “Adding Logical Units to the Operating System” on page 46 to add the configured logical units to your system.

Modifying and Displaying Drive Status

You can add a drive, delete a drive, fail a drive, or reconstruct drive data in a SCSI RAID array. There are eight possible values for the status of a drive. The FC SCSI RAIDiant Disk Array Manager displays the following values for drive status in the drive matrix:

Optimal The drive is functioning normally.

Non-Existent No drive is physically connected to the array at this position.

Spare The drive is connected to the array, but not configured into a logical unit.

Failed The drive was failed by the array controller or by the user and must be replaced.

Replaced The drive has just been replaced.

Mismatch The array controller sensed that the drive has either a sector size, capacity, serial number, SCSI channel, or ID different from what the array controller expected.

Formatting The drive is currently being formatted or is currently reconstructing.

Wrong Drive The wrong drive was replaced, and the logical unit is no longer accessible. See “Recovering from a ’Wrong Drive Replaced’ Error” on page 67 to revive the drive.

The following table describes the various actions performed and the new status of the drive after the action is complete.

Table 4. Modifying Drive Status

Action:	Status Becomes:
Mark a drive as operational	Spare
Delete a spare drive	Non-Existent
Fail an optimal drive	Failed
Mark a drive as a hot spare	Hot Spare
Mark a drive as a spare	Spare
Fail a drive	Failed
Format, replace, and reconstruct data on a failed drive	Replaced then Optimal
Replace and reconstruct data on a failed drive	Replaced then Optimal

Viewing the Current Drive Status

To view the current drive status in SMIT, select the List All SCSI RAID Arrays option from the FC SCSI RAIDiant Disk Array Manager menu.

A display similar to the following appears:

```
Move cursor to desired item and press Enter.

hdisk4 Raid 0 00-02-00-1,4    1 MB Status Optimal
       hdisk4  32 Channel 3   ID 2  Optimal
hdisk0 Raid 5 00-02-00-0,0 3835 MB Status Optimal
       hdisk0  11 Channel 1   ID 1  Optimal
       hdisk0  21 Channel 2   ID 1  Optimal
       hdisk0  31 Channel 3   ID 1  Optimal
hdisk1 RAID 1 00-02-00-0,1 1917 MB Status Optimal
       hdisk1  20 Channel 2   ID 0  Optimal
       hdisk1  50 Channel 5   ID 0  Optimal
hdisk2 RAID 1 00-02-00-0,2 1917 MB Status Optimal
       hdisk2  30 Channel 3   ID 0  Optimal
       hdisk2  40 Channel 4   ID 0  Optimal
```

Adding a Drive

Adding a drive through software causes the array controller to change the drive status from Non-Existent to Spare, thus making the drive available for you to create a logical unit.

1. Select the FC SCSI RAIDiant Disk Array Manager option from the RAIDiant Disk Array menu.
2. Select the Change/Show Drive Status option. Each disk array controller and its associated disks is displayed.
3. Select a drive with the Non-Existent status and press Enter.
4. Select Add New Drive from the list of options in the Drive Status field.
5. The status of the added drive changes from Non-Existent to Spare.

   Repeat the steps in this procedure until you have added all the drives you want to add.
6. Exit the FC SCSI RAIDiant Disk Array Manager.

   You may now create a logical unit using the drive.

Deleting a Drive

A drive must be deleted if it is to be removed from the array subsystem and not subsequently replaced. Deleting a drive through software causes the array controller to change the drive status from Spare to Non-Existent. You cannot delete a drive that is part of a logical unit. You must first delete all the logical units contained on the drive. See "Deleting Logical Units from the Operating System" on page 47 to delete the logical unit. Next, delete the drive and physically remove the drive from the subsystem.

If you are going to delete a drive that is currently in a logical unit, and there is data on the logical unit, back up the data in the logical volumes and file systems on the logical unit and delete the logical unit from its volume group. If there is more than one logical unit in the drive group, back up the data on these other logical units and delete the AIX structures from them.

1. Select the FC SCSI RAIDiant Disk Array Manager option from the RAIDiant Disk Array menu.
2. Select the Change/Show Drive Status option. All disk array controllers and associated disks are displayed.
3. Select the drive to be deleted.
4. Select Delete Drive from the list of options in the Drive Status field. The status of the deleted drive changes from Spare to Non-Existent.
5. Repeat the steps in this procedure until you have deleted all the drives you want to delete.
6. Exit the FC SCSI RAIDiant Disk Array Manager.
7. Remove the drives you deleted from the array.

Failing a Drive

Attention: Do not fail a drive in a RAID level 0 logical unit unless you want to replace the drive. RAID level 0 has no parity, and once a drive is failed in a RAID level 0 logical unit, you cannot recover the data. Do not fail a drive in a RAID level 1 or 5 logical unit if the logical unit is already degraded. See "Restoring and Recovering Logical Units" on page 59 for instructions on which drives to fail and when to fail them.

1. Select the FC SCSI RAIDiant Disk Array Manager option from the RAIDiant Disk Array menu.
2. Select the Change/Show Drive Status option. Each disk array controller and its associated disk is displayed.
3. Select the drive to be failed.
4. Select Fail Drive from the list of options in the Drive Status field. The status of the selected drive changes to Failed.

   Attention: Do not fail more than one Warning drive unless the recovery procedure you are using specifically states that you can fail more than one drive at a time (see "Restoring and Recovering Logical Units" on page 59). Failing two drives with this procedure could result in the loss of all data on all the logical units on the drive.
5. Repeat the steps in this procedure until you have deleted all the drives you want to fail.
6. Exit the RAIDiant Disk Array Manager.

7. After finishing this procedure, go to “Restoring and Recovering Logical Units” on page 59 to reconstruct the drive data, or go to “Replacing Failed Drives (RAID Levels 1 and 5)” on page 63 or “Replacing Failed Drives (RAID Levels 1 and 5)” on page 63 to reformat the logical unit.

Marking a Drive as a Hot Spare

This procedure will mark a drive as a hot-spare drive, which will take the place of any failed drive that is part of a RAID 1 or 5 array. If there is already a failed drive when a hot-spare drive is added, the hot-spare drive will immediately begin filling in for it.

To return a hot-spare drive to the spare state, simply follow the same steps, but choose Mark Drive Spare instead. The hot-spare drive must not be currently filling in for a failed drive.

1. Select the FC SCSI RAIDiant Disk Array Manager option from the RAIDiant Disk Array menu.
2. Select the Change/Show Drive Status option. All disk array controllers and associated disks are displayed.
3. Select the drive to be deleted.
4. Select Mark Drive Hot Spare from the list of options in the Drive Status field. The status of the hot-spare drive changes from Spare to Hot Spare (or, if there is already a failed drive, to Hot Spare Drive In Use).
5. Repeat the steps in this procedure until you have marked all the drives that you wish as Hot Spare.
6. Exit the FC SCSI RAIDiant Disk Array Manager.

Setting a Drive’s Visual Identification

Use this procedure to switch the visual identification light for a drive to either on or off. The visual identification light is a blinking light located just above the drive. You can use the light to identify a drive for service purposes.

Use the same steps to switch the visual identification on or off. Only the option you select varies.

Attention: Use the same controller when changing the visual identification status for a given drive. The system maintains the visual identification status on a per-controller basis.

1. Select the FC SCSI RAIDiant Disk Array Manager option from the RAIDiant Disk Array menu.
2. Select the Change/Show Drive Status option. All disk array controllers and associated disks are displayed.
3. Select the drive to be identified.
4. Select either the Turn on Drive Visual Identification option or the Turn Off Drive Visual Identification option from the list of options in the Drive Status field.

   The status of the visual identification does not change, but the identifier light changes.
5. Exit the FC SCSI RAIDiant Disk Array Manager.

Adding a Drive to an Existing Drive Group

Use this procedure to add a drive to an existing drive group. The capacity of the drive group will be increased, allowing you to either add subarrays or change the RAID level.

Notes:

1. Adding drives does not increase the size of the LUNs in the drive group. The LUNs are spread out over the added drives. However, after adding drives, the drive group will have the capacity to accommodate a new LUN.
2. There is no way to remove a drive from a drive group aside from deleting and recreating the drive group.
3. This procedure cannot be performed from SMIT. It must be performed from the command line.
4. This procedure is subject to change in later releases of the Fibre Channel Storage Manager for AIX.

From the AIX command line, use the following command to add a drive to the drive group:

```
fraidmgr -M -y ChanID -l hdisk#
```

where *ChanID* is the channel and ID of the drive to be added. For example, to add a drive at 5,0 to hdisk5, the command would be:

```
fraidmgr -M -y 50 -l hdisk5
```

Note: Only one drive may be added at a time. Because RAID 1 requires an even number of drives, this procedure does not allow drives to be added to a RAID 1 array. One way to add drives to a RAID 1 array is to convert the array to RAID 5 (which is possible only if the RAID 1 array has more than two drives), add two drives to the resulting RAID 5 array, then convert back to RAID 1. There is currently no recommended way to add drives to a two-drive, RAID 1 array.

Reconstructing a Drive

You can reconstruct a drive after replacing a failed drive in a degraded RAID level 1 or 5 logical unit. The procedure should be performed in the context of overall logical unit restoration. Refer to “Restoring and Recovering Logical Units” on page 59 for a description of logical unit restoration for each RAID level.

Adding Logical Units to the Operating System

Attention: The AIX Logical Volume Manager has a limit of 1016 physical partitions in a physical volume. When creating AIX volume groups, you need to select a physical partition size sufficiently large that you don’t have more than 1016 physical partitions per physical volume. For example, if you are creating a volume group composed of a RAID 5 LUN composed of 5 physical disks of 2 GB each, you will need to use a physical partition size of 8 MB minimum, the default size of 4 MB will exceed the limit.

Attention: When using the SMIT interface to set the characteristics of the logical volume, the **Enable BAD BLOCK relocation** must be set to **NO**.

You will know you have exceeded the 1016 physical partitions per physical volume (hdisk) if you receive this message while trying to create a volume group on a 2102 Storage Server hdisk:

```
0516-862 mkvg:  Unable to create volume group
```

The following procedure gives the basic steps required to create a file system on a configured logical unit so that you can use it. This procedure makes the logical unit its own volume group and creates a single logical volume or file system on it. If you want to add the logical unit to an existing volume group, extend a logical volume or file system onto it, or change other file system options, see the procedures covered in your AIX system documentation.

1. Enter the following SMIT fast path:

   ```
   smit mkvg
   ```

2. Enter values for the Add A Volume Group menu fields.
3. Press Enter to save these values. Then press the F3 key until you return to the Physical & Logical Storage menu.
4. Select the File Systems option from the Physical & Logical Storage menu.
5. Select the Add/Change/Show/Delete File Systems option.
6. Select the Journaled File System option. A list of volume groups is displayed. from the list of names displayed.
7. Select Add a Journaled File System.
8. Select the volume group that you just created from the list of names displayed.
9. Enter values for the Journaled File System menu fields.
10. Press Enter to create the journaled file system, then exit SMIT.

After completing this procedure, you can mount and use the newly created file system, using the specified mount point.

Deleting Logical Units from the Operating System

Attention: Failure to back up any data before performing this procedure could result in the loss of all data on the logical units.

It is necessary to delete a logical unit whenever you want to change or modify logical unit parameters. The procedures required are basically the reverse of those used to add logical units.

For each logical unit you want to delete, perform the following procedure:

1. Back up any data on the logical unit. There may be more than one logical volume and file system on the unit. To find out which logical volumes or file systems are on the logical unit, see “Identifying Hard Disks and Logical Units” on page 15.
2. Unmount any file systems on the logical unit.

Delete any file systems on the logical unit using the following steps:

3. Enter the following SMIT fast path:

   ```
   smit rmjfs
   ```

4. Select the file system to be removed.

Delete any logical volumes on the logical unit using the following steps:

5. Enter the following SMIT fast path:

```
smit rmlv
```

6. Select the logical volume to be removed.

Remove the logical unit from its volume group using the following steps:

7. Enter the following SMIT fast path:

   ```
   smit vg
   ```

8. Select either the Remove a Volume Group or Set Characteristics of a Volume Group option, depending on whether there are other physical volumes in the volume group.

   To find out which volume group the logical unit is in, see “Identifying Hard Disks and Logical Units” on page 15. Then go to “Deleting a SCSI RAID Array” on page 39 to delete the logical unit using the Fibre Channel Storage Manager, or go to “Fibre Channel RAID Array Configuration” on page 38 to reconfigure the logical unit.

Changing Disk Array Ownership

The disk array or LUN ownership between two active controllers is stored in the device drive. If an error condition or a change by another host in a dual-host configuration occurs, a swap may take place without notifying or updating the user. In these cases, the Object Data Manager (ODM) information for the active controllers and the lun_bitmap attributes are not always up to date. Using SMIT, the ODM is updated with the current device driver state.

Note: The swap between controllers is indicated by the occurrence of an FCP_ARRAY_ERR10 error being logged. This is a normal function of the software and is an information-logging message. It is not an indication of an actual error.

To change disk array ownership in SMIT, perform the following steps:

1. Select the Disk Array Controller Configuration option from the RAIDiant Disk Array menu in SMIT.
2. Select the Change/Show Disk Array Ownership option.
3. When prompted, select a disk array controller. Select the drive group that you want to switch. A list of drive groups can be displayed for selection.

   The selected drive group is now switched to the other active controller.

Using Dynamic Load Balancing

For a complete description of load balancing, see “Load Balancing” on page 23.

1. Select the Disk Array Router Configuration option from the RAIDiant Disk Array menu in SMIT.
2. Select the Change/Show Disk Array Router option.
3. Select the appropriate dar from the Disk Array Router pop-up.
4. When the Change/Show Disk Array Router menu is displayed, select the Load Balancing option and change the value to yes to activate dynamic load balancing or no to deactive dynamic load balancing.

   Note: IBM does not recommend setting the Load Balancing field to yes in a multiple host environment.

5. If you are activating dynamic load balancing, you can also select the frequency at which load balancing will occur by setting the value of the Load Balancing Frequency option.

   Load balancing can also be activated from the command line using the chdev command. To activate the dynamic load balancing option using chdev, enter the following command:

   ```
   chdev -l dar# -a load_balancing=yes
   ```

   Note: The load balancing results in an FCP_ARRAY_ERR10 error being logged. This is a normal function of the software and is an information logging message and not an indication of an actual error. There must be at least four drive groups for load balancing to occur. Load balancing is not recommended for multi-host operation.

Changing the Controller Health-Check Frequency

The **fcparray** device driver provides a monitoring capability, known as passive controller health checks, for testing components that are not actively in use. These checks include actual data transfers through the passive controller which ensure a complete path to the attached drives is functioning properly.

To set the Health-Check Frequency option in SMIT, perform the following steps:

1. Select the Disk Array Router Configuration option from the RAIDiant Disk Array menu.
2. Select the Change/Show Disk Array Router option.
3. Select the appropriate dar from the Disk Array Router pop-up.

   When the Change/Show Disk Array Router menu is displayed, select the Health Check Frequency option and select the frequency at which health check will occur.

   The Health Check Frequency option can also be activated from the command line using the chdev command.

   To change the frequency of the controller health checks, enter the following command:

   ```
   chdev  -l  dar#  -a healthcheck_freq=seconds
   ```

   where *seconds* is the number of seconds selected for frequency checks. Valid values are 1 to 9999. The default value is 600.

Changing the Automatic Error Notification Frequency

The **fcparray** device driver provides a monitoring capability, known as *automatic error notification* (AEN). This option allows you to set the frequency at which polled AEN request sense commands are issued to the controllers in the selected Storage Server. Any detected errors are logged by the system error logger.

To set the Automatic Error Notification Frequency option in SMIT, perform the following steps:

1. Select the Disk Array Router Configuration option from the RAIDiant Disk Array menu.
2. Select the Change/Show Disk Array Router option.
3. Select the appropriate dar from the disk Array Router pop-up.

When The Change/Show Disk Array Router menu is displayed, select the Automatic Error Notification Frequency option and select the frequency at which it will occur. The AEN frequency option can also be activated from the command line using the chdev command.

To change the frequency of the controller health checks, use the following command string:

```
chdev -l dar0 -a aen_freq=seconds
```

where *seconds* is the number of seconds selected for frequency checks. Valid values are 1 to 9999. The default value is 600, which indicates that all LUNs will be checked at evenly spaced intervals every 600 seconds.

Disabling or Enabling Write Cache

Note: Follow these instructions only when changing the write cache settings from what you set when creating the SCSI RAID array. See "Creating a SCSI RAID Array" on page 40 for instructions on initially setting the value of write cache.

To enable or disable write cache using SMIT, perform the following steps:

1. Select the Change/Show SCSI RAID Array option from the RAIDiant Disk Array menu. settings must be the same on both host machines.
2. Select the appropriate hdisk from the Change/Show SCSI RAID Array menu.
3. Enter values for the Change/Show SCSI RAID Array menu fields.

Changing 2102 Fibre Channel RAID Storage Server Settings

To change properties of the 2102 Storage Server, such as the RAID level, segment size, reconstruction rate, and reservation locking, perform the following steps.

1. Select the Change/Show SCSI RAID Array option from the RAIDiant Disk Array menu.
2. Select the appropriate hdisk from the Change/Show SCSI RAID Array menu.
3. Select the options to be changed and enter their new values.
4. Press Enter to make the changes, then exit the RAIDiant Disk Array Manager.

Chapter 5. Problem Determination and Recovery

This chapter provides information to assist you in determining and recovering from error conditions. The following topics are discussed in detail:

- Removing incorrectly configured LUNs at installation
- Adding a disk array controller to the operating system
- Drive failures and arrays
- Reconstructing RAID levels after a drive failure
- Checking and repairing array parity
- Restoring and recovering logical units

“Appendix B. FC SCSI RAIDiant Disk Array Manager Messages” on page 93 describes the error messages generated by the Fibre Channel Storage Manager for AIXand the fcparray software.

You may also refer to “Appendix A. Application Programming Interface (API)” on page 75 for the error messages returned from the **fcparray** SCSI device driver.

Removing Incorrectly Configured LUNs at Installation

Attention: The Fibre Channel Storage Manager for AIX software must be installed before attaching the 2102 Fibre Channel RAID Storage Server controller hardware. If the Fibre Channel Storage Manager software is not installed first, the LUNs appear as `Other FC SCSI disk` to the operating system when the system is powered up with the 2102 Storage Server. To correct this, the LUNs must be removed from the operating system. After removing the incorrectly installed hard disks and installing the Fibre Channel Storage Manager software, the LUNs appear as `2102-F10 Disk Array Device` after the reboot that follows installation.

If the 2102 Storage Server controller hardware was attached before the software installation, the logical units must be removed from the operating system before continuing with configuration.

Following are the steps for removing the LUNs using SMIT.

1. Enter the following SMIT fast path:
   ```
   smit rmvdsk
   ```
2. Select the disk to be removed. The device type is listed as `Other FC SCSI disk`. Do not remove any devices that should be listed as `Other FC SCSI disk`.
3. Select the **KEEP definition in database** option from the Remove a Disk menu, and change the value to **No**.

The incorrectly installed disk is removed from the system configuration. Repeat the procedure as necessary to remove all incorrectly installed disks.

After removing the incorrectly installed hard disks and installing the RAIDiant Array software, the LUNs appear as `2102-F10 Disk Array Device` after the reboot that follows installation.

Adding a 2102 Fibre Channel RAID Storage Server Controller to the Operating System

Note: This procedure is normally run automatically when the system is restarted during the installation and is not a required step of the installation process.

The following two procedures describe how to add a new Storage Server controller to the system and how to reconfigure a Storage Server controller that had been previously removed from the system.

Adding a New Controller to the System

Run the configuration manager using the **cfgmgr** command to configure a 2102 Storage Server controller. The 2102 Storage Server controller may be configured manually using SMIT with the following steps:

1. Select the **Disk Array Controller Configuration** option from the RAIDiant Disk Array menu.

   Note: For complete instructions on accessing the appropriate SMIT menus, see "Accessing the FC RAIDiant Disk Array Manager Menu" on page 35.
2. Select the **Add a Disk Array Controller** option. The system displays a list of adapters similar to the following example:
   ```
   fscsi1  Available  00-01-01  FC SCSI I/O Controller Protocol Device
   fscsi3  Available  00-03-01  FC SCSI I/O Controller Protocol Device
   ```
3. Select the adapter that will be the parent of the controller you are adding. After you select the adapter, the system displays the Add a Disk Array Controller menu.
4. Specify the **CONNECTION Address** of the controller you wish to add by using the tab key to cycle through the available choices, or press the F4 key to display a list of valid connection address values from which you can select.
5. Continue by pressing the appropriate button or key.

After the controller is added to the system, repeat the steps if additional controllers are to be added.

Reconfiguring a Controller in the Defined State

If a controller is in a DEFINED state, you can reconfigure the controller using SMIT.

To use SMIT, perform the following steps:

1. Select the **Disk Array Controller Configuration** option from the RAIDiant Disk Array menu.

   Note: For complete instructions on accessing the appropriate SMIT menus, see "Accessing the FC RAIDiant Disk Array Manager Menu" on page 35.
2. Select the **Configure a Defined Disk Array Controller** option.

   A panel displays showing you the disk array controllers that are defined, but not yet available to the system.
3. Select the appropriate Storage Server controller. The Storage Server controller status is changed from **Defined** to **Available** and is now available to the system.

Putting Disk Array Controllers into Active Mode

This procedure is to be used at installation of a 2102 Storage Server, at installation of an additional disk array controller, or following replacement of a failed controller.

For normal operation, both disk array controllers (DACs) in the 2102 Storage Servershould be in the active state. This dual-active mode of operation optimizes performance by allowing each controller to handle I/O traffic for the drive groups assigned to it. When a failure occurs on one controller, the remaining controller handles all I/O at a reduced performance level. When the failed controller is replaced, this procedure causes drive groups to be returned to the control of the Storage Server controller to which they were originally assigned.

Run the command:

```
chdev -l dar#
```

where *dar#* is the device name of each disk array router (dar) on the system. The dar device name for a single 2102 Storage Serverattached to a host system is usually `dar0`.

Refer to “Verifying the 2102 Fibre Channel RAID Storage Server Software Installation” on page 35 for instructions on using the SMIT interface to determine the dar device names and to confirm that all dacs are in the active state.

Drive Failures and Arrays

The Fibre Channel Storage Manager handles drive failures differently, depending upon the RAID level involved. For information on how to recover after a drive failure, see “Restoring and Recovering Logical Units” on page 59.

Drive Status

The Fibre Channel Storage Manager displays the following status for drives:

Failed The drive was failed by the array controller or by the user and must be replaced.

Formatting The drive is currently being formatted or is currently reconstructing.

Mismatch The Storage Server controller sensed that the drive has either a sector size, capacity, serial number, SCSI channel, or ID different from what the Storage Servercontroller expected.

Non-Existent No drive is physically connected to the array at this position.

Optimal The drive is functioning normally.

Replaced The drive has just been replaced.

Spare The drive is connected to the Storage Server, but not configured into a logical unit.

Wrong Drive The wrong drive was replaced, and the logical unit is no longer accessible. See “Recovering from a ’Wrong Drive Replaced’ Error” on page 67 to revive the drive.

Drive Failure Types

This section contains information on drive failures, including:
- Drive Failures on RAID Level 0
- Drive Failures on RAID Level 1
- Drive Failures on RAID Level 5

A status code is given in parentheses with the **Failed Drive** message. These codes provide you with additional information about the type of drive failure:

Status	Definitions
03	Failed Drive–Cause Unknown
13	Failed Drive–Read Failure
23	Failed Drive–No Response
33	Failed Drive–Format or Reconstruct Failure
43	Failed Drive–Write Failure
53	User Failed via Mode Select
73	Failed Drive–Controller Storage Failure

Drive Failures on RAID Level 0

Whenever one or more drives in a RAID level 0 LUN has experienced read or write errors, you should attempt to back up the files on all logical volumes and file systems on the LUN. If the error is not serious, you may be able to recover some data. Regardless of whether the backup succeeds, you must use the Problem Determination option in the diagnostics to identify the disk drive on which the error occurred. You must then delete all the logical volumes and file systems on the LUN, remove the LUN from the volume group, fail the drive identified as having the error (using the FC SCSI RAIDiant Disk Array Manager), and replace the newly failed drive. You must then add the LUN back to a volume group, recreate the logical volumes and file systems on it, and copy data back to the restored LUN from your backup media.

Note: Remember that a given drive may contain more than one LUN. If a failed drive does contain more than one LUN, you will have to perform the recovery procedure above on each LUN on the affected drive.

Drive Failures on RAID Level 1

This section describes drive failures on a RAID level 1 LUN, and the LUN and drive status values that result from such failures.

Single-Drive Failures

Whenever a single drive in a RAID level 1 LUN experiences read or write errors, the Storage Server controller changes the status of the drive to **Failed** if the errors are serious (loss of drive power, drive component failure, and so on). The LUN status changes to **Degraded**. You can continue to access the LUN, because the data on the failed drive can be reconstructed from the mirrored pair.

Whenever a drive is marked as **Failed**, you should replace it as soon as possible. The Storage Server controller will then reconstruct the data on the drive automatically. If you cannot replace the drive immediately, back up the files on all logical volumes and file systems on the LUN and replace the drive when you can.

Multiple-Drive Failures

Whenever the second drive in a RAID level 1 LUN experiences read or write errors, the Storage Server controller takes the following actions:

- If the error is a recoverable read error, then the drive status remains **Optimal** and an entry is posted in the error log. You must use the **Problem Determination** option in the diagnostics to identify the disk drive on which the error occurred.
- If the error is an unrecoverable error, and the drive is the mirrored pair of a drive-that is already failed, the drive status changes to **Failed**. The LUN status changes to **Dead**.

Note: Only one failed drive per RAID level 1 LUN is supported on the 2102 Storage Server.

Although a RAID level 1 LUN may have more than one failed drive and still remain degraded as long as none of the failed drives are a mirrored pair, this situation is not supported. For example, in a six-drive RAID level 1 LUN, it is possible for as many as three drives to fail while the LUN remains accessible (in degraded mode). You should not, however continue to operate a RAID level 1 LUN with any number of failed drives for longer than it takes to replace the drives.

Note: Remember that a given drive may contain more than one LUN. If a failed drive does contain more than one LUN, you will have to perform the preceding recovery procedure on each LUN on the affected drive.

Drive Failures on RAID Level 5

This section describes drive failures on a RAID level 5 LUN, and the LUN and drive statuses that result from such failures.

Single-Drive Failure

Whenever a single drive in a RAID level 5 LUN experiences read or write errors, the Storage Server controller changes the status of the drive to **Failed** if the errors are serious (for example, loss of drive power or drive component failure). The LUN status changes to **Degraded**. You can continue to access the LUN because the data on the failed drive can be reconstructed using parity and data on the other drives.

Whenever a drive is marked as Failed, you should replace it as soon as possible. The Storage Server controller will then reconstruct the data on the drive automatically. If you cannot replace the drive immediately, back up the files on all logical volumes and file systems on the LUN and replace it when you can.

Multiple-Drive Failures

If a second drive in a RAID level 5 LUN experiences read or write errors, the Storage Server controller takes the following actions:

- If the error is a recoverable read error, then the drive status remains Optimal and an entry is posted in the error log. You must use the **Problem Determination** option in Diagnostics to identify the disk drive on which the error occurred.
- If the errors are unrecoverable, the drive status changes to **Failed**. The LUN status changes to **Dead**. If the status is **Dead**, all data on the LUN has been lost.

If the LUN status is **Degraded**, you will probably be able to reconstruct the data on the LUN by replacing the **Failed** drive, and then reconstructing the data on the drive. If the LUN status is **Dead**, you will have to replace the failed drives and then delete and recreate the LUN (after deleting the logical volumes and file systems on the LUN and removing the LUN from the volume group).

Note: Remember that a given drive may contain more than one LUN. If a failed drive does contain more than one LUN, you will have to perform the recovery procedure above on each LUN on the affected drive.

Reconstructing RAID Levels after a Drive Failure

Reconstruction is a process used to restore a degraded RAID level 1 or 5 LUN to its original state after a single drive has been replaced. During reconstruction, the Storage Server controller recalculates the data on the drive that was replaced, using data and parity from the other drives in the LUN. The controller then writes this data to the replaced drive. Although RAID level 1 does not have parity, the Storage Server controller can reconstruct data on a RAID level 1 LUN by copying data from the mirrored disk.

Normally, the Storage Server controller automatically initiates the reconstruction process after you replace a **Failed** drive in a degraded RAID level 1 or 5 LUN. In some circumstances, you must manually start reconstruction on the replaced drive. Once reconstruction is initiated (either by you or the Storage Server controller), the Storage Server controller completes the following actions:

- Formats the new drive (if the Storage Server controller determines it is necessary)
- Copies special array software files to the new drive
- Recalculates the data and parity from the data and parity on the other LUN drives
- Writes the recalculated data and parity to the new drive

Reconstruction can take place while the LUN is in use. You do not need to shut the LUN down.

While the FC SCSI RAIDiant Disk Array Manager is formatting the new drive, there is no additional effect on system I/O performance (the LUN continues to operate in degraded mode). Once data reconstruction begins, you can adjust the rate of reconstruction so that it does not interfere significantly with system performance.

If there are multiple-drive failures in a RAID level 1 or 5 LUN (at least two **Failed** drives), you will not be able to reconstruct the drive data. To restore a RAID level 1 or 5 LUN to a working state in this case, you need to replace the drives and reformat the LUN (after deleting the logical volumes and file systems on the LUN and removing the LUN from the volume group).

The Reconstruction Rate

The rate of reconstruction depends on two parameters, the delay interval and the blocks-per-delay interval.

The delay interval is the time, in tenths of a second, between reconstruction operations. When the Storage Server controller reconstructs data on a drive, it divides its time between reconstruction and normal I/O operations. System I/O operations take place during this delay interval. As the delay interval increases, system I/O performance also increases, but so does reconstruction time.

The blocks-per-delay interval is the number of blocks the Storage Servercontroller reconstructs during each reconstruction operation. The parameter can be any number between 1 and 65,535. The more blocks there are, the longer it will take to reconstruct them. This is time that cannot be used to perform system I/O. Therefore, the larger the reconstruction amount is, the more the system performance is degraded.

The delay interval parameter and the blocks-per-delay interval parameter interact to determine the overall rate of data reconstruction. The reconstruction rate is also linked to the amount of I/O on the logical unit while it is being reconstructed. The faster the rate of reconstruction is, the less I/O the logical unit can handle; but the faster the reconstruction (and the I/O degradation) will be completed.

These parameters control only the rate of data reconstruction, not total reconstruction time. The first step in reconstruction, formatting the replaced drive, can take a long time (an hour or more for some drives). During this time, however, you can continue to access the array with no decrease in current system performance (the array remains in the degraded mode).

Table 5 gives some suggested rates and sample timings:

Table 5. Reconstruction Rates

Rate (relative)	Reconstruction Frequency (in tenths of a second)	Reconstruction Amount (in blocks)	Time for Completion (in minutes/GB)*
Slow	6	250	20
Moderate	1	256	7
Fast	1	1024	3

Note: *Timings are approximate and apply only to data reconstruction; formatting time is not included. Times will be slower if other processes are accessing the array during reconstruction.

In general, reconstruction times tend to level off after block sizes of 20 KB. Although higher block sizes (up to 64 KB) are possible, there is no great reduction in time at the higher levels. Moreover, higher numbers may cause system response time to be significantly lower when data on the reconstructing logical unit is being accessed.

Changing Reconstruction Rate Parameters

You can use the SMIT procedure to change the Reconstruction Rate parameters.

Note: You cannot change the other parameters using this procedure. To change these parameters, see “Changing the RAID Level, Drive Selection, or Size of Array Parameters” on page 38.

Perform the following to use SMIT:

1. Select the **RAIDiant Disk Array Manager** option from the RAIDiant Disk Array menu.
2. Select the **Change/Show SCSI RAID Array** option.
3. When prompted, select a disk array.
4. When the Change/Show SCSI RAID Array menu is displayed, select the **Reconstruction Rate Delay Interval** option, and enter the new value for this field.
5. Select the Blocks Per Reconstruction Rate Delay Interval option and enter the new value for this field.
6. Press Enter to make the changes, then exit the FC SCSI RAIDiant Disk Array Manager.

Checking and Repairing Array Parity

Note: The **Parity Check/Repair** option applies only to RAID levels 1 and 5. RAID level 0 does not have array parity, and therefore can not be checked and repaired. RAID level 1 does not really have parity either, but a parity check compares data on the mirrored drives. In addition, the RAID level 1 or 5 logical unit must be at **Optimal** status in order to run **Parity Check/Repair.**

Parity Check/Repair performs the following functions:

- Scans the logical unit and checks the array parity for each block in the logical unit. See "RAID Levels" on page 5 for a description of array parity. On a RAID level 1 logical unit, Parity Check compares the data on each mirrored pair, block by block.
- Repairs any array parity errors found during the parity check. On a RAID level 1 logical unit, the array controller changes the data on the mirror disk to make it match the data on the data disk. On a RAID level 5 logical unit, the controller changes the parity so that it is consistent with the data.

If array parity errors result from corrupted data, the data is not repaired, only the array parity. Moreover, you may still lose some data as a result of a power failure or abnormal shutdown, especially if you do not have an uninterruptible power supply (UPS). Data cached in buffers will be lost and cannot be reconstructed if you do not have a UPS. Therefore, you should always maintain backup files, even with a redundant array.

Array parity is used in RAID 5 logical units to enable data to be reconstructed if a single drive fails. Checking and repairing parity helps ensure that you will be able to recover after a drive failure. You should check and repair parity after an abnormal system shutdown.

Run **Parity Check/Repair** manually after an abnormal system or array shutdown. As the result of such a shutdown, required array parity may not have been updated, resulting in potential data corruption.

Running Manual Parity Check/Repair

You should run **Parity Check/Repair** manually after an abnormal shutdown (for example, a power failure or system crash). You may still lose some data as a result of the power failure or abnormal shutdown, especially if you do not have a UPS.

Data cached in buffers will be lost and cannot be reconstructed if you do not have a UPS. This is one of the reasons you should always maintain backup files, even with RAID level 1 or 5 logical units.

To run **Parity Check/Repair** in SMIT, perform the following steps:

1. Start the RAIDiant Disk Array Manager from SMIT by entering the following fast path:

   ```
   smit fraidmgr
   ```

2. Select the **Parity Check/Repair** option in the RAIDiant Disk Array menu.
3. Select the device (hdisk) name of the logical unit you want to check and repair.

   Note: You must run Parity Check/Repair on each logical unit separately.
4. When all the entries are correct, press Enter to perform the check on the selected logical units.
5. When the parity check and repair is complete, a parity report is written to the report file **/tmp/parity.chk.**
6. Repeat step 3 through step 5 as needed to check and repair additional logical units.

Restoring and Recovering Logical Units

Logical units can be restored from a working state drive failure using several methods. Procedures for logical unit restoration are provided for each of the four RAID levels, because each level has a unique recovery process. To restore a logical unit, refer to “Restoring a RAID Logical Unit” on page 62.

Procedures are also provided for various logical unit status values within RAID levels 1 and 5. Refer to the following sections:

- “Replacing Failed Drives (RAID Levels 1 and 5)” on page 63
- “Replacing Optimal Drives (RAID Levels 1 and 5)” on page 64
- “Replacing Both Failed and Optimal Drives” on page 65
- “Replacing a Dead LUN” on page 66
- “Recovering from a ’Wrong Drive Replaced’ Error” on page 67

Restoring a logical unit is the process of returning the logical unit to a working state after one or more drive failures. The tasks involved depend on the RAID level of the logical unit and on how many drives have failed or have experienced errors:

- Single-drive failures in RAID levels 1 and 5 are easily restored without losing data and without shutting down the logical unit.
- Multiple-drive failures, or single-drive failures on RAID level 0, are treated like non-array drive failures in AIX. This involves deleting the logical volumes, file systems, and volume groups on the logical unit, as well as replacing the drive, reformatting the logical unit, and recreating the AIX file structures.

  Note: Single-drive failures on redundant RAID levels (1 and 5) are easy to recover from. Identify the failed drive, physically remove it from the subsystem, and replace it with another drive. The subsystem then starts reconstructing the failed drive. This task is intended to be performed by the system administrator and involves no reorganization or deletion of file systems or volume groups.

You need to replace a drive when one of the following events occurs:

– The AIX operating system sends a message indicating a logical unit failed or is no longer accessible.
– You see a logical unit status other than **Optimal** displayed by the **List all SCSI RAID Arrays** option.
– You see a drive status other than **Optimal** displayed by the **List all SCSI Raid Arrays** option.

The recovery tasks are summarized in the RAID Recovery Tasks Summary tables that follow.

Table 6. RAID Level 0 Recovery Tasks Summary

Logical Unit and Drive Status	Recovery Steps
The LUN status is **Optimal**. The problem determination procedure in the diagnostics indicates that a drive should be replaced.	1. Stop using all file systems and logical volumes on the affected logical units immediately. Attempt to back up all files in those logical volumes. 2. Delete the file systems and logical volumes on the affected logical units and remove the logical units from their volume groups. (See "Deleting Logical Units from the Operating System" on page 47.) 3. Fail all drives (using the FCRDAM) that were called out by the diagnostics, replace the failed drives, and reformat the affected logical units. 4. Create new logical volumes and file systems on the restored units. (See "Adding Logical Units to the Operating System" on page 46.) 5. Copy backed-up data to the restored logical volumes or file systems.
The LUN status is **Dead**. All failed drives failed due to a single channel failure event.	See "Reviving a LUN That Failed Due to a Channel Failure" on page 67.

Note: See "Restoring a RAID Logical Unit" on page 62 for more detailed information.

Table 7. RAID Level 1 Recovery Tasks Summary

Logical Unit and Drive Status	Recovery Steps
The LUN status is **Optimal**. The problem determination procedure in the diagnostics indicates that a drive should be replaced.	1. If there are two or more drives within a LUN that were called by the diagnostics, stop using all file systems and logical volumes on the affected logical units immediately, and attempt to back up all files in those logical volumes before continuing. 2. One at a time, fail each drive (using the FCRDAM) that was called out by the diagnostics, replace the failed drive, and reconstruct the data on it.
The LUN status is **Degraded**. The drive status shows one or more **Failed** drives. **Note:** Only one failed drive per RAID level 1 LUN is supported on the 2102 Storage Server.	1. Replace the failed drives (all drives can be replaced at the same time). 2. One at a time, reconstruct the data on the replaced drives by using the FCRDAM(if the array does not auto-reconstruct).

Table 7. RAID Level 1 Recovery Tasks Summary (continued)

Logical Unit and Drive Status	Recovery Steps
The LUN status is **Degraded**. The drive status shows one **Failed** and the problem determination procedure in the diagnostics indicates that a drive other than the **Failed** drive should be replaced.	1. Stop using all file systems and logical volumes on the affected logical units immediately, and back up all files in those logical volumes before continuing. If the backup is unsuccessful, delete the file systems and logical volumes on the affected logical units and remove the logical units from their volume groups. (See "Deleting Logical Units from the Operating System" on page 47.) 2. Replace and reconstruct the failed drive by using the FCRDAM. 3. After the failed drive has been reconstructed, then one at a time fail, replace, and reconstruct each drive (by using the FCRDAM) that was called out by the diagnostics.
The LUN status is **Dead** The drive status shows two or more **Failed** drives.	1. Delete the file systems and logical volumes on the affected logical units, and remove the logical units from their volume groups. (See "Deleting Logical Units from the Operating System" on page 47.) 2. Fail all drives (using the FCRDAM) that were called out by the diagnostics, replace the failed drives, and reformat the logical units. 3. Create new logical volumes and file systems on the restored units. (See "Adding Logical Units to the Operating System" on page 46.) 4. Copy backed-up data to the restored file systems or logical volumes.
The LUN status is **Dead**. All failed drives failed due to a single channel failure event.	See "Reviving a LUN That Failed Due to a Channel Failure" on page 67.

Note: See "Restoring a RAID Logical Unit" on page 62.

Table 8. RAID Level 5 Recovery Tasks Summary

Logical Unit and Drive Status	Recovery Steps
The LUN status is **Optimal**. The problem determination procedure in the diagnostics indicates that a drive should be replaced.	1. If there are two or more drives within a LUN that were called out by the diagnostics, stop using all file systems and logical volumes on the affected logical units immediately and back up all files in those logical units before continuing. 2. One at a time, fail each drive using the FCRDAM, replace the failed drive, and reconstruct the data on it.
The LUN status is **Degraded**. The drive shows one **Failed** drive.	1. Replace the failed drive. 2. Reconstruct the data on the replaced drive by using the FCRDAM.

Table 8. RAID Level 5 Recovery Tasks Summary (continued)

Logical Unit and Drive Status	Recovery Steps
The LUN status is **Degraded**. The drive status shows one **Failed** drive and the problem determination procedure in the diagnostics indicates that a drive other than the **Failed** drive should be replaced.	1. Stop using all file systems and logical volumes on the affected logical units immediately, and back up all files in those logical volumes before continuing. If the backup is unsuccessful, delete the file systems and logical volumes on the affected logical units and remove the logical units from their volume groups. (See "Deleting Logical Units from the Operating System" on page 47.) 2. Replace and reconstruct the failed drive by using the FCRDAM. 3. After the failed drive has been reconstructed, then one at a time fail, replace, and reconstruct each drive (by using the FCRDAM) that was called out by the diagnostics.
The LUN status is **Dead**. The drive status shows two or more **Failed** drives.	1. Delete the file systems and logical volumes on the affected logical units, and remove the logical units from their volume groups. (See "Deleting Logical Units from the Operating System" on page 47.) 2. Fail all drives (using the FCRDAM) that were called out by the diagnostics, replace the failed drives, and reformat the logical units. 3. Create new logical volumes and file systems on the restored units. (See "Adding Logical Units to the Operating System" on page 46.) 4. Copy backed-up data to the restored file systems or logical volumes.
The LUN status is **Dead**. All failed drives failed due to a single channel failure event.	See "Reviving a LUN That Failed Due to a Channel Failure" on page 67.

Note: See "Restoring a RAID Logical Unit" for more information.

REQTEXT

Restoring a RAID Logical Unit

Use the following procedure to restore a RAID logical unit:

1. Determine the degree to which the drive fault has affected the 2102 Storage Server. This can be accomplished by first running the **Problem Determination** option in diagnostics. This will report any drives that have had errors. See the *Diagnostic Information for Multiple Bus Systems* manual for complete instructions on running system diagnostics.
2. Using SMIT, display the status for each LUN and the status of each drive within the LUN.
   - Enter the following fast path:

     ```
     smit fraidmgr
     ```
   - Select the **RAIDiant Disk Array Manager** option, then select the **List all SCSI RAID ARRAY** option.

3. If a LUN status is **Optimal**, yet diagnostics has called out one of the drives that it contains, then that drive has had one or more soft errors, or it is a RAID level 0 LUN. Go to “Replacing Optimal Drives (RAID Levels 1 and 5)” on page 64 and complete the procedure given there.
4. If a LUN status is **Degraded**, and diagnostics only calls out the same drive that the **List all SCSI RAID ARRAY** option shows as **Failed**, then only one drive in the LUN has had an error. Go to “Replacing Failed Drives (RAID Levels 1 and 5)” and complete the procedure given there.
5. If the LUN status is **Degraded**, and diagnostics calls out a drive within the **Degraded** LUN, yet different from the drive that the **List all SCSI RAID ARRAY** option shows as **Failed**, then the LUN has suffered both an unrecoverable error and a recoverable error. Depending upon the circumstances, it is possible that data has been lost. The steps below should be used to recover from this scenario.
   a. Stop using all file systems and logical volumes on the affected logical units immediately, and back up all files in those logical volumes before continuing.
   b. Go to “Replacing Failed Drives (RAID Levels 1 and 5)” and complete the procedure given there.
   c. Go to “Replacing Optimal Drives (RAID Levels 1 and 5)” on page 64 and complete the procedure given there for each drive that was reported by Diagnostics.
   d. If the Reconstructions do not work, go to step 6.
6. If the LUN status is **Dead**, then go to “Replacing a Dead LUN” on page 66.

Replacing Failed Drives (RAID Levels 1 and 5)

Use the following procedure to locate and replace one or more failed drives in a RAID level 1 or 5 logical unit. The logical unit will have a status of **Degraded** and at least one drive with a status of **Failed**.

1. Perform the following in SMIT:
   a. Enter the following fast path:
      ```
      smit fraidmgr
      ```
   b. Select the **RAIDiant Disk Array Manager** option, then select the **Change/Show Drive Status** option.
2. Run hardware diagnostics in concurrent mode on the system. See the *Diagnostic Information for Multiple Bus Systems* manual for complete instructions on running system diagnostics.
3. After running hardware diagnostics, an error code should have been generated by the diagnostics program.
4. After you have physically replaced the drive, the subsystem should automatically start reconstruction on that drive. The status of the replaced drive should change to **Replaced**. If you replaced more than one drive, only one will start reconstructing.

   Attention: The 2102 Storage Server controller will only reconstruct a single drive at a time. If more than one drive is replaced, then the drive status will change to **Replaced**, and the LUN status will remain as **Degraded**. When the first reconstruction completes, the controller will start the reconstruction process for the second drive. It will then change the LUN status to **Reconstructing**.
5. If reconstruction does not start automatically, for SMIT, select the **Replace and Reconstruct** option from the **Drive Status field** of the Change/Show Drive Status menu to start reconstruction manually.

If more than one logical unit is on the replaced drive, each unit will be reconstructed in turn. The status of each affected logical unit changes as the unit is reconstructed.

You can change the reconstruction rate to speed up and slow down reconstruction, depending on the level of system I/O. See “The Reconstruction Rate” on page 57 for more information on the reconstruction rate and “Changing Reconstruction Rate Parameters” on page 57 for information on how to change the rate.

Wait for reconstruction to finish on all affected logical units. The drive status of the replaced drive changes to **Optimal**, and the statuses of all logical units in the drive change to **Optimal** as well.

6. If you have to replace more than one drive (RAID level 1 only), you must now select and start reconstruction manually for each one in turn. Reconstruction does not start automatically for the second and subsequent drives.
7. After you have replaced and reconstructed all the failed drives, exit the RAIDiant Disk Array Manager.

Replacing Optimal Drives (RAID Levels 1 and 5)

Attention: Do not use this procedure to replace drives in a RAID level 0 array. For that procedure, see “Replacing a Dead LUN” on page 66. If there are two or more drives that need to be replaced, stop using all logical volumes and file systems on the logical unit immediately. Back up all logical volumes on the logical unit before continuing. If there is more than one logical unit on the affected drives, stop using the logical volumes and file systems on them as well.

Continuing to use the logical unit with two drives with recovered errors could result in the failure of both drives and the loss of all data. You may resume using the logical unit after you have replaced and reconstructed all the drives but one.

1. Refer to the list of drives that Diagnostics called out, yet were not marked as **Failed** drives.
2. Use the following SMIT procedure to locate and fail one of the **Optimal** drives. Multiple failed drives are possible only in RAID level 1.

   Attention: Do not fail more than one drive at a time, even if you need to replace more than one drive. Fail and reconstruct the drives one at a time. After you fail a drive, physically replace it. Replacing all the drives at once will result in the loss of all data on the logical unit.

   In SMIT, perform the following:

   a. Enter the SMIT fast path `smit fraidmgr`.
   b. Select the **RAIDiant Disk Array Manager** option.
   c. Select the **Change/Show Drive Status** option.
   d. Locate the Optimal drive you want to fail.
   e. Select the **Fail Drive** option from the **Drive Status** field. See “Failing a Drive” on page 44 if you need more information.
3. Run hardware diagnostics on the system. See the *Diagnostic Information for Multiple Bus Systems* manual for complete instructions on running system diagnostics.
4. After running hardware diagnostics, an error code should have been generated by the diagnostics program.

5. After you have physically replaced the drive, the subsystem should automatically start reconstruction on that drive. The status of the replaced drive should change to **Replaced**. If you replaced more than one drive, only one will start reconstructing.

   Attention: The 2102 Storage Server controller will only reconstruct a single drive at a time. If more than one drive is replaced, then the drive status will change to **Replaced**, and the LUN status will remain as **Degraded**. When the first reconstruction completes, the controller will start the reconstruction process for the second drive. It will then change the LUN status to **Reconstructing**.
6. If reconstruction does not start automatically, perform the appropriate action depending on which method you choose.

   For SMIT, select the **Replace and Reconstruct** option from the Change/Show Drive Status menu to start reconstruction manually.

   You can change the reconstruction rate to speed up and slow down reconstruction, depending on the level of system I/O. See “The Reconstruction Rate” on page 57 for more information on the reconstruction rate and “Changing Reconstruction Rate Parameters” on page 57 for information on changing the rate.

   Wait for reconstruction to finish on all affected logical units. The drive status of the replaced drive changes to **Optimal**, and the status of all logical units in the drive changes to **Optimal** as well.
7. If there are more drives that need to be replaced, go back to step 2 on page 64 and repeat this procedure for each drive.
8. After you have replaced and reconstructed all drives, exit the Fibre Channel Storage Manager.

Replacing Both Failed and Optimal Drives

Use the following procedure to replace one or more failed and one or more **Optimal** drives that was called out by the diagnostics in a RAID level 1 or 5 logical unit. The logical unit will have a status of **Degraded** and one drive with a status of **Failed** and one or more drives that have been called out by the diagnostics and have a status of **Optimal**.

Attention: On a RAID level 1 logical unit, this procedure only works if you were able to successfully back up all logical volumes and file systems on the logical unit (indicating that there has not been a failure of a mirrored pair of drives). If the backup failed on at least one logical volume or file system, use the procedure in “Replacing a Dead LUN” on page 66. Stop using all logical volumes and file systems on the logical unit immediately, if they are still accessible. If there is more than one logical unit on the affected drives, stop using the logical volumes and file systems on them, as well. Continuing to use the logical unit in a **Degraded** state could result in the failure of the **Optimal** drive and the loss of all data. You may resume using the logical unit after you have replaced and reconstructed all the drives but one.

1. Go to “Replacing Failed Drives (RAID Levels 1 and 5)” on page 63 and replace the failed drives one by one.
2. After the failed drives have been reconstructed, go to “Replacing Optimal Drives (RAID Levels 1 and 5)” on page 64 and replace all the Optimal drives one by one.

Replacing a Dead LUN

This procedure is essentially the same procedure you would use to replace a failed non-array disk drive in the AIX operating system. Use this procedure to restore logical units in the following circumstances:

- All RAID level 0 logical units.
- A RAID level 1 logical unit with a logical unit status of **Degraded** and drive status of **Failed**. Steps in this procedure direct you to another recovery procedure if you are able to back up the files on the RAID level 1 logical unit.
- A RAID level 1 logical unit with a status of **Dead**.
- A RAID level 5 logical unit with a status of **Dead**.

Attention: Stop using the logical volumes and file systems on the affected logical unit immediately, if they are still accessible. If there is more than one logical unit on the affected drives, stop using the logical volumes and file systems on them as well. Continuing to use the logical volumes and file systems on a logical unit having drives that were called out by the diagnostics could result in the complete failure of the drives and the loss of all data.

1. For all RAID levels, attempt to back up the files in all the logical volumes and file systems on all affected logical units.

   If this is a RAID level 1 or 5 logical unit and all the backups were successful, go to “Replacing Both Failed and Optimal Drives” on page 65 and complete the procedure given there.
2. Remove all logical volumes and file systems on all logical units in the drive group affected by the bad drives. See “Deleting Logical Units from the Operating System” on page 47.
3. Use the following SMIT procedure to locate and fail one of the drives. Multiple failed drives are possible only in RAID Level 1.
   a. Enter the SMIT fast path `smit fraidmgr`
   b. Select the **RAIDiant Disk Array Manager** option.
   c. Select the **Change/Show Drive Status** option.
   d. Locate the drive you want to fail.
   e. Select the **Fail Drive** option from the Drive Status field. See “Failing a Drive” on page 44 if you need more information.
4. After you fail the drives, run concurrent hardware diagnostics on the system. See the *Diagnostic Information for Multiple Bus Systems* for complete instructions on running system diagnostics.
5. After running hardware diagnostics, an error code should have been generated by the diagnostics program. to complete the hardware diagnosis of the failing drive.
6. After you have replaced the failed drives, choose the appropriate action to begin reformatting the repaired unit. In SMIT:
   a. Select the **Change/Show Drive Status** from the RAIDiant Disk Array Manager menu.
   b. Select the **Failed** drive and then select the **Add New Drive option** from the **Drive Status field**.
7. Repeat step 6 for each **Failed** drive in the LUN.
8. Create new logical volumes and file systems on the restored logical units. (See “Adding Logical Units to the Operating System” on page 46.)

9. Copy backed-up data back to the restored logical volumes and file systems.

Recovering from a 'Wrong Drive Replaced' Error

If you mistakenly remove a drive in an Optimal state from a degraded logical unit (either a RAID level 1 or 5), the array controller will mark the logical unit Dead, and mark the drive as **Wrong Drive**.

To recover from a 'wrong drive replaced' error, you must revive the drive by completing the steps given in the following procedure:

Attention: Never revive a drive as a replacement for reconstructing the drive data. Reviving a drive only changes a **Failed** status to **Optimal**, and does not rebuild the data and parity.

1. Reinstall the good drive (which still has the data on it) in its original drive slot.
2. Start the RAIDiant Disk Array Manager by entering the following SMIT fast path:
```
smit fraidmgr
```
3. Select the **Change/Show SCSI RAID Arrays** option from the RAIDiant Disk Array Manager menu. A display similar to the following appears:
```
hdisk4  Raid 0 00-02-00-1,4     1 MB Status Optimal
    hdisk4  32 Channel 3 ID 2   Location LR-1  Optimal
hdisk0  Raid 5 00-02-00-0, 0  3835 MB Status Dead
    hdisk0  11 Channel 1 ID 1   Location LR-4  Optimal
    hdisk0  21 Channel 2 ID 1   Location LR-7  Wrong Drive Replace
    hdisk0  31 Channel 3 ID 1   Location LF-2  Failed Drive
hdisk1  Raid 1 00-02-00-0, 1  1917 MB Status Optimal
    hdisk1  20 Channel 2 ID 0   Location LR-6  Optimal
    hdisk1  50 Channel 5 ID 0   Location LF-3  Optimal
hdisk2  Raid 1 00-02-00-1, 2  1917 MB Status Optimal
    hdisk2  30 Channel 3 ID 0   Location LF-1  Optimal
    hdisk2  40 Channel 4 ID 0   Location LF-6  Optimal
```
4. Identify the line describing the **hdisk** affected by the drive replacement error. The drive status will be indicated as **Wrong Drive Replaced**.
5. Select the **Change/Show Drive Status** option from the RAIDiant Disk Array Manager menu.
6. Select the **Wrong Drive Replace** drive.
7. Select the **Add New Drive** option from the **Drive Status** field of the Change/Show Drive Status to revive the drive.
8. Check the drive status of the drive that you just revived. The drive status should now be **Optimal**.
9. Go back to the reconstruction procedure you were using and complete it.

Reviving a LUN That Failed Due to a Channel Failure

Attention: Do not use this procedure to revive a LUN that has failed for any reason other than a channel failure. In any other case, the failed drives will not be consistent, and there is an extremely high risk of data corruption.

Use this procedure to return a LUN to the Optimal status after a channel failure caused multiple drives to fail and put a LUN into a Dead status.

Notes:

1. This procedure cannot be performed from SMIT; it must be performed from the command line.

2. This procedure is subject to change in later releases of the Fibre Channel Storage Manager for AIX.

Enter the following command from the command line to add a drive to the drive group:

```
fraidmgr -V -l hdisk#
```

Chapter 6. Diagnostic Aids

This chapter provides guidelines for running diagnostic aids for the 2102 Fibre Channel RAID Storage Server and detailed steps to run each diagnostic service aid.

Diagnostic Applications

Attention: Do not run diagnostics on the 2102 Fibre Channel RAID Storage Server while running the array configuration utility. Otherwise, unexpected results may occur.

Diagnostics for the 2102 Storage Server are available through the standard diagnostics interface. To start diagnostics, enter **diag** and follow the instructions displayed on the screen.

Notes:

1. Diagnostics for the 2102 Storage Server can be run concurrently with the customer's normal operations. Unless you need to stop the using system for some other reason, do not stop the using system to run diagnostics on the 2102 Storage Server.
2. Only error log analysis can be run on a logical disk drive (hdisk*n* where *n* is the number of the LUN).
3. Running diagnostics on the disk array controller (DAC) causes diagnostics to be run on the attached disk drives.
4. The diagnostics can run only on an active DAC. To find out whether a DAC is active, enter the following AIX command:

   ```
   lsattr -El darn
   ```

   where *n* is the number of the disk array router (dar); for example, dar0.

If diagnostics indicate a problem with the RAID subsystem, refer to the *FCRDAM and Expandable Storage Unit, Service Guide, IBM 2102 Model F10 and Model D00* for information about service procedures.

Diagnostic Service Aids

This section identifies the steps to take to run the diagnostic service aids for the 2102 Fibre Channel RAID Storage Server. The names of the diagnostic service aid tasks are:

- Display Microcode Level
- Certify LUN
- Certify Spare Physical Disk
- Download Microcode to Array Controller
- Download Microcode to Physical Disk
- Format Failed Physical Disk
- Update EEPROM
- Replace Controller

The following sections identify the specific steps required to perform each of these tasks. For additional information about diagnostics, refer to the *Diagnostic Information for Multiple Bus Systems* book.

Displaying the Microcode Level

To display the current level of controller microcode, perform the following steps:

1. Enter **diag** to start the diagnostics. The Operating Instructions screen appears.
2. Press Enter to continue. The Function Selection menu appears.
3. Select the Task Selection option. The Tasks Selection List appears.
4. Select the Display Microcode Level option. The Resource Selection List appears.
5. Select the controller for which you want to display the microcode level.

Certifying a LUN

The Certify LUN service aid checks a logical disk for defective sectors. Before running the Certify LUN service aid, ensure that diagnostics can be run in the system verification mode on the array controller without errors. (Refer to the *Diagnostic Information for Multiple Bus Systems* book.)

To check a logical disk for defective sectors, perform the following steps:

1. Enter **diag** to start the diagnostics. The Operating Instructions screen appears.
2. Press Enter to continue. The Function Selection menu appears.
3. Select the Task Selection option. The Tasks Selection List appears.
4. Select the Certify Media option or the Fibre Channel RAID Tasks option. The Resource Selection List appears.
5. Select the controller.
6. Select the Certify LUN option.
7. Follow the instructions displayed on the screen.

The time it takes to complete the certification varies depending on the size of the LUN.

Certifying a Spare Physical Disk

The Certify Spare Physical Disk service aid checks the disk for defective sectors. Before running the Certify Spare Physical Disk service aid, ensure that diagnostics can be run without errors in the system verification mode on the array controller. (Refer to the *Diagnostic Information for Multiple Bus Systems* book.) To check a spare disk for defective sectors, perform the following steps:

1. Enter **diag** to start the diagnostics. The Operating Instructions screen appears.
2. Press Enter to continue. The Function Selection menu appears.
3. Select the Task Selection option. The Tasks Selection List appears.
4. Select the Certify Media option or the Fibre Channel RAID Tasks option. The Resource Selection List appears.
5. Select the controller.
6. Select the Certify Spare Physical Disk option.
7. Follow the instructions displayed on the screen.

The time it takes to complete the certification varies depending on the size of the physical disk.

Downloading Microcode to Array an Controller

Attention: Do not run this service aid while I/O operations are running to the devices attached to the controller. Otherwise, unpredictable results may occur.

You can only download microcode to an active disk array controller (DAC). To find out whether a DAC is active, enter the following AIX command:

```
lsattr -El darn
```

where *n* is the number of the disk array router (dar); for example, dar0.

Use the system management interface tool (SMIT) to switch a DAC to active. See “Putting Disk Array Controllers into Active Mode” on page 53 for the instructions.

You may download the microcode from a diskette or from a hard disk. Ensure that the microcode is in the directory /etc/array/fw. If you are downloading from a diskette, ensure that the microcode file is in the backup format.

To download microcode to the array controller, perform the following steps:

1. Enter **diag** to start the diagnostics. The Operating Instructions screen appears.
2. Press Enter to continue. The Function Selection menu appears.
3. Select the Task Selection option. The Tasks Selection List appears.
4. Select the Download Microcode option or the Fibre Channel RAID Tasks option. The Resource Selection List appears.
5. Select the controller.
6. Select the Download Microcode to Array Controller option.
7. Follow the instructions displayed on the screen.

Downloading Microcode to a Physical Disk

You may download the microcode from a diskette or from a hard disk. Ensure that the microcode is in the directory /etc/array/fw. If you are downloading from a diskette, ensure that the microcode file is in the backup format.

To update microcode on spare physical disks, perform the following steps:

1. Backup the data from the affected disk drive modules. (For general information about backing up data, refer to the *AIX 4.3 System Management Guide: Operating System and Devices* book.)
2. Enter **diag** to start the diagnostics. The Operating Instructions screen appears.
3. Press Enter to continue. The Function Selection menu appears.
4. Select the Task Selection option. The Tasks Selection List appears.
5. Select the Download Microcode option or the Fibre Channel RAID Tasks option. The Resource Selection List appears.
6. Select the controller.
7. Select the Download Microcode to Physical Disk option.
8. Follow the instructions displayed on the screen.

Formatting a Failed Physical Disk

Attention: The Format Failed Physical Disk service aid destroys all data on the disk. Formatting the wrong disk drive might destroy valuable data.

Use this task to format a failed physical disk. If the disk drive to be formatted is not already in the failed state, put it into this state by using the Fibre Channel Storage Manager for AIX. See “Failing a Drive” on page 44 for the procedure.

To format a failed physical disk, perform the following steps:

1. Backup the data from the affected disk drive modules. (For general information about backing up data, refer to the *System Management Guide: Operating System and Devices* book.)
2. Enter **diag** to start the diagnostics. The Operating Instructions screen appears.
3. Press Enter to continue. The Function Selection menu appears.
4. Select the Task Selection option. The Tasks Selection List appears.
5. Select the Format Media option or the Fibre Channel RAID Tasks option. The Resource Selection List appears.
6. Select the controller.
7. If you selected the Fibre Channel RAID Tasks option, select the Format Failed Physical Disk option from the list.
8. Follow the instructions displayed on the screen.

If you are not sure about the location codes for the disk drive modules that you want to format, see *Fibre Channel RAID Storage Server and Expandable Storage Unit, Service Guide, IBM 2102 Model F10 and Model D00* for an explanation of the codes.

The time it takes to complete the formatting varies depending on the size of the physical disk. If the format operation completes successfully, run the Certify Spare Physical Disk task on the disk drive before using it. If the format operation does not complete successfully, follow the instructions displayed on the screen.

Updating the EEPROM

Use this task to update the array controller’s user configuration NVSRAM region. The EEPROM data must be in the directory /etc/array/fw.

To update the EEPROM, perform the following steps:

1. Enter **diag** to start the diagnostics. The Operating Instructions screen appears.
2. Press Enter to continue. The Function Selection menu appears.
3. Select the Task Selection option. The Tasks Selection List appears.
4. Select the Fibre Channel RAID Tasks option. The Resource Selection List appears.
5. Select the controller.
6. Select the Update EEPROM option.
7. Follow the instructions displayed on the screen.

Replacing the Controller

Use this task to put an array controller into the held-in-reset state.

Attention: You must use this task whenever you are going to remove an array controller that is not in the failed/held-in-reset state from the 2102 Storage Server. Otherwise, problems might occur in the 2102 Storage Server when the controller is removed and replaced.

To put the array controller into the held-in-reset state and to replace it, perform the following steps:

1. Enter **diag** to start the diagnostics. The Operating Instructions screen appears.
2. Press Enter to continue. The Function Selection menu appears.
3. Select the Task Selection option. The Tasks Selection List appears.
4. Select the Fibre Channel RAID Tasks option. The Resource Selection List appears.
5. Select the controller.
6. Select the Replace Controller option.
7. Follow the instructions displayed on the screen.

Note: Status and error messages may have been logged on the host during the controller replacement procedure because the host continues to poll for the controller. If advanced diagnostics in system verification mode run without error and there are no fault lights active after the replacement procedure, you can disregard these messages.

Appendix A. Application Programming Interface (API)

This programming information includes an overview of the software components required to support the 2102 Fibre Channel RAID Storage Server, followed by detailed information on these components:

- Device drivers
  - Programming interfaces to the device drivers
  - I/O control (ioctl) operations that the device drivers support
  - Errors returned by the device drivers
- Special files
- SCSI commands supported by the 2102 Fibre Channel RAID Storage Server

API Overview

In the standard AIX SCSI device support model, the host system communicates with the devices connected to the SCSI bus through a SCSI adapter. To provide access to the SCSI adapters connected to the system, the operating system creates a special file for each adapter, such as: `fscsi0`, `fscsi1`, `fscsi2`.

The AIX operating system divides device driver support for SCSI devices into two layers. The top layer, called the SCSI device driver, is responsible for handling the requirements of a particular type of SCSI device, such as fixed disk and CD-ROM. The top-layer device driver queues SCSI requests, building the SCSI command block, specifying the address of the target SCSI device, and handling errors. For example, the **scdisk** SCSI device driver is the top-layer SCSI device driver for fixed disks.

The bottom-layer, called the SCSI adapter device driver, is responsible for managing the connection with the SCSI bus. This layer arbitrates to obtain control of the bus and sends the SCSI command out onto the SCSI bus. The various top layer SCSI device drivers communicate with the bottom layer SCSI adapter driver by filling a **sc_buf** structure with all the SCSI request information and passing the structure as an argument to the **devstrat** kernel service. The kernel routes the request to the SCSI adapter through which the target SCSI device can be reached.

To support the definition of multiple paths between a host and a 2102 Storage Server, a layer has been added between the existing layers of the AIX SCSI device support model. This middle layer, a pseudo-device called the disk array router (DAR), is responsible for managing the multiple paths that are defined between the host and the device and managing a switch over between the paths in the event a component failure makes a path unusable. To provide access to the pseudo-device, the operating system creates a special file name, such as `dar0`, `dar1`, `dar2`. The system supports the DAR pseudo-device with the **fcparray** device driver.

As in the existing AIX SCSI support model, the top-layer device driver, called the **fcparray** SCSI device driver for the 2102 Storage Server, passes the I/O command down the currently active path described by the **router_info**, **controller_info**, and **adapter_info** structures. The **arrayinfo** structure associated with an hdisk uses its pointer to a **router_info** structure to determine the current path and the state of the path. During normal, error-free operation, ″walking″ the **hdisk** path generates a SCSI adapter device number value which is placed in the **sc_buf**. When an error occurs that temporarily blocks the path to the array, the **arrayinfo** structure contains information about the state of the array. For example, during a controller switch from the active/passive state to the dual active state, the **router_info** structure may be used to quiesce all **hdisk** activity on the array. As another example, during a LUN switch from one controller to another, the **router_info** structure temporarily quiesces a single LUN until the LUN switch over completes. Other LUNs on the array continue to process I/O requests in the normal fashion.

Figure 11 compares the two-layer model to the three-layer model used to support the 2102 Storage Server.

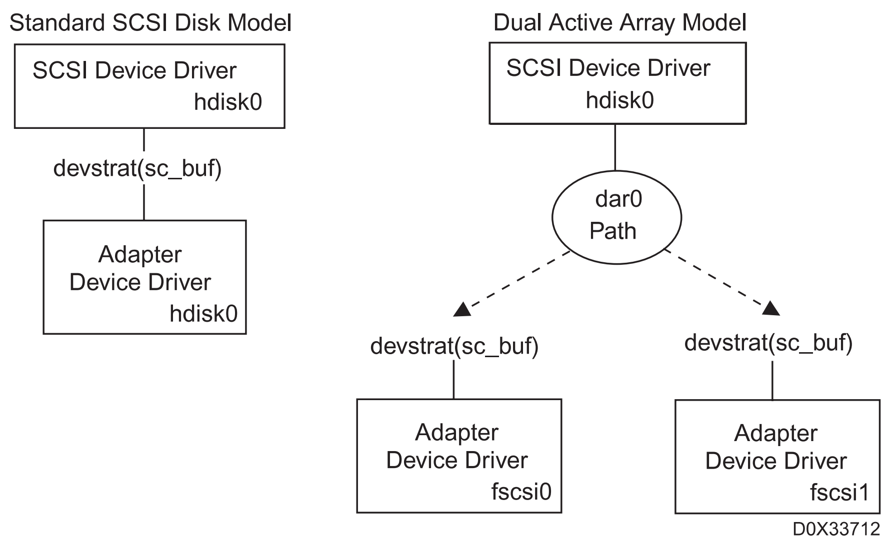

Figure 11. Multiple Array Paths

fcparray FC SCSI Device Driver

Purpose

Support the FC SCSI physical volume (fixed disk) array device driver.

Syntax

#include <sys/devinfo.h>
#include <sys/scsi.h>
#include <sys/scsi_buf.h>
#include <sys/fcparray.h>
#include <sys/errids.h>

Description

Attention: Potential exists for data corruption or system crashes. Data corruption, loss of data, or loss of system integrity will occur if devices supporting paging, logical volumes, or mounted file systems are accessed using block special files. Block special files are provided for logical volumes and disk devices and are solely for system use in managing file systems, paging devices, and logical volumes. These files should not be used for other purposes. Additional information concerning the use of special files is in *AIX Version 4 Kernel Extensions and Device Support Programming Concepts*.

The **fcparray** device driver provides all of the functions of a SCSI head device driver in terms of request queueing and error recovery. The device driver detects controller hardware state changes, adapter hardware errors, and LUN ownership changes and issues the appropriate commands to change the hardware state and/or modifies the data path to a device. In addition, the **fcparray** device driver initiates state changes received via the FC SCSI RAIDiant Disk Array Manager.

The device driver provides a monitoring capability which is used to test components not actively in use (passive controller health checks), thereby ensuring the active components have backups. These checks include actual data transfers through the passive controller which ensure a complete path to the attached drives in functioning properly.

In dual active mode, the **fcparray** device driver provides enhanced switching mechanisms. For example, the host which is not issuing the controller or LUN switch commands can gracefully quiesce the controller activity and remain idle until the switch is complete.

Devices Configure, Open and Close

Since the **fcparray** device driver supports these different types of logical devices: **dars**, **dacs** and **hdisks**, the driver allocates device resources defined by **arrayinfo** and **router_info** at configuration time based on the device type. Although a **dac** device can share the same SCSI ID and LUN value as an **hdisk**, it has a separate **arrayinfo** structure because it is logically a separate device. LUN sharing is accomplished when **dac** commands are executed by quiescing the corresponding **hdisk** before executing the **dac** command.

The **dar** configuration accepts the **array_ddi** configuration data from the **cfgfdar** configuration method. Information passed in this structure identifies the configuration as a **dar** configuration and contains no other information. A **router_info** structure allocates and initializes for the **dar** instance. Unconfiguring the **dar** instance sets a flag indicating that this array is no longer configured, causing the monitor process to terminate.

The **dac** configuration accepts the **array_ddi** configuration data from the **cfgfdac** configuration method. The **array_ddi** structure is filled out with all device information as a normal **hdisk** would be. Information in the **arrayinfo** structure identifies this configuration as a **dac** device configuration, and the **adapter_devno** field in the **diskinfo** structure is set to the SCSI parent device number. This value is maintained so that command accesses through the **dac** device can be directed at a specific adapter and are not rerouted as they are for **hdisk** devices.

The **hdisk** configuration accepts the **array_ddi** configuration data from the **cfgfcparray** configuration method. The **array_ddi** structure is filled out with all device information, and a field in the structure identifies this as an **hdisk** device instance. To properly pass I/O requests along different paths, the **arrayinfo** structure for the **hdisk** will contain a pointer in its **arrayinfo** structure for **lun_pair**, which is set to the **dac** device. The **dac arrayinfo** has a similar pointer back to the **hdisk** being configured.

To eliminate the issuing of multiple controller switching **ioctls** from different sources, the **dar** device can only be opened for a single instance. Using the **dar** device does not cause any SCSI commands to be issued to the array controller unless specified via the **dar**-only **ioctls** to add a path, set the router state for an array, or set the controller assignments for a LUN. The device entry points for reading, writing, strategy, and dump are not valid accesses for the **dar**.

The **dac** device can only be opened in normal mode. The **dac** device driver entry points are not valid for reads, writes, strategy, or dump. When the device is opened, the LUN assignment is not made, as it is not needed until a command is executed. There is no dependency between the **dac** device opens and the **hdisk** device that may share the same LUN assignment. Opens in normal mode should use the **DKIOLCMD_RS ioctl** that processes the command as a tagged request and does a Request Sense operation should a Contingence Allegiance Condition (CAC) occur. Diagnostic opens can use the normal **DK_IOLCMD** passthrough mechanisms as they are assured exclusive SCSI ID and LUN access. Opening of a **dac** device does not cause the start sequence of SCSI command to be issued to the device because there may not be a LUN configured at the SCSI ID/LUN location if an **hdisk** is not yet defined.

The **hdisk** device can be opened in the normal manner. Opening an **hdisk** in normal mode the first time causes the start sequence of SCSI commands to be issued on the first open even if a **dac** has already been opened at the same location.

The **dar** close routine does not issue any SCSI commands or close any SCSI devices. This close routine tracks the number of opens to ensure exclusive access.

With a few exceptions, the **dac** and **hdisk** close routines are similar to the current close mechanisms. Since AIX only calls the close device routine on the last closure, device locations with both a **dac** and an **hdisk** require special processing. A release sequence is not attempted for a **dac** device because a LUN may not exist at the device location. This is done only when the **hdisk** is closed.

Device-Dependent Subroutines

Typical fixed-disk operations are implemented using the **open, read, write,** and **close** subroutines.

open and close Subroutines: The **openx** subroutine is intended primarily for use by the diagnostic commands and utilities. Appropriate authority is required for execution. Attenpting to execute this subroutine without the proper authority results in a return value of -1, with the **errno** global variable set to a value of **EPERM.**

The *ext* parameter passed to the **openx** subroutine selects the operation to be used for the target device. The **/usr/include/sys/scsi.h** file defines possible values for the *ext* parameter. The parameter can contain any combination of the following flag values logically ORed together:

SC_DIAGNOSTIC
Places the selected device in Diagnostic mode. This mode is singularly entrant; that is, only one process at a time can open it. When a device is in Diagnostic mode, SCSI operations are performed during **open** or **close** operations, and error logging is disabled. In Diagnostic mode, only the **close** and **ioctl** operations are accepted. All other device-supported subroutines return a value of -1, with the **errno** global variable set to a value of **EACCES.**

A device can opened in Diagnostic mode only if the target device is not currently opened. If an attempt is made to open a device in

Diagnostic mode and the target device is already open, a value of -1 is returned and the errno global variable is set to a value of **EACCES**.

SC_FORCED_OPEN
Forces a bus device reset (BDR), regardless of whether another initiator has the device reserved. The SCSI bus device reset is sent to the device before the **open** sequence begins; otherwise, the **open** executes normally.

SC_RETAIN_RESERVATION
Retains the reservation of the device after a **close** operation by not issuing the release. This flag prevents other initiators from using the device unless they break the host machine's reservation.

For more specific information on the open operations, see *AIX Version 4 Kernel Extensions and Device Support Programming Concepts*.

readx and writex Subroutines: The **readx** and **writex** subroutines provide additional parameters affecting the raw data transfer. These subroutines pass the *ext* parameter, which specifies request options. The options are constructed by logically ORing zero or more of the following values:

HWRELOC Indicates a request for hardware relocation (safe relocation only)

UNSAFEREL Indicates a request for unsafe hardware relocation

WRITEV Indicates a request for write verification

ioctl Subroutine: The following ioctl operation is available for any device:

IOCINFO Returns the **devinfo** structure defined in the **/usr/include/sys/devinfo.h** file. The **IOCINFO** operation is the only operation defined for all device drivers that use the **ioctl** subroutine. The remaining operations are all specific to the physical volume device.

The following ioctl operations are available for **hdisks:**

DKIOLCMD When the device has been successfully opened in the Diagnostic mode, this operation provides the means for issuing any SCSI command to the specified device. If the **DKIOLCMD** is issued when the device is not in Diagnostic mode, a value of **-1** is returned and the **errno** global variable set to a value of **EACCES**. The device driver performs no error recovery or logging on failures of this **ioctl** operation.

The SCSI status byte and the adapter status bytes are returned through the *arg* parameter, which contains the address of a **scsi_iocmd** structure (defined in the **/usr/include/sys/scsi_buf.h** file). If the **DKIOLCMD** operation returns a value of **-1** and the **errno** global variable is set to a nonzero value, the requested operation has failed. In this case, the caller should evaluate the returned status bytes to determine why the operation was unsuccessful and what recovery actions should be taken.

The **devinfo** structure defines the maximum transfer size for the command. If an attempt is made to transfer more than the maximum, a value of -1 is returned and the **errno** global variable set to a value of **EINVAL**. Refer to the *Small Computer System Interface (SCSI) Specification* for the applicable device to get request sense information.

DKIOLRDSE
Provides a means for issuing a **read** command to the disk and obtaining the target device sense data on error. If the **DIKIOLRDSE** operation returns a value of **-1 and the**

`status_validity`

field has **sc_valid_sense** set, then valid sense data is returned. Otherwise, target sense data is omitted.

The **DKIOLRDSE** operation is provided for diagnostic use. It allows the limited use of the target device while operating in an active system environment. The *arg* parameter to the **DKIOLRDSE** operation contains the address of a **scsi_rdwrt** structure. This structure is defined in the **/usr/include/sys/scsi_buf.h** file.

The **devinfo** structure defines the maximum transfer size for a read. If an attempt is made to transfer more than the maximum, a value of **-1** is returned and the **errno** global variable set to a value of **EINVAL**. Refer to the *Small Computer System Interface (SCSI) Specification* for the applicable device for the format of request sense information.

DKIOLWRSE
Provides a means for issuing a **write** command to the disk and obtaining the target device sense data on error. If the **DKIOLWRSE** operation returns a value of **-1** and the

`status_validity`

field has **sc_valid_sense** set, then valid sense data is returned. Otherwise, target sense data is omitted. The **DKIOLWRSE** operation is provided for diagnostic purposes to allow for limited use of the target device while operating in an active system environment. the *arg* parameter to the **DKIOLWRSE** operation contains the address of a **scsi_rdwrt** structure. This structure is defined in the **/usr/include/sys/scsi_buf.h** file. The **devinfo** structure defines the maximum transfer size for a write. If an attempt is made to transfer more than the maximum, a value of **-1** is returned and the **errno** global variable set to a value of **EINVAL**. Refer to the *Small Computer System Interface (SCSI) Specification* for the applicable device to get the particular request sense information.

The following **ioctl** operations are available for **hdisks** and **dacs**:

DKIOLCMD_RS
When the device has been successfully opened, this operation provides the means for issuing any SCSI command to the device and obtaining the target device request sense data on error. If the **DKIOLCMD_RS** returns a value of **-1**, the **errno** global variable is set to a value of **EACCES**. The device driver performs error recovery on failures of this **ioctl** operation.

The SCSI status byte and the adapter status bytes are returned through the *arg* parameter, which contains the address of a **scsi_iocmd** structure (defined in the **/usr/include/sys/scsi_buf.h** file). If the **DKIOLCMD_RS** operation returns a value of **-1** and the

errno global variable is set to a nonzero value, the requested operation has failed. In this case, the caller should evaluate the returned status bytes to determine why the operation was unsuccessful and what recovery actions should be taken.

The **devinfo** structure defines the maximum transfer size for the command. If an attempt is made to transfer more than the maximum, a value of **-1** is returned and the **errno** global variable set to a value of **EINVAL**. Refer to the *Small Computer System Interface (SCSI) Specification* for the applicable device to get request sense information.

DKIOLCMD_PT
When the device has been successfully opened, this operation provides the means for issuing a passthrough SCSI command to a drive that is attached to the specified device and obtaining the target device request sense data on error. If the **DKIOLCMD_PT** returns a value of **-1**, the **errno** global variable is set to a value of **EACCES**. The device driver performs error recovery on failures of this **ioctl** operation.

The SCSI status byte and the adapter status bytes are returned through the *arg* parameter, which contains the address of a **scsi_iocmdpt** structure (defined in the **/usr/include/sys/fcparray.h** file). If the **DKIOLCMD_PT** operation returns a value of **-1** and the **errno** global variable is set to a nonzero value, the requested operation has failed. In this case, the caller should evaluate the returned status bytes to determine why the operation was unsuccessful and what recovery actions should be taken.

The **devinfo** structure defines the maximum transfer size for the command. If an attempt is made to transfer more than the maximum, a value of **-1** is returned and the **errno** global variable set to a value of **EINVAL**. Refer to the *Small Computer System Interface (SCSI) Specification* for the applicable device to get request sense information.

Error Conditions

In addition to those errors listed, **ioctl, open, read**, and **write** subroutines against this device fail in the following circumstances:

EACCES An attempt was made to open a device currently opened in Diagnostic mode.

EACCES An attempt was made to open a diagnostic session on a device already opened.

EACCES The user attempted a subroutine other than an **ioctl** or **close** subroutine while in Diagnostic mode.

EACCES A **DKIOLCMD** operation was attempted on a device not in Diagnostic mode.

EBUSY The target device is reserved by another initiator.

EINVAL The **read** or **write** subroutine supplied an *nbyte* parameter that is not and even multiple of the block size.

EINVAL A sense data buffer length of greater than 255 bytes is not valid for a **DKIOLWRSE** or **DKIOLRDSE** operation.

EINVAL The data buffer length exceeded the maximum defined in the **devinfo** structure for a **DKIOLRDSE, DKIOLWRSE**, or **DKIOLCMD ioctl** operation.

EINVAL An unsupported **ioctl** operation was attempted.

EINVAL The data buffer length exceeded the maximum defined for a **strategy** operation.

EIO The target device cannot be located or is not responding.

EIO The target device has indicated an unrecovered hardware error.

EMEDIA The target device has indicated an unrecovered media error.

EMFILE An **open** was attempted for an adapter that already has the maximum permissible number of opened devices.

ENXIO The command could not be completed and should be retried.

ENXIO A **read** or **write** command was attempted beyond the end of the disk.

EPERM The attempted subroutine requires appropriate authority.

Reliability and Serviceability Information

The following errors are returned from SCSI disk array devices:

ABORTED COMMAND
Indicates the device ended the command.

ADAPTER ERRORS
Indicates the adapter returned an error.

GOOD COMPLETION
Indicates that the command completed successful.

HARDWARE ERROR
Indicates that an unrecoverable hardware failure occurred during command execution or during a self-test.

ILLEGAL REQUEST
Indicates and illegal command or command parameter.

MEDIUM ERROR
Indicates that the command ended with an unrecovered media error condition.

NOT READY Indicates that the logical unit is offline.

RECOVERED ERROR
Indicates that the command was successful after some recovery applied.

UNIT ATTENTION
Indicates the device has been reset of the power has been turned on.

Disk Array Error Record Values

The following error types are defined for the RAIDiant Array and are logged into the AIX error log. In some cases, notification messages are sent to the systems console.

FCP_ARRAY_ERR1

Error Description: ARRAY OPERATION ERROR

This error is logged by the **fcparray** adapter device driver in the event of a permanent hardware error involving the disk array media. Either these errors cannot be retried, or all allowed retries have been exhausted. It is unlikely that any operations to the **fcparray** adapter can continue after one of these errors.

Detail Data

The sense data consists of failure data, which is analyzed by the diagnostic programs.

FCP_ARRAY_ERR2

Error Description: ARRAY OPERATION ERROR

This error is logged by the **fcparray** adapter device driver in the event of a permanent hardware error involving the disk array device hardware or related hardware. Either these errors cannot be retried, or all allowed retries have been exhausted. It is unlikely that any operations to the **fcparray** adapter can continue after one of these errors.

FCP_ARRAY_ERR3

Error Description: ARRAY OPERATION ERROR

This error is logged by the **fcparray** adapter device driver in the event of a permanent hardware error involving the array device and all related hardware that was detected by the disk array adapter. Either these errors cannot be retried, or all allowed retries have been exhausted. It is unlikely that any operations to the **fcparray** adapter can continue after one of these errors.

Detail Data: The sense data consists of failure data, which is analyzed by the diagnostic programs.

FCP_ARRAY_ERR4

Error Description: ARRAY OPERATION ERROR

This error is logged by the **fcparray** device driver in the event of a potentially catastrophic hardware error involving the disk array or its related hardware, or both. Either these errors cannot be retried, or all allowed retries have been exhausted. Since these errors are not necessarily catastrophic, operations may or may not continue successfully after the error.

Detail Data: The sense data consists of failure data, which is analyzed by the diagnostic programs.

FCP_ARRAY_ERR5

Error Description: UNDETERMINED ERROR

This error is logged by the **fcparray** device driver in the event of a potentially permanent hardware failure. Either these errors cannot be retried, or all allowed retries have been exhausted. Since these errors are not necessarily catastrophic, operations may or may not continue successfully after the error.

Detail Data: The sense data consists of failure data, which is analyzed by the diagnostic programs.

FCP_ARRAY_ERR6

Error Description: SUBSYSTEM COMPONENT FAILURE

A degradation condition has been detected on a disk array controller. Either the disk has become degraded or the disk controller has detected an error.

Detail Data: The sense data consists of failure data, which is analyzed by the diagnostic programs.

FCP_ARRAY_ERR7

Error Description: CONTROLLER HEALTH CHECK FAILURE

A health check on the passive controller has failed on a disk array controller.

Detail Data: The sense data consists of failure data, which is analyzed by the diagnostic programs.

FCP_ARRAY_ERR8

Error Description: ARRAY ACTIVE CONTROLLER SWITCH FAILURE

A condition has occurred which has caused an array controller to be switched for device **dac0**.

Note: This is not an error condition, but is logged for information purposes.

Detail Data: The sense data consists of additional data.

FCP_ARRAY_ERR9

Error Description: ARRAY CONTROLLER SWITCH FAILURE

An array controller switch has failed on a disk array controller.

Detail Data: The sense data consists of failure data, which is analyzed by the diagnostic programs.

FCP_ARRAY_ERR10

Error Description: ARRAY CONFIGURATION CHANGED

A condition has occurred which has caused an array configuration to change for a disk array controller

Note: This is not an error condition, but is logged for information purposes.

Detail Data: The sense data consists of additional data.

FCP_ARRAY_ERR11

Error Description: IMPROPER DRIVE TYPE FOR DUAL ACTIVE MODE

An array drive replacement problem has occurred on device **dac0**.

Detail Data: The sense data consists of failure data, which is analyzed by the diagnostic programs.

FCP_ARRAY_ERR12

Error Description: POLLED AEN FAILURE

An automatic error notification has failed on device **dac0**.

Detail Data: The sense data consists of failure data, which is analyzed by the diagnostic programs.

FCP_ARRAY_ERR13

Error Description: ARRAY INTER-CONTROLLER COMMUNICATIONS FAILURE

An array inter-controller communications problem has occurred on device **dac0**.

Detail Data: The sense data consists of failure data, which is analyzed by the diagnostic programs.

FCP_ARRAY_ERR14

Note: This disk array error record value is available with AIX Version 4.1 only.

Error Description: ARRAY DRIVE FAILURE

A disk drive has failed.

Detail Data: The sense data consists of failure data, which is analyzed by the diagnostic programs.

FCP_ARRAY_ERR15

Error Description: CACHE BATTERY LOW/DATA LOSS POSSIBLE

This error will occur after the 2102 Storage Server has been powered off while there was data in the write cache that needs to be written to the disks and the batteries which supply power to the RAM to maintain the data will have drained below the accepted safe level.

FCP_ARRAY_ERR16

Error Description: CACHE BATTERY CHARGE BELOW 87.5%

This error is the result of the controller detecting that the charge left on the batteries is insufficient to guarantee the data held in cache will remain intact should a power failure occur.

Error Record Format

This section identifies the fields that are defined in the error record for degradation errors.

An error log report contains the following information:

Note: Not all errors will generate information for each of the following categories.

LABEL Predefined name for the event

ID Numerical identifier for the event

Date/Time Date and time of the event

Sequence Number
Unique number for the event

Machine ID Identification number of your system processor unit

Node ID Mnemonic name of your system

Class General source of the error. The possible error classes are:

- **H** Hardware. (When you receive a hardware error, refer to your system operator guide for information about performing diagnostics on the problem device or other piece of equipment. The diagnostics program tests the device and analyzes the error log entries related to it to determine the state of the device.)
- **S** Software.
- **O** Informational messages.
- **U** Undetermined (for example, a network).

Type Severity of the error that has occurred. Six types of errors are possible.

PEND The loss of availability of a device or component is imminent.

PERF The performance of the device or component has degraded to below an acceptable level.

PERM Condition that could not be recovered from. Error types with this value are usually the most severe errors and are more likely to mean that you have a defective hardware device or software module. Error types other than PERM usually do no indicate a defect, but they are recorded so that they can be analyzed by the diagnostics programs.

TEMP Condition that was recovered from after a number of unsuccessful attempts. This error type is also used to record informational entries, such as data-transfer statistics for DASD devices.

UNKN It is not possible to determine the severity of the error.

INFO The error log entry is informational and was not the result of an error.

Resource Name
Name of the failing resource (for example, a device type of hdisk).

Resource Class
General class of the failing resource (for example, a device class of disk),

Resource Type
Type of the failing resource (for example, a device type of 355 MB).

Location Code
Path to the device. There may be up to four fields, which refer to drawer, slot, connector, and port, respectively.

VPD
Vital product data, The contents of this field, if any, vary. Error log entries for devices typically return information concerning the device manufacturer, serial number, Engineering Change levels, and Read Only Storage levels.

Description
Summary of the error.

Probable Cause
List of some of the possible sources of the error.

User Causes
List of possible reasons for errors due to user mistakes. An improperly inserted disk and external devices (such as modems and printers) that are not turned on are examples of user-caused errors.

Recommended Actions
Description of actions for correcting a user-caused error.

Install Causes
List of possible reasons for errors due to incorrect installation or configuration procedures. Example of this type of error include hardware and software mismatches, incorrect installation of cables or cable connections becoming loose, and improperly configured systems.

Recommended Actions
Description of actions for correcting an installation-caused error.

Failure Causes
List of possible defects in hardware or software.

Note: A failure-causes section in a software error log usually indicates a software defect. Logs that list user or install causes or both, but not failure causes, usually indicate that the problem is not a software defect.

If you suspect a software defect, or are unable to correct user or install causes, report the problem to your software service department.

Recommended Actions
Description of actions for correcting the failure. For hardware errors `PERFORM PROBLEM DETERMINATION PROCEDURES` is one of the recommended actions listed. For hardware errors, this will lead to running the diagnostic programs.

Detailed Data
Failure data that is unique for each error log entry, such as device sense data. See “Interpreting Detail Sense Data Format”.

Interpreting Detail Sense Data Format

Refer to the *FCRDAM and Expandable Storage Unit, Service Guide, IBM 2102 Model F10 and Model D00* for information about how to interpret sense data.

DAC Special File

Purpose

Provides access to a SCSI disk array controller (DAC) on a 2102 Storage Server device.

Description

The **DAC** special file is primarily used by RAIDiant Array diagnostic programs. In addition, the **fcparray** device driver uses the **dar** special file to maintain multiple paths to a 2102 Storage Server device.

Implementation Specifics

The **fcparray** SCSI device driver provides further information and implementation specifics.

Files

/dev/dac0,/dev/dac1,.../dev/dac*n*

Provide access to a SCSI controller on a 2102 Storage Server.

DAR Special File

Purpose

Provides access to a disk array router (DAR) on a RAIDiant Array device.

Description

The **dar** special file is primarily used by the **fcparray** device driver to maintain multiple paths to a RAIDiant Array device.

Implementation Specific

The **fcparray** SCSI device driver provides further information and implementation specifics.

Files

/dev/dar0,/dev/dar1,.../dev/dar*n*

Provide access to a DAR pseudo-device.

SCSI Commands for the 2102 Fibre Channel RAID Storage Server

The set of SCSI commands that support the 2102 Storage Server are used by both the active and standby controllers. The following operations have the maximum command execution times shown:

Start/Stop Spindle
: 60 seconds

Format Unit Depends on the drive capacity

Reassign Block
: 2 minutes

Send Diagnostics
: 120 seconds

Read/Write Commands
: 30 seconds

Mode Select 120 seconds

Write Buffer 120 seconds

Other SCSI Commands
: 30 seconds

If a drive fails to respond to a command within the defined time interval, then the system issues a SCSI Bus Reset. The command that caused the timeout condition is logged in the system error log as a Permanent (non-recoverable) Equipment Check.

The following SCSI commands must be supported by the system software (device driver) in order for the drive to be tested:

Format Unit Provides a method for reinitializing 2102 Storage Server media. The command causes the ID fields as well as the data fields to be rewritten.

Before issuing the **Format Unit** command, the **Mode Select** command should be sent to the drive so that the correct format parameters are used.

Inquiry Provides a method for the system to obtain the following vital product data from the disk drive:

```
Bytes Field     0-7
ANSI Defined Fields     8-15
Vendor ID       16-31
Product Identification  32-35
Product revision Level
Page Code 0x80 and EVPD = 1    4-19
Unit Serial Number
```

Serial Number
: A unique number must be provided for each drive. It does not have to be the same as the bar code serial number on the label

Vendor Must be IBM. Distinguishes an IBM assembly from a vendor assembly.

Product ID May contain a single Product Type field or contain Product Type and Model Number fields.

Product Revision Level xxyy
: Used to determine what level of microcode to download.

Mode Sense or **Mode Select**
: Provides a method for the program to determine various device

parameters from the disk drive (via SCSI **Mode Sense** command) and to specify how those parameters, which are changeable to the LUNs/disk drive, (via SCSI **Mode Select** command) should be set. Not all parameters listed in the table are changeable by all drives. Only the parameters that are defined as changeable by each device type can be changed.

Read
Requests that the target transfer data to the initiator starting at the LBA specified in the command and that the transfer last for as many blocks as specified in the length field. Both the **Read** (08h) and **Read Extended** (28h) commands need to be supported. The Device Driver will always set the DPO and FUA bits to zero when using the **Read** (10) command.

Read Capacity
Provides a method to determine the Logical Block Address and the Block Length (in bytes) of the last logical data block on the disk drive.

The **Read Capacity** command can also be used to determine, for a given Logical Block Address, the Block Length (in bytes) and the Logical Block Address of the last logical data block that can be accessed prior to requiring a physical seek.

Reassign
Requests the drive to reassign a logical block to an available spare location. The Reassign Block operation must complete successfully for the file to be usable. Therefore if the command is interrupted by a reset or power outage, the command will automatically continue from the last checkpoint when the drive has recovered from the interruption. No data will be lost or sector destroyed (ECC Uncorrectable Error).

It is suggested that the system always preserve the data before reassignment (if possible) and always write the reassigned sector immediately after the **Reassign** block command has completed.

Request Sense
Provides a method for transferring sense data from the disk drive to the test program. If should be used anytime CHECK CONDITION Status is returned from the disk drive.

Note: Not all SCSI-2 drives will return the same number of sense bytes. The 2102 Storage Server returns up to 18 bytes of information, when microcode is not download. It is recommended that FFh be used as the allocation length.

Reserve or **Release**
Reserves or releases an LUN for an initiator.

Send Diagnostics
Provides a method for the test program to request that the RAIDiant Array subsystem perform diagnostic tests on itself or to the drives after setting the passthrough mode. SlfTst = 1, and PF=0.

Set Passthrough Mode
Sends commands to the drives in the 2102 Storage Server. The drive is identified by the channel number and drive SCSI ID in the command. A SCSI command can be sent to the drive by sending the **Set Passthrough Mode** command followed by the SCSI command to the 2102 Storage Server. Every SCSI command to the

drive requires a **Set Passthrough Mode** command before it . To receive the Request Sense status after a check condition, send the **Request Sense** command to the 2102 Storage Server without the **Set Passthrough Mode** command.

Start/Stop
Provides a method to start or stop the spindle motor of the disk drive.

Test Unit Ready
Provides a method for the test program to determine if the disk drive logical unit is Ready. This command may be used after issuing the **Start/Stop** unit commands with an Immediate bit of 1 to determine when the motor has reached proper speed.

The **Test Unit Ready** command is not intended to be a diagnostic. No self-analysis is performed by the logical unit as a result of this command.

Write
Requests that the target write the data transferred by the initiator to the medium starting at the LBA defined in the command and lasting for the number of blocks specified in the length field. The device driver will always set the DPO and FUA bits to zero when using the **Write** (10) command. The Erase Bypass bit in the **Write** command should be supported. This would improve write performance when used during writes to pre-erased sectors.

Write/Read Buffer
Tests target memory and SCSI bus integrity in conjunction with the READ BUFFER command.

Write and Verify
Signals the drive to write data and then verify that it is correctly written. The verify operation is an ECC check only.

For BytChk=0, if an ECC error is detected, then Check Condition status is returned with a Medium Error Sense key. This command will be used as a part of the process to qualify a sector to be relocated in that any sector that meets the DRP level defined in Appendix B for a read error will be rewritten using the **Write and Verify** command. If any data error occurs during the verify, the sector will be reassigned. The DPO bit should be 0.

Appendix B. FC SCSI RAIDiant Disk Array Manager Messages

The Fibre Channel Storage Manager for AIX and the disk array software generate two basic types of error messages:

- RAIDiant Disk Array Manager error messages. These messages occur while you are performing 2102 Storage Server operations. These messages are displayed on the command screen as soon as they occur.
- Standard AIX error messages. These messages indicate disk availability, mount status, and so on. They are also used for disk devices other than the array subsystem. See *AIX Version 4 Messages Guide and Reference* for an explanation of these messages.

Fibre Channel Storage Manager for AIX Messages

The Fibre Channel Storage Manager for AIX reports error messages to the command screen as they occur. These error messages indicate that one of the following errors has occurred:

- An invalid value for an option was entered.
- An incorrect device name was identified.
- The device specified is mounted.
- A miscommunication occurred between the array controller and the Fibre Channel Storage Manager.
- Root-user permissions were needed to execute a command.
- An installation error occurred.
- An array-controller error occurred.
- The controller firmware or error message files do not exist.

Error Message Definitions

This section lists the error messages displayed by Fibre Channel Storage Manager utilities and the actions to take when they occur. When more than one option is offered in the ″Action to Take″ instructions, perform Option 1 first. If that does not correct the problem, try Option 2, and so on.

Device is not available

Probable Cause

There is a communication problem between the Fibre Channel Storage Manager and the array controller.

Action to Take

Option 1:

Make sure that all cables are securely connected, and retry.

Option 2:

Call your service representative.

SCSI Mode Select command to add a 2102 Storage Server array has failed

Probable Cause

There is a communication problem between the 2102 Storage Server and the array controller.

Action to Take

Option 1:

Make sure that all cables are securely connected, and retry.

Option 2:

Call your service representative.

SCSI Mode Select command to format a 2102 Storage Server array has failed

Probable Cause

There is a communication problem between the 2102 Storage Server and the array controller.

Action to Take

Option 1:

Make sure that all cables are securely connected, and retry.

Option 2:

Call your service representative

SCSI Mode Select command to modify a drive status has failed

Probable Cause

There is a communication problem between the 2102 Storage Server and the array controller.

Action to Take

Option 1:

Make sure that all cables are securely connected, and retry.

Option 2:

Call your service representative

SCSI Mode Select command to modify a 2102 Storage Server array has failed

Probable Cause

There is a communication problem between the 2102 Storage Server and the array controller.

Action to Take

Option 1:

Make sure that all cables are securely connected, and retry.

Option 2

Call your service representative

Acquisition of SCSI Mode Sense Data from the device has failed

Probable Cause

There is a communication problem between the 2102 Storage Server and the array controller.

Action to Take

Option 1

Make sure that all cables are securely connected, and retry.

Option 2:

Call your service representative

SCSI Mode Sense/Select commands to Check Available Capacity has failed

Probable Cause
There is a communication problem between the 2102 Storage Server and the array controller.

Action to Take
Option 1

Make sure that all cables are securely connected, and retry.

Option 2

Call your service representative

SCSI Mode Select command to delete a 2102 Storage Server array has failed

Probable Cause
There is a communication problem between the 2102 Storage Server and the array controller.

Action to Take
Option 1:

Make sure that all cables are securely connected, and retry.

Option 2:

Call your service representative

Configuration method invoked to change lun ownership has failed

Probable Cause
OS command is failing.

Action to Take
Option 1

Reboot the system.

Option 2:

Call your service representative

System call ’rmdev’ for RAIDiant array hdisk has failed

Probable Cause
OS command is failing.

Action to Take
Option 1

Reboot the system.

Option 2

Call your service representative

Configuration method invoked to change ’queue depth’ or ’reserve lock’ has failed

Probable Cause
OS command is failing.

Action to Take

Option 1

Reboot the system.

Option 2

Call your service representative

System call ’mkdev’ for RAIDiant array hdisk has failed

Probable Cause
OS command is failing.

Action to Take

Option 1:

Reboot the system.

Option 2:

Call your service representative

Parity Check/Repair has failed

Probable Cause
While performing a Parity Check and Repair, another communication error has occurred.

Action to Take

Option 1:

Make sure that all cables are securely connected, and retry.

Option 2:

Call your service representative.

Parity Check/Repair cannot be performed on a RAID 0 device

Probable Cause
You have selected an **hdisk** which is currently configured as a RAID 0 device on which to perform a Parity Check. This is not permitted.

Action to Take
Perform Parity Checks only on hdisks which are configured with parity drives defined in their RAID level.

You are attempting to perform a Parity Check & Repair on a

logical unit which is currently degraded. The logical unit must be in OPTIMAL status before a Parity Check&Repair will be performed.

Probable Cause
You have selected an **hdisk** which currently has a status other than OPTIMAL on which to perform a Parity Check. this is not permitted.

Action to Take
Perform Parity Checks only on hdisks which are OPTIMAL.

Please wait - Formatting the Array ...typically this format will

take less than 30 minutes...in some cases, it may take up to 4 hours.

Probable Cause

This is not an error condition. It is a warning that the formatting of a new LUN (upon creation) may take a significant amount of time.

Action to Take

Option 1:

If the format takes longer than four hours, make sure all cables are securely connected, and retry...

Option 2: Call your service representative.

No disks have been defined for this array

Probable Cause

No disks were selected on the previous panel.

Action to Take

Return to previous panel and make valid drive selections.

You have attempted to exceed the maximum number of hdisks possible for this subsystem

Probable Cause

An attempt has been made to create more than the maximum allowed LUNs per subsystem.

Action to Take

If you want a new logical unit, delete an existing logical unit. Remember to back up any data on the logical unit before it is deleted.

Glossary

array controller. The function (hardware and software) within a 2102 Fibre Channel RAID Storage Server that controls the flow of data, commands, and status between the SCSI connection to the host and the disk drives within the 2102 Storage Server.

bank. A physical set of up to ten disk drives, one attached to each of the six SCSI channels within a 2102 Fibre Channel RAID Storage Server. All drives in a bank have the same SCSI ID.

DAC. disk array controller.

dar. disk array router

DASD. direct access storage device.

direct access storage device. A disk drive.

disk array. The set of physical disk drives within a 2102 Fibre Channel RAID Storage Server.

disk array controller (DAC). An array controller. The operating system uses the name **DAC** for an array controller.

disk array router (dar). A pseudo-device used in the definition of multiple paths to a 2102 Fibre Channel RAID Storage Server. Disk array routers control the automatic switching between paths that is necessary when the failure of a component makes one path unusable. The operating system uses the name **dar** for a disk array router.

drive group. A set of 1 to 20 drives that have been configured into one or more logical units.

fast load. A 2102 Fibre Channel RAID Storage Server option that is typically used instead of the fast-write option when loading data from archive storage.

fast write. A 2102 Fibre Channel RAID Storage Server option that offers additional protection of cached data by mirroring the data in the cache of the second controller.

FC. fibre channel.

FCAL. fibre channel arbitrated loop.

FCP. fibre channel for SCSI protocol.

Fibre Channel Storage Manager for AIX. A program that is accessed via SMIT and used to manage the logical units of an array system.

FC SCSI RAIDiant Disk Array Manager (FCRDAM). A set of SMIT menus in the Fibre Channel Storage Manager for AIX program.

FCRDAM. FC RAIDiant Disk Array Manager.

fibre channel. A communications system characterized by a large bandwith, shared bus and (often) optical fibre connections.

fibre channel arbitrated loop. A type of FC network in which the hosts and devices are in a loop configuration, with each node communicating with the next node and messages being passed along to the target.

fibre channel for SCSI protocol. An implementation of the SCSI interface command set for FC networks.

group. A logical set of up to ten disk drives within a 2102 Fibre Channel RAID Storage Server. A group can contain more than one LUN.

hard disk. A fixed disk drive that is defined to the AIX operating system. A LUN within a 2102 Fibre Channel RAID Storage Server appears as a single hard disk to the operating system. The operating system uses the name **hdisk** for a hard disk.

hot-spare drive. A drive, which is not part of a drive group, that is set to the hot-spare status.

location code. The eight-digit code by which the operating system addresses I/O devices.

logical unit (LUN). A disk array that is treated as a single disk drive by the AIX operating system. In a 2102 Fibre Channel RAID Storage Server, a LUN is split across up to ten physical disk drives, all of which must be in the same group.

LUN. logical unit.

RAID. Redundant Array of Independent Disks.

RAID level. One of a standard set of definitions of how a RAID is configured and controlled. Each level provides different performance and a different degree of protection against the affects of the failure of one or more of the disk drives within the RAID.

RAIDiant Disk Array Manager. A set of SMIT menus in the Fibre Channel Storage Manager for AIX.

rank. A bank of disk drives.

Redundant Array of Independent Disks. A set of disk drives configured and controlled to provide enhanced availability.

SCSI. small computer system interface.

SCSI adapter. The function (hardware and software), residing in the host, to which the SCSI attaches. Sometimes referred to as the SCSI controller.

SCSI bus. The physical connections between devices on a SCSI. In a 2102 Fibre Channel RAID Storage Server, a SCSI channel.

SCSI channel. In a 2102 Fibre Channel RAID Storage Server, one of the six physical connections between the disk drives and the array controller. Up to ten disk drives can be attached on each SCSI channel.

SCSI ID. In a 2102 Fibre Channel RAID Storage Server, the number (0 through F) of a disk drive on a SCSI channel.

small computer system interface. A standard (physical and logical) connection between I/O devices and a host.

write cache. A function that improves array performance by keeping recently written information in memory located in the array controllers.

Index

Numerics

A

B

C

D

E

F

G

H

U

V

W

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

Fibre Channel Storage Manager
Installation and User’s Guide
for AIX Operating System

Publication No. SC26-7290-00

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

	Very Satisfied	Satisfied	Neutral	Dissatisfied	Very Dissatisfied
Overall satisfaction	☐	☐	☐	☐	☐

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

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

Please tell us how we can improve this book:

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

When you send comments to IBM, you grant IBM a nonexclusive right to use or distribute your comments in any way it believes appropriate without incurring any obligation to you.

Name

Address

Company or Organization

Phone No.

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

SC26-7290-00

Cut or Fold
Along Line

Fold and Tape **Please do not staple** Fold and Tape

NO POSTAGE
NECESSARY
IF MAILED IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST-CLASS MAIL PERMIT NO. 40 ARMONK, NEW YORK

POSTAGE WILL BE PAID BY ADDRESSEE

International Business Machines Corporation
RCF Processing Department
G26/050
5600 Cottle Road
San Jose, CA 95193-0001

Fold and Tape **Please do not staple** Fold and Tape

Cut or Fold
Along Line

Printed in the United States of America
on recycled paper containing 10%
recovered post-consumer fiber.

SC26-7290-00

Spine information:

Fibre Channel Storage Manager

Fibre Channel Storage Manager, Installation and User's Guide for AIX Operating System