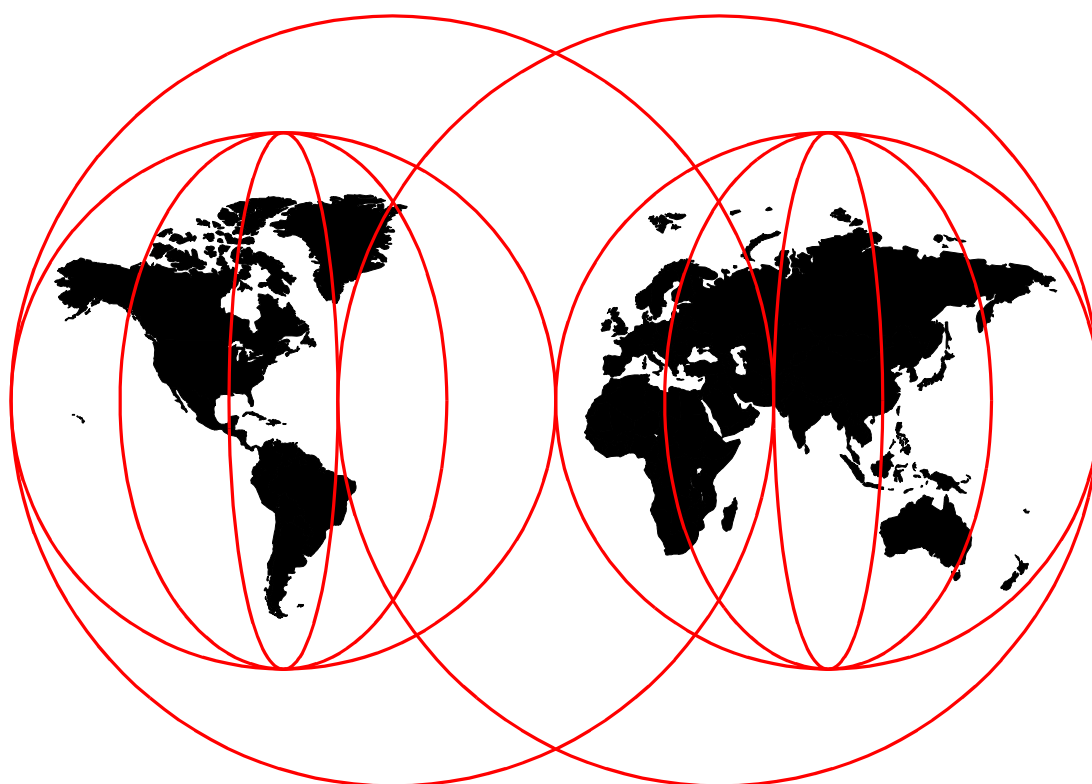


Configuring and Implementing the IBM Fibre Channel RAID Storage Server

Mark Blunden, Guillermo Duffoo, Seth Mason, Ling Pong



International Technical Support Organization

<http://www.redbooks.ibm.com>



International Technical Support Organization

SG24-5414-00

**Configuring and Implementing the
IBM Fibre Channel Storage Server**

April 1999

Take Note!

Before using this information and the product it supports, be sure to read the general information in Appendix C, "Special Notices" on page 123.

First Edition (April 1999)

This edition applies to the IBM Fibre Channel RAID Storage Server. See the PUBLICATIONS section of the IBM Programming Announcement for the IBM Fibre Channel RAID Storage Server for more information about what publications are considered to be product documentation.

Comments may be addressed to:
IBM Corporation, International Technical Support Organization
Dept. QXXE Building 80-E2
650 Harry Road
San Jose, California 95120-6099

When you send information to IBM, you grant IBM a non-exclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© Copyright International Business Machines Corporation 1999. All rights reserved

Note to U.S Government Users - Documentation related to restricted rights - Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract with IBM Corp.

Contents

Figures	vii
Tables	ix
Preface	xi
The Team That Wrote This Redbook	xi
Comments Welcome	xii

Part 1. Product Components 1

Chapter 1. Overview	3
1.1 Introduction	3
1.2 Product Description	4
1.3 Architecture	5
1.4 Components	5
1.4.1 Hardware Components	5
1.4.2 Physical Requirements	12
1.4.3 Software Components	14
1.5 Supported Platforms	15
1.5.1 Windows NT	15
1.5.2 Hewlett Packard	16
1.5.3 Sun Enterprise	16
1.5.4 IBM RS/6000	16
1.5.5 Systems Attachment Limitations	17
1.5.6 Summary of Requirements	17
Chapter 2. Positioning	19
2.1 Current IBM Storage Products	19
2.1.1 IBM 7204 External Disk Drive	19
2.1.2 IBM 7131 SCSI or Serial Disk System	20
2.1.3 IBM 7133 Serial Disk System	20
2.1.4 IBM Versatile Storage Server	20
2.2 Storage Area Network	20
IBM Seascapes Storage Area Networks	21
2.3 SCSI and SSA	22
2.3.1 SCSI Technology	22
2.3.2 SSA	23
2.3.3 RAID Technology	24
Chapter 3. Fibre Channel	29
3.1 Fibre Channel Layers	29
3.1.1 Physical and Signaling Layers	29
3.1.2 Upper Layers	30
3.2 Fibre Channel Topologies	31
3.2.1 Point-to-Point	31
3.2.2 Fabric	31
3.2.3 Loop	32
3.3 Fibre Channel Hardware	32
3.3.1 Connectors	32
3.3.2 Concentrators	33

Chapter 4. Performance Considerations	35
4.1 Overview	35
4.1.1 Test Configuration	36
4.1.2 Measurement Results	36
4.1.3 Additional Considerations	40
4.2 Conclusions	41
Chapter 5. Application Considerations	43
5.1 General Considerations	43
5.1.1 Availability	43
5.1.2 Performance	43
5.1.3 Cost Benefit	44
5.2 Database Management System (DBMS) Specific Considerations	44
5.2.1 Common Considerations	44
5.2.2 DBMS Functional Characteristics	45
5.3 OLTP Specific Considerations	45
5.4 Data Warehousing Applications	45
5.5 SAP, BAAN or Other General Business Applications	46

Part 2. Implementation Processes 47

Chapter 6. Installation	49
6.1 Components	49
6.1.1 2101-100 IBM Seascope Solutions Rack	49
6.1.2 2102-F00 FC-RSS—Control Unit	49
6.1.3 2102-D00 Fibre Channel Storage Expandable Unit	49
6.1.4 2103-H07 Fibre Channel Storage Hub	49
6.1.5 FC GBICs	50
6.1.6 Fibre Cables	50
6.2 Configurations	50
6.3 Cabling	52
6.3.1 Cabling the Rack	52
6.3.2 Cabling the Hub	52
6.3.3 Cabling the Control Unit	53
6.3.4 Cabling the FC Storage Unit	53
6.4 Host System Requirements and Installation Procedures	61
6.4.1 Windows NT System	61
6.4.2 HP - UX System	67
6.4.3 Sun Solaris System	69
6.4.4 RS/6000 System	71
Chapter 7. Configuration	75
7.1 Solaris, HP-UX and Windows NT	75
7.1.1 Storage Allocation	75
7.1.2 Modifying LUN and Drive Group Attributes	80
7.1.3 Networking	84
7.1.4 Expansion	93
7.2 AIX	95
7.2.1 Preparation	95
7.2.2 Storage Allocation	96
7.2.3 Modifying Array Attributes	97
7.2.4 Expansion	97

Chapter 8. Management	99
8.1 Overview and Philosophy	99
8.2 StorWatch Fibre Channel RAID Specialist	100
8.3 Command Line	100
8.4 General Functions	100
8.4.1 Initial Setup	101
8.4.2 Operations	101
8.4.3 Monitoring	102
8.4.4 Modifications	102
8.4.5 Maintenance	103
8.4.6 Migration Considerations	103
8.4.7 Clustering Considerations	104
Chapter 9. Performance Tuning	105
9.1 Fibre Channel	105
9.1.1 Basic Configurations	105
9.1.2 Advanced Configurations	105
9.2 Cache	106
9.2.1 Read Ahead Prefetching	106
9.2.2 Cache Flushing	108
9.2.3 Cache Block Size	109
9.3 Disk Groups and LUNs	109
9.3.1 Choosing Disks for a Drive Group	109
9.3.2 Balancing LUNs	110
9.3.3 LUN Segment Sizes	111
Chapter 10. Backup, Recovery and Availability	113
10.1 Backup	113
10.2 Types of Failures	113
10.2.1 Fibre Connection Failures	113
10.2.2 Controller Failures	114
10.2.3 Cache Failures	114
10.2.4 Disk Drive Failures	115
10.2.5 Power Failures	115
10.3 Data Integrity	115
10.4 System Availability	115
10.4.1 Planning for Maximum Availability	115
<hr/>	
Part 3. Appendices	117
Appendix A. SMIT Menus	119
Appendix B. Support	121
B.1 Installation Support	121
B.2 Field Support	121
Appendix C. Special Notices	123
Appendix D. Related Publications	125
D.1 International Technical Support Organization Publications	125
D.2 Redbooks on CD-ROMs	125
D.3 Other Publications	125

How to Get ITSO Redbooks	127
How IBM Employees Can Get ITSO Redbooks	127
How Customers Can Get ITSO Redbooks	128
IBM Redbook Order Form	129
List of Abbreviations	131
Index	133
ITSO Redbook Evaluation	137

Figures

1. The Team: Guillermo, Seth, Ling and Mark	xii
2. IBM Fibre Channel RAID Storage Server	3
3. IBM Fibre Channel RAID Storage Server Architecture	5
4. FC-RSS Controller Module	6
5. Fibre Channel Expandable Storage Unit	8
6. Rear View of IBM Fibre Channel Expandable Storage Unit.	8
7. IBM Seascope Solutions Rack.	10
8. Fibre Channel Storage Hub	11
9. PCI FC Adapter	12
10. SAN Architecture	22
11. SSA Loop Configuration.	24
12. RAID-0	25
13. RAID-1	26
14. RAID-3	26
15. RAID-5	27
16. RAID-10	28
17. Fibre Channel Layers.	29
18. Fibre Channel Topologies	31
19. Flexible Use of Gigabit Copper or Fibre Optics	33
20. Sequential Throughput for 256 KB Records	37
21. Cache Hit Performance for 256 KB Records	37
22. Random I/O, RAID-5 for 256 KB Records	38
23. Sequential Throughput for 4 KB Records	39
24. Cache Hit Performance for 4 KB Records	39
25. Random Access Throughputs for 4 KB Records	40
26. Fibre Hub	50
27. Single Host.	51
28. Independent Host.	51
29. Multi Host (One FC-RSS).	51
30. Control Unit (Rear View)	53
31. Fibre Channel Storage Unit (Rear view)	54
32. One Storage Drawer	55
33. Two Storage Drawers	56
34. Three Storage Drawers	57
35. Four Storage Drawers	58
36. Five Storage Drawers	59
37. Six Storage Drawers	60
38. Create LUN, Options	76
39. Creating a LUN, Drive Selection	77
40. Configuration, One LUN Created	78
41. Creating a LUN from an Existing Drive Group	79
42. Creating a Hot Spare - Options	80
43. Ld -a Output Under Windows NT	84
44. Add Module Screen	89
45. Add Name for RAID Module.	89
46. Two-Channel Drawer	94
47. Drawer (Rear View)	94
48. Single-Channel Drawer	95
49. GUI Function Selection Icons.	100
50. Configuration Icon	101

51. Status Monitoring Icon	102
52. Maintenance and Tuning Icon	102
53. Maintenance and Tuning Icon	103
54. Example of an Advanced Configuration	106
55. SMIT Menus	119

Tables

1. Power Requirements	12
2. Environmental Requirements	12
3. Seascope Rack Specifications	13
4. Support Matrix	17
5. LEDs on HUB	52
6. Summary of Tasks for Configuring Controllers on the Network	85
7. Sample Table for Recording Module and Controller Information	85
8. Table for Recording Module and Controller Information	86
9. Required And Optional Entries for Setting up the BOOTP or DHCP Server . .	87
10. Contents of an SNMP Trap Message	90
11. Changeable Parameters in rparams File	92
12. AIX and HP-UX/Solaris/Windows NT Terminology	95

Preface

This redbook describes a new, compact, economical, open storage disk subsystem offering designed to address market requirements in the lower-GB segment.

Technology, function, design, and configuration options are described so that effective cost, availability, and performance criteria can be balanced with clear technical evaluation and market positioning. Comparisons with other open storage subsystem offerings are described, and the various supported server environments and product capabilities are contrasted to ensure that customer expectations are correctly set when multiple-choice configurations are offered.

This redbook is intended to help customer management, storage administrators, and technical support staff understand, plan, implement and administer the new IBM Fibre Channel RAID Storage Server.

The Team That Wrote This Redbook

This redbook was produced by a team of specialists from around the world working at the International Technical Support Organization San Jose Center. Figure 1 on page xii shows the team that wrote this book.

Mark Blunden is the project leader for Open Systems Storage at the International Technical Support Organization, San Jose Center. He has coauthored six previous redbooks and comanged two other redbooks. Mark also teaches IBM classes worldwide on all areas of storage. He has worked for IBM for 18 years in many areas of the IT business. Before joining the ITSO in 1998, Mark worked in Sydney, Australia, as an Advisory Storage Specialist.

Guillermo Duffoo is a Staff Technical Support Specialist for PSG Products in Peru. His areas of expertise include Netfinity Servers and PC Servers, Windows NT, Netware and OS/2. He also is on staff in the PC Institute, teaching Netfinity courses. This is Guillermo's first ITSO Redbook.

Seth Mason is a Product Engineer and a member of the Center of Competency for Open Systems Storage located in San Jose. He holds a bachelor's degree in Computer Engineering from Auburn University. His areas of expertise include Fibre Channel, Linux, AIX, Solaris and Windows NT.

Ling Pong is a Certified I/T Specialist in Storage in the USA. He has over 26 years of experience in IBM in the I/T industry as a systems programmer, a systems engineer, an Image and Workflow consultant, and a program manager. He has been a member of the Storage Systems Advanced Technical Support Center since 1997.

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

John Murtagh, IBM Dallas

Kevin Smith, IBM San Jose

Richard Chelberg, IBM San Jose

Bill Ogroski, IBM Tucson

David Watts, IBM Cary

Marcee Cress, IBM San Jose

Dan Braden, IBM Dallas

David Muscarella, IBM Rochester

Lon Cowden, IBM Rochester

Barry Mellish, IBM UK

Comments Welcome

Your comments are important to us!

We want our redbooks to be as helpful as possible. Please send us your comments about this or other redbooks in one of the following ways:

- Fax the evaluation form found in "ITSO Redbook Evaluation" on page 137 to the fax number shown on the form.
- Use the electronic evaluation form found on the Redbooks Web sites:

For Internet users

<http://www.redbooks.ibm.com>

For IBM Intranet users

<http://w3.itso.ibm.com>

- Send us a note at the following address:

redbook@us.ibm.com



Figure 1. The Team: Guillermo, Seth, Ling and Mark

Part 1. Product Components

Chapter 1. Overview

In this chapter we present a high-level overview of the IBM Fibre Channel RAID Storage Subsystem (FC-RSS) and its supporting software. The FC-RSS consists of dual controllers, up to six drawers of disk drives, and optionally, one or two Fibre Channel hubs mounted in an IBM Seascope Solutions Rack (see Figure 2).



Figure 2. IBM Fibre Channel RAID Storage Server

1.1 Introduction

The FC-RSS is an entry-level to mid-range solution for customers with a need for a multiserver storage subsystem that has a capacity of 18 gigabytes to one terabyte of RAID protected storage in a single rack. It provides mission-critical, easy-to-manage high availability storage with redundancy in the controllers, RAID disk storage, power supplies, and cooling units. Customers with storage needs that are supported across multiple server platforms and that require Fibre Channel Arbitrated Loop (FC-AL) attachment should consider the FC-RSS. The IBM 2105 Versatile Storage Server and the IBM 7133 Serial Disk Systems should be considered for large enterprise-wide heterogeneous platforms, or for RS/6000 clustered environments where appropriate adapters may offer the right functions and storage with price performance characteristics that better meet customers' requirements. The FC-RSS provides users a Fibre Channel based storage subsystem that allows centralization of disk storage, systems management, and a facility for data sharing in a homogeneous system environment.

The FC-RSS offers a high-performance, high-availability storage solution for servers running Microsoft Windows NT, HP-UX, Sun Solaris, or AIX. Its industry standard Fibre Channel host attachment provides configuration flexibility for distances of up to 10 kilometers. It offers enhanced data availability through the implementation of various RAID techniques, disk drive module sparing, failover protection and dual paths to data.

1.2 Product Description

The FC-RSS consists of a 2101-100 rack, the 2102-F10 controller, one to six 2102-D00 drawers, and optionally one or two 2107-H07 hubs. The heart of the FC-RSS is the 2102-F10, which is a configuration with dual active controllers. The FC-RSS offers a choice of RAID protection levels RAID-0, RAID-1, RAID-3, RAID-5 and RAID-0+1. These RAID levels are described in detail in Chapter 2.3.3, "RAID Technology" on page 24. Each of the dual controllers supports 128MB or 256 MB of battery-backed read/write cache for a total of 512MB per server. With the write cache mirroring function, an additional level of system availability is provided, albeit at a cost to performance. The FC-RSS supports up to six drawers, each of which can hold four to ten disk drive modules (DDMs). Thus, a fully configured FC-RSS provides up to one terabyte (TB) of disk storage (the actual amount of user accessible storage will vary, as approximately 40MB of space is reserved on each drive for system use). This is ideal for business-critical applications where continuous availability and high performance are required.

The IBM Fibre Channel Expandable Storage Unit (hereafter referred to as a *drawer*) supports IBM's high performance, reliable Ultrastar disk drives in sizes of 4.5, 9.1, and 18.2 GB. Each drawer comes with a minimum of four DDMs of the same size, and can be populated to full capacity (10 drives) with DDMs of any size. For ease of management, we recommend that all DDMs within an FC-RSS be of the same size.

The new IBM Fibre Channel Storage Hub (hereafter referred to as a *hub*) provides flexibility when configuring multiple host attachments and storage servers, and provides support for distances beyond 500 meters. It provides connectivity for seven FC-AL ports, a base of four shortwave multimode GBICs with capability of an additional three GBIC ports, longwave or shortwave. Shortwave multimode fibre technology allows cable lengths of up to 500 meters. Longwave single mode fibre allows cable lengths of up to ten kilometers. Host attachments must be shortwave GBIC connections.

Included with the FC-RSS is the Fibre Channel Storage Manager software (for a selected operating system), and optionally the StorWatch FibreChannel RAID Specialist. These software tools provide ease of setup and management of the FC-RSS, and along with the FC Specialist, they allow multiple FC-RSS servers to be managed over a networked environment from a single workstation.

The FC-RSS offers:

- Concurrent multiple homogeneous server attachments
- 100 MB Fibre Channel support with remote data storage of up to 10 kilometers
- A high-availability solution with redundant controllers, power supplies, cooling units, offering protection against loss of data availability
- High performance using high-speed IBM disk drives with Ultra2 SCSI interfaces, predictive cache algorithms, and high-speed RAID controllers
- StorWatch Fibre Channel RAID Specialist and Fibre Channel Storage software that provides for ease of Storage Server configuration, dynamic changes, performance monitoring, error reporting, and management operations
- A solution ready for Storage Area Networks utilizing Fibre Channel attachment

This combination of leading-edge hardware and software technologies in the FC-RSS delivers high performance, investment protection, and industry-leading scalability.

1.3 Architecture

Figure 3 on page 5 shows the architecture of the FC-RSS in schematic form.

Fibre Channel RAID Storage Serve Architecture

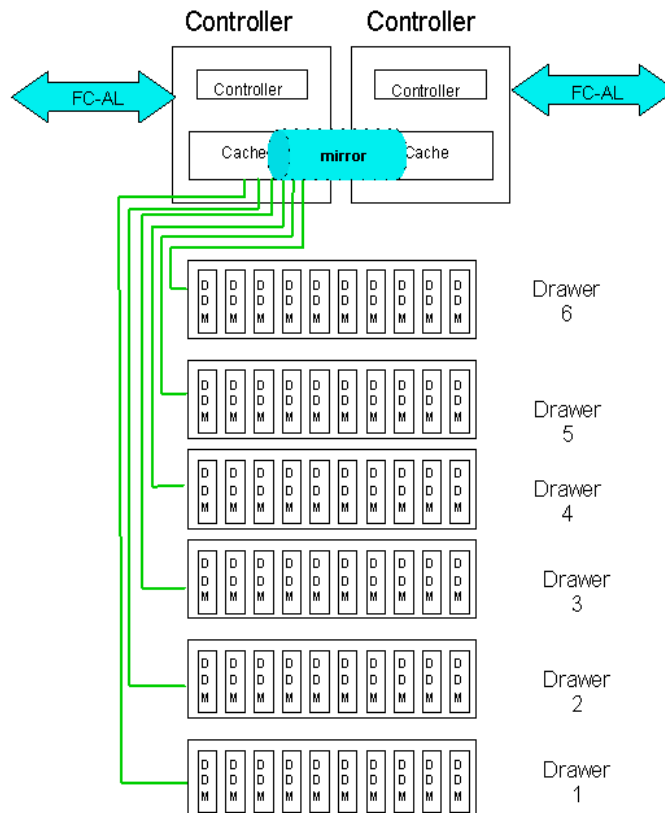


Figure 3. IBM Fibre Channel RAID Storage Server Architecture

1.4 Components

This section lists the components of the FC-RSS — the controller module, the storage module, the rack, and the optional hub.

1.4.1 Hardware Components

This section describes the modules of the FC-RSS.

1.4.1.1 FC-RSS (2102 Model F10)

Figure 4 on page 6 shows the individual components within the control unit module.

Control Unit Components

- 1) FC-RSS Controllers
- 2) Fans
- 3) Battery
- 4) Power Supplies
- 5) Fans

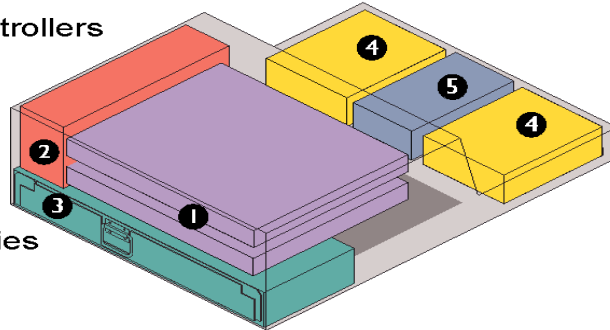


Figure 4. FC-RSS Controller Module

The FC-RSS is a rack-mountable drawer that requires 4U space. It can be installed in a 2101-100 IBM IBM Seascapes Solutions Rack, or equivalent industry-standard 19-inch rack. Up to six 2102-D00 Expandable Storage units can be attached to the FC-RSS, providing capacities from 18 GB up to one (1.092) Terabytes of disk storage. Its dual active controllers offer a variety of choices in RAID protection (0, 1, 3, 5, and 0+1) with global hot sparing. RAID-1 arrays can range from a single mirrored pair to 15 mirrored pairs of drives, RAID-5 arrays can span from 2+p to 19+p configurations, and RAID-0+1 can span up to 20 disk drives. Each controller supports up to 256 MB of battery-backed read/write cache providing both high performance and high data integrity. In addition, high-availability features like redundant hot pluggable power supplies and cooling units are standard.

The 2102-F10 base configuration consists of:

- Dual processors with failover protection
- 256 (2 x 128) MB read/write cache
- Two FC-AL ports (upper interface)
- Six Ultra2 SCSI LVD ports (lower interface)
- Support for up to six disk drawers (2102-D00)
- RAID 0, 1, 3, 5, and 0+1 support
- Dual power supplies and cooling units
- NT, HP, Sun, and RS/6000 host support
- Rack mount hardware

Features:

FC-RSS Optional Remote Service

The FC-RSS includes a set of optional features designed to provide rapid diagnostic support capability. You are responsible for providing an analog

phone line next to the FC-RSS enclosure to enable remote diagnostic support. The Remote Support Modem/Switch provides the modem / switch (which supports up to seven FC-RSS), cables, and connectors required to attach the first subsystem. The Remote Support Attachment feature provides additional cables and connectors needed to attach the FC-RSS to an existing modem/switch. Each FC-RSS requires two ports (one per controller) to enable the optional remote service capability. If a modem/switch has at least two available ports, and is within 150 feet of the FC-RSS, the Remote Support Attachment can be used to attach to those available switch ports.

The modem/switch is the same as used with the IBM Versatile Storage Server (2105-B09), the IBM Virtual Tape Server (3494-B18), and the IBM Magstar 3590 Controller (3590-A50).

Remote Service:

- Allows the service provider to initiate a call to the FC-RSS processor and remotely analyze potential or actual failures and determine the type of failure and proper fix.
- Allows correction of many types of microcode problems remotely. When a problem is reported, fixes can be created and applied from the remote location.

Cache :

The FC-RSS base system includes 256 MB of read/write cache. Additional cache can be added to the storage server. You can select additional cache with feature number 4001, increasing the total cache of 512 MB for the two controllers.

Host Attachment Cables:

Fibre optic cables are required for the 2102-F10 to connect to the host system adapters, or the 2103-H07 Hub's short wave Gigabit Interface Converter (GBIC) ports. FC cables can be customer supplied or ordered as needed. Each controller on the 2102-F10 provides one host connection (FC-AL) port. The 2102-F10, with two controllers, thus has two host connections (ports). Feature numbers 9701 and 9702 provide a fibre optical 50.0 micron multimode fibre cable, either 5 meters or 25 meters in length. Customer-provided cables, 50.0 micron multimode fibre cables to 500 meters, or 62.5 micron multimode fibre up to 175 meters in length, can be used. A duplex SC connector type is required.

1.4.1.2 IBM Fibre Channel Expandable Storage Unit (2102 Model D00)

The IBM Expandable Storage Unit is a rack-mountable drawer (3U space) that can be housed in the Seascapes Solutions Rack, or equivalent industry-standard 19-inch rack, with the 2102-F10. It supports up to ten hot swappable IBM Ultra2 SCSI (LVD) drives in capacities of 4.5 GB, 9.1 GB, or 18.2 GB per unit. At least four drives of the same capacity are required per drawer. Up to six Expandable Storage Units can be attached to each FC-RSS. The dual power supplies and cooling units provide redundancy, allow for installation of separate power boundaries (requires Feature Number 6001), and are hot pluggable, allowing for continuous operation when replacement is required. Figure 5 on page 8 shows the individual components within the storage unit module.

1. Drive module (DDM)
2. Power supply
3. Fan
4. Environmental Service Module (ESM)

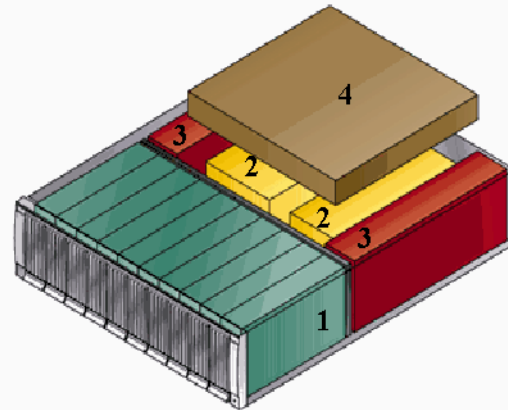


Figure 5. Fibre Channel Expandable Storage Unit

The Expandable Storage Unit contains a single backplane that can be configurable into one or two SCSI buses. When less than six drawers are attached to the FC-RSS, some of the drawers can be configured using the dual bus option, providing for shorter SCSI disk drive strings. The single or dual bus configuration is determined by the setting of a DIP switch at the rear of the drawer.

There are two standard VHDI .8mm 16 bit Ultra2 SCSI ports on the back of the unit as shown in figure Figure 6 on page 8.

There are two possible configurations for the storage unit, two independent 5-drive buses or one 10-drive bus. The two-bus configuration is selected when the dip switch in the switch block is set to on (up). Conversely, when the dip switch is off (down), the storage unit is set to the single 10-drive bus configuration (see Figure 6).

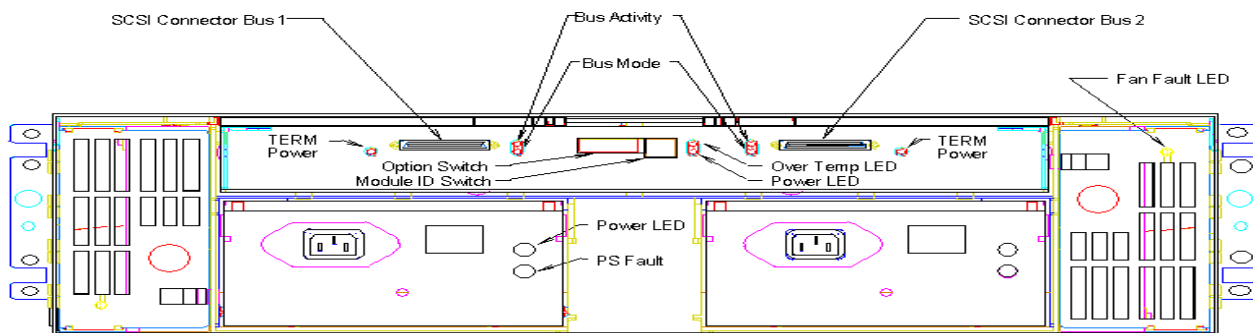


Figure 6. Rear View of IBM Fibre Channel Expandable Storage Unit

The disk drive carriers are hot pluggable. Additional disk drives of any capacity, up to the drawer maximum of ten, can be installed without interrupting operations.

When additional Expandable Storage Unit drawers are installed, or SCSI bus configurations are changed, the FC-RSS must be interrupted, and procedures defined in the *User's Guide* (SC26-7288) accurately followed to assure that there

is no loss of data. The Expandable Storage Units must be installed in the same rack as the FC-RSS.

When feature number 9201 is specified, the plant will install the Expandable Unit in the 2101-100 rack, and depending upon the total number of drawers, you can install one or two Ultra2 SCSI cables per drawer. When feature number 9202 is specified, for field installation, two 2-meter Ultra2 SCSI cables will be supplied with the drawer. The actual number of cables used per 2101-D00 Expandable Storage Unit (one or two) will depend on the total number of 2102-D00 drawers attaching to the FC-RSS.

This drawer contains two redundant hot pluggable power units. A Y-styled power cord is supplied to connect the two power units to the rack PDU. If feature number 6001 is specified, then two inline power cords are provided instead of the Y-styled power cord, so that each drawer power unit is attached independently to different PDUs.

1.4.1.3 IBM Seascope Solutions Rack (2101 Model 100)

The IBM Seascope Solutions Rack is an industry standard 19-inch rack that holds the FC-RSS, the Expandable Storage Unit(s), and the FC Storage Hub(s). It provides a power distribution unit (PDU) for all of the attached drawer and the option for a second PDU with separate power cords enabling each of the drawers to operate on two different power boundaries. Power cords are provided based PDU selected and specified country feature number.

Figure 7 on page 10 shows the outside of the rack, which houses all of the components that make up the FC-RSS.

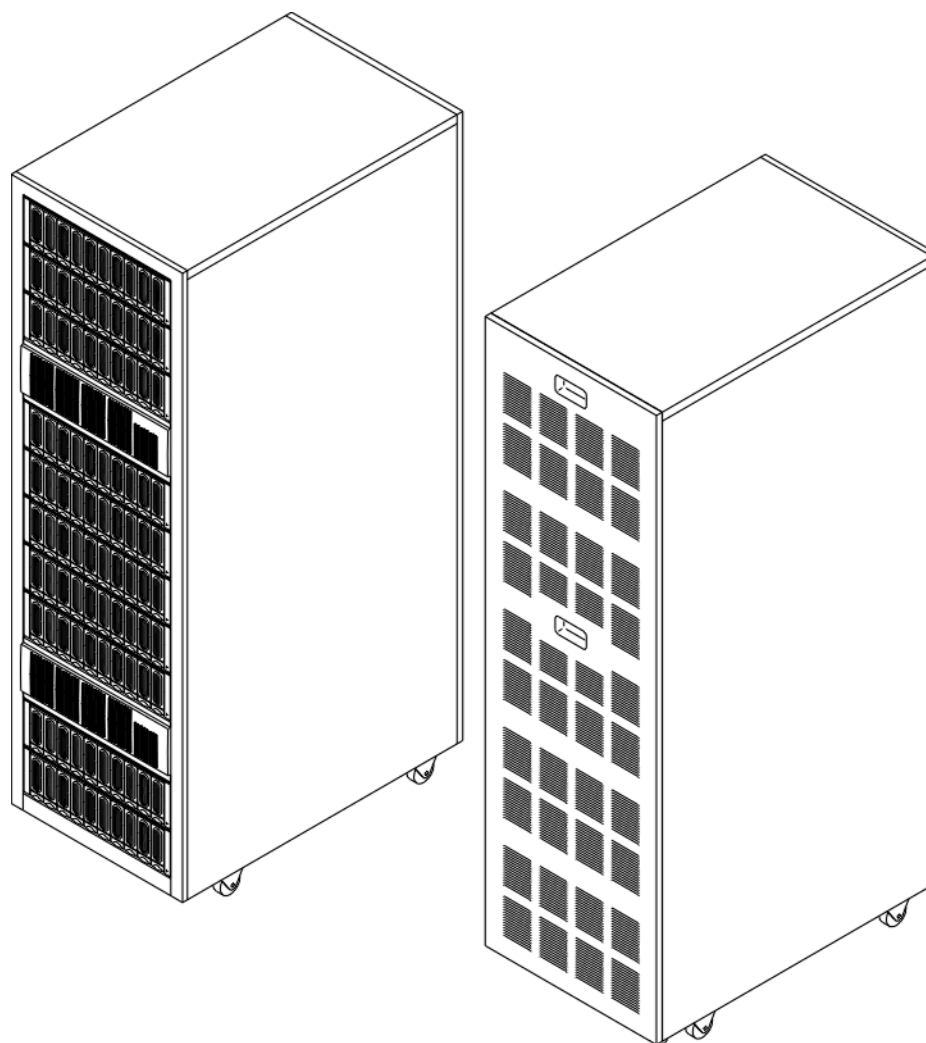


Figure 7. IBM Seascope Solutions Rack

The 2101-100 provides 32U of mounting space and includes front and rear access doors.

The 2101-100 contains the following features:

- Dual Power Distribution Unit: A second power distribution unit available to provide higher availability. Power cords are provided based on the PDU selected and specified country feature number.
- Disk Storage: Up to six IBM Expandable Storage Units can be installed in an IBM Seascope Solutions Rack.
- Hubs: Up to two IBM Fibre Channel Storage Hubs can be installed in the IBM Seascope Solutions Rack.

1.4.1.4 IBM Fibre Channel Storage Hub (2103 Model H07)

The IBM Fibre Channel Storage Hub can be used to enable the configuration of FC-AL loops, and to add greater distances to the FC-AL topology. The hub provides four shortwave optical Gigabit Interface Converter (GBIC) ports, and the option to add up to three additional longwave or shortwave optical GBIC ports.

The short wave GBIC ports allow fibre cable connection of up to 500 meters to the host based FC-AL initiator adapter, or to the FC-AL port on the FC-RSS. Either shortwave or longwave GBIC ports can be used to connect two hubs, extending the distance up to an additional 500 meters with the shortwave GBIC, or up to an additional 10 kilometers with the longwave GBIC. The hub comes in a rack mounted tray that requires 1U of rack space. If two hubs are ordered they will both be assembled in the same 1U rack mounted tray. An option (feature 9205) is available for a non-rack mounted, tabletop usage with an external country unique power cord. All ports in the hub are hot-pluggable.

Figure 8 on page 11 shows the hub and its fibre channel connections.



Figure 8. Fibre Channel Storage Hub

This unit contains a single power supply. A single Y-styled power cord is supplied when mounting one or two hubs at the plant in a 2101-100 rack. This power cord enables connecting the hub(s) to the rack PDU, and requires one outlet on the PDU. If feature number 6001 is specified, then one in-line power cord is provided instead of the Y-styled power cord, with each hub for connecting the hub to the PDU. If feature code 9205 is specified, indicating that the hub is not mounted within a rack and will be used externally, then the appropriate convenience cord is supplied, depending upon the country code (SC: 94xx) indicated.

1.4.1.5 Fibre Optic cables and GBICs

Fibre optic cables are required for the 2103-H07 hub to the host system adapters, or to another 2103-H07 hub. Fibre channel cables can be customer-supplied or ordered as needed. Each GBIC on the hub provides one FC-AL connection (port). The shortwave GBIC ports can use either feature number 9701 cables for 5 meters, feature number 9702 for 25 meters, or customer-provided cables (50.0 micron multimode fibre) for up to 500 meters. For the longwave GBIC ports, connecting two 2103-H07 hubs up to 10 Km, a customer-provided 9 micron single mode fibre cable must be used. A duplex SC connector type is required.

1.4.1.6 IBM PCI Fibre Channel Adapter

This PCI-to-Fibre-Channel connection allows access to the storage server. The data rate goes up to 100 MB/s or 1 Gbit. The PCI adapter is a 32/64 bit adapter card and is installed as a SCSI controller. The On-board Fast!Util configuration

and testing application is used to verify and find all devices in the FC-AL network. Figure 9 on page 12 shows the PCI FC adapter.

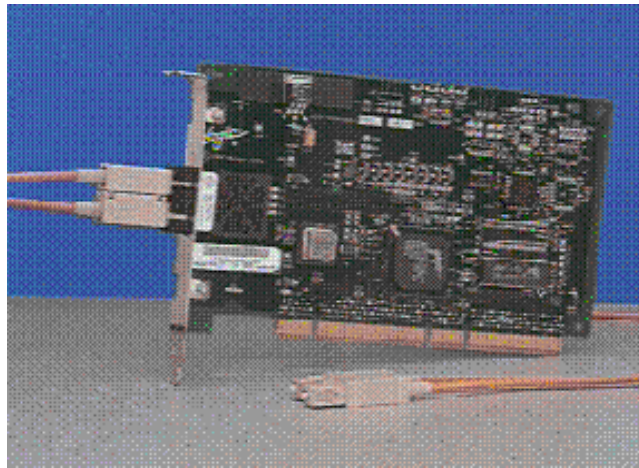


Figure 9. PCI FC Adapter

1.4.2 Physical Requirements

This section provides details of the physical requirements of the FC-RSS. It shows information about power, environments, and floor space.

1.4.2.1 Power Requirements

Table 1 on page 12 provides the power requirements for the FC-RSS.

Table 1. Power Requirements

Input Power Connection	50 Hz			60 Hz		
	Nominal	Minimum	Maximum	Nominal	Minimum	Maximum
Single-Phase Line to Neutral	100	90	107	100	90	107
	127	114	136	120	104	127
	220	198	235			
	230	207	246			
	240	216	264			
	200	180	214	200	180	214
Single-Phase Line to Line				208	180	220
				240	208	264

1.4.2.2 Common Environmental Requirements

Table 2 on page 12 provides the environmental requirements for the FC-RSS.

Table 2. Environmental Requirements

Element	Unit of Measure	Requirement
Air Flow		Air flows from front to back

Element	Unit of Measure	Requirement
Altitude	Operating Range	30.5 m (100 ft) to 3,000 m (9,840 ft)
	Storage Range	30.5 m (100 ft) to 3,000 m (9,840 ft)
	Transit Range	30.5 m (100 ft) to 12,000 m (40,000 ft)
Temperature	Optimum	22° C (72° F)
	Operating Range	10° C to 40° C (50° F to 104° F)
	Storage Rang	-10° C to 50° C (14° F to 122° F)
	Transit Range	-40° C to 60° C (-40° F to 140° F)
Temperature Change (maximum allowed)	Operating Range	10° C (18° F) per hour
	Storage Rang	15° C (27° F) per hour
	Transit Range	20° C (36° F) per hour
Relative Humidity (no condensation)	Operating Range	20% to 80%
	Storage Rang	10% to 90%
	Transit Range	5% to 95%
	Maximum Dew Point	28° C (79° F)
	Maximum Humidity	10% per hour
If you plan to operate the controller module at altitudes between 1000m and 3000 m (3280 ft and 9850 ft), you must lower the environmental temperature 1.7° C (3.3° F) for every 1000 m (3280 ft) above sea level.		

1.4.2.3 Specifications

Table 3 on page 13 provides the specifications for the FC-RSS, detailing the physical characteristics and requirements.

Table 3. Seascope Rack Specifications

Specification	Model	Model
	2101-100	2102-F10
Width	650 mm 25.5 in	445 mm 17.5 in
Depth	1030 mm 40.5 in	635 mm 25.0 in
Height	1580 mm 62.0 in	175 mm 6.88 in
Weight	160 kg (empty) 352 lb (empty)	36 kg 79 lb
	2102-D00	2103-H07
Width	480 mm 18.9 in	219 mm 8.6 in

Specification	Model	Model
Depth	575 mm 22.6 in	367 mm 14.4 in
Height	132 mm 5.2 in	44 mm 1.7 in
Weight	32/42 kg (min/max) 69/92 lb(min.max)	4 kg 8 lb
Temperature	10 - 40 degrees C 50 - 140 degrees F	
Relative Humidity (non-condensing)	8 to 80 %	
Wet bulb temperature	27 C 80.6 F	
Exhaust Capacity	xxx KCal/hr (max) 9250 BTU/hr (max)	
Noise level	6.6 bels	
Power consumption	Standard Configuration: 0.6 kVA Maximum Configuration: 2.2 kVA System Limit: 4.8 kVA Operating Voltage: 200 to 240 V ac Frequency: 50 or 60 Hz	

1.4.3 Software Components

The FC-RSS software provides for configuration setup and subsystem management.

The IBM Fibre Channel Storage Manager provides platform specific automatic failover management. FC-RSS systems administration and the IBM Storwatch Fibre Channel Raid Specialist allows you to perform systems administration functions as well as manage multiple FC-RSSs from a single network attached workstation.

The dynamic FC-RSS management software includes the following features:

- Logical unit (LUN) creation and deletion
- Dynamic capacity expansion
- RAID level migration
- Segment size changes
- Defragmentation
- Status and performance monitoring
- Programmable error notification
- Maintenance and tuning utilities
- Step-by-step recovery procedures

1.4.3.1 IBM Fibre Channel Storage Manager

The FC Storage Manager provides configuration setup and management (including LUN creation and deletion, dynamic capacity expansion, RAID level migration, and segment size changes and defragmentation), status and performance monitoring programmable error notification (log, e-mail, and SNMP traps), maintenance and tuning utilities, and step-by-step recovery procedures when errors occur. The operating system specific FC Storage Manager software, which resides on the host attached to the FC RAID Storage Server, also provides automatic I/O path failover when one controller of a dual controller pair fails.

The FC Storage Manager provides two interfaces, a GUI interface and a command line interface. The functions provided by the GUI interface are also accessible through the StorWatch FC RAID Specialist described below. The FC Storage Manager command line interface has additional functions such as the **raidutil**, which allows you to set cache parameters like flush or destage thresholds or **rdacutil**, which defines LUN load balancing.

1.4.3.2 IBM StorWatch Fibre Channel RAID Specialist

The StorWatch Fibre Channel RAID Specialist is a network-based integrated storage management tool that lets the storage administrator dynamically configure, monitor, and manage multiple FC RAID Storage Servers from a single Windows 95 or NT workstation. It provides the same functions as the FC Storage Manager GUI interface, but extends it to multiple platforms. This capability requires that two IP addresses be assigned to each FC-RSS: one for each of the controllers; and two ethernet connections per , each with a 15-pin AUI connector. The ethernet network connection allows a system administrator, at a single location, to configure, monitor and management multiple FC-RSS Servers throughout an enterprise.

1.5 Supported Platforms

The FC-RSS supports these IBM and non-IBM systems:

- Intel-based servers running Windows NT Server 4.0
- Sun Microsystems SPARCserver and Enterprise servers with Sun Solaris 2.5.1 and 2.6
- Hewlett Packard 9000 K Series server with HP/UX V10.20, 11.0
- RS/6000 with AIX 4.3.2

The systems supported at the time of announcement are listed below. Details on models and operating system releases are available on the Web at URL:

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

1.5.1 Windows NT

Intel-based servers with Microsoft Windows NT Server 4.0 (Server or Enterprise Editions) with Service Pack 4 or later level or maintenance and at least 64MB of system memory.

1.5.1.1 IBM Netfinity

Netfinity 5000, 5500, and 7000 systems with the ,QLogic QLA2100F Fibre Channel PCI Adapter, or IBM Netfinity Fibre Channel PCI Adapter(01K7297).

Both the QLogic and the Netfinity host adapters require firmware level 1.35 and device driver level 1.21.02.

1.5.1.2 Compaq

Proliant 3000 and above with the QLogic QLA2100F Fibre Channel PCI Adapter.

1.5.1.3 Hewlett Packard

HP LD Pro, LH, LX Series and Netervers with the QLogic QLA2100F Fibre Channel PCI Adapter.

1.5.1.4 Other Servers

For details, check the web at URL:

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

1.5.2 Hewlett Packard

Hewlett Packard HP9000 K series UNIX servers with HP-UX 10.20 or HP-UX 11.0

Note: HP-UX 10.20 requires patch bundle
/cdrom/10.X/800/10.20/XSW800GR1020 or /cdrom/XSW800GR1020 depending
on the Extention Software Date Code, later versions use the shorter form.

HP Fibre Channel HSC Adapter A3404A

1.5.3 Sun Enterprise

Sun SPARCservers 1000/1000E, 2000/2000E, Ultra Enterprise 2, 150, 3000, 4000, 5000, 6000 with the following operating systems platforms:

- Solaris 2.5.1
 - Sun service recommended Patch Cluster dated 1/26/98 or later
 - SunOS 5.5.1 kernel patch 103640-20 or later
 - SunOS 5.5.1 kernel/drv/isp patch 103934-08 or later
 - SunOS 5.5.1 kernel/drv/sd patch 103622-09 or later
 - SunOS 5.5.1 kernel/drv format driver patch 104312-03 or later
- Solaris 2.6
 - SunOS 5.6 kernel update patch 105181-06 or later
 - SunOS 5.6 kernel/drv/isp patch 105600-06 or later
 - Note: in Ultra 2 to support multiple SCSI adapters, SunOS 5.6 SBus driver patch 105222-01 or later is needed

Jaycor Networks Inc. SBus Fibre adapter (FC64-1063-IBM) with version 2.2.0 of the Jaycor adapter driver is needed.

1.5.4 IBM RS/6000

RS/6000 S70, S7A with AIX Version 4.3.2

Note: Please check with AIX support for latest fixes required for Fibre Channel and FC-RSS support.

RS/6000 Gigabit Fibre Channel Adapter for PCI bus, feature code 6227

1.5.5 Systems Attachment Limitations

The following restrictions apply at the time of announcement:

- Only similar host operating systems can be attached to the FC-RSS.
- A maximum of two Windows NT initiators, and two FC-RSSs can be attached to a single FC-AL.
- A maximum of two supported Sun system initiators, and two FC-RSSs can be attached to a single FC-AL.
- A maximum of two supported HP system initiators, and two FC-RSSs can be attached to a single FC-AL.
- Only one supported RS/6000 system initiator and maximum of two FC-RSSs can be attached to a single FC-AL loop.
- Maximum of two FC Storage Hubs can be configured on a single loop for Windows NT, Sun, and HP systems.
- Long Wave GBIC connection of up to ten kilometers between Fibre Channel Storage Hubs is supported on Windows NT, Sun Solaris, and HP-UX configurations.

1.5.6 Summary of Requirements

Table 4 on page 17 shows a summary of support requirements.

Table 4. Support Matrix

Server	Models	Adapters	Operating Systems
Intel	See list at 1.5.1 on page 15	QLogic QLA21000F Fibre Channel PCI adapter IBM Netfinity Fibre Channel PCI adapter (01K7297)	Windows NT 4.0 Server and Enterprise Server Editions, Service Pac 4 or later
Sun	SPARCserver 1000/1000E 2000/2000E Enterprise 2,150,3000,4000, 5000 and 6000	Jaycor SBus Fibre (FC64-1063-IBM)	Solaris 2.5.1, 2.6
Hewlett Packard	9000 K-Series	HP Fibre Channel HSC Adapter A3404A	HP-UX 10.20,11.0
RS/6000	S70,S7A	RS/6000 Gigabit Fibre Channel Adapter for PCI bus (f/c 6227)	AIX 4.3.2 with the most current FC-RSS and FC maintenance
Other UNIX and NT based servers	Consult the Web for additional support	Consult the Web for additional support	Consult the Web for additional support

Consult the Web at URL:

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

Chapter 2. Positioning

The information explosion and the need for high-performance communications for server-to-storage and server-to-server have been the focus of much attention during the 1990s. Performance improvements in storage, along with the move to distributed architectures, have spawned increasingly complex data management and networking applications. The interconnect between systems and their input/output devices demands a new level of performance in reliability, speed, and distance.

Fibre Channel, a reliable technology, allows concurrent communications among workstations, mainframes, servers, and data storage systems, using SCSI and IP protocols. This technology provides interconnect systems for multiple topologies and heterogeneous platforms that can scale to a total system bandwidth on the order of a terabit per second. Fibre Channel delivers a new level of reliability and throughput.

Fibre Channel has been adopted by IBM as another technology for enterprise storage. It eliminates the distance, bandwidth, scalability, and reliability issues of SCSI. With today's increasingly sophisticated business environment, applications such as business intelligence, e-business, and collaborative groupware make reliable data storage more critical. More than ever, the competitive advantage for an enterprise hinges on fast access to data. This data must be stored and managed—not just for daily operations, but for all types of applications that must leverage data across multiple platforms. Serial storage is a key building block of the IBM Seascape Storage Enterprise Architecture, the blueprint for innovative storage solutions that help you manage and share information across your global enterprise. Seascape takes advantage of the best technology advancements as they become available to build comprehensive storage solutions tuned to your requirements.

The FC-RSS is an entry-level to mid-range solution for customers who have mission-critical applications and a group of servers of different platforms and who are looking for an enterprise solution with Fibre Channel and RAID technology that offers excellent performance, scalability, high capacity, high availability and easy tools to management.

2.1 Current IBM Storage Products

In many large enterprises, a movement is underway to consolidate UNIX and Windows NT servers into a single location under similar IT management. The benefits are clear: improved centralized storage management, easy maintenance, and great performance and capacity. Following is a list of the IBM products available today.

2.1.1 IBM 7204 External Disk Drive

The IBM 7204 is a small external disk enclosure. Depending on the model, it can house either a 9.1 GB or 18.2 GB single disk drive. It attaches to SCSI-2 F/W differential adapters in the RS/6000. Increased data capacity enables you to add new application programs and expand the number and size of data files that are stored online and readily accessible.

2.1.2 IBM 7131 SCSI or Serial Disk System

The IBM 7131 is a low cost, stand-alone SCSI (Model 105) or Serial (Model 405) storage tower that provides expansion storage for hard drives as well as tape or CD-ROM drives. In total, the tower provides up to 145 GB of storage. The media bays can be used to house an additional two non-hot-swappable disk drives, giving even more capacity.

2.1.3 IBM 7133 Serial Disk System

IBM 7133 Serial System provides high-performance, highly available storage for servers. Features are provided to add any combination of supported SSA disk drive modules providing up to 291 GB of disk storage capacity. Disk sharing is enabled by allowing multiple UNIX and Windows NT hosts to simultaneously attach to and share high-performance IBM 7133 Serial Disk Systems. Simplified storage management is available when the IBM StorWatch Serial Storage Expert is used with 7133 models.

2.1.4 IBM Versatile Storage Server

The IBM Versatile Storage Server (VSS) is a centralized shared storage solution for concurrently attached UNIX, Windows NT and AS/400 servers. Technologies like SSA and dual 4-way SMP processors provide excellent performance, data integrity and availability with storage server failover and RAID-5 protection. The VSS scales to over 4 TB of usable capacity. The VS Specialist interface enables users to monitor and manage storage resources from work, home, or on the road through a company-secured intranet.

2.2 Storage Area Network

Today data is scattered across the enterprise. It is stored on isolated islands of separated, distributed servers, each with different methods of storing, managing and protecting its piece of enterprise data. A Storage Area Network (SAN) is a shared storage repository attached to multiple host servers via a storage interface. It simplifies all the management, maintenance and monitoring actions to keep the data safety. All application data can be implemented by a wide range of interconnection technologies such as SCSI, FC-AL, SSA, ESCON and FICON.

The network behind the servers links one or more servers to one or more storage systems. Each storage system could be RAID, tape backup, tape library, CD-ROM library, or "just a bunch of disks" (JBOD). Fibre Channel networks are robust and resilient with these features:

- Shared storage among systems
- Scalable network
- High performance over great distances
- Robust data integrity and reliability
- Fast data access and backup

In a Fibre Channel network, legacy storage systems are interfaced using a Fibre Channel to SCSI bridge. TCP/IP is used for server to server and client/server communications. Storage networks operate with both SCSI and networking (IP) protocols. Servers and workstations use the Fibre Channel network for shared access to the same storage device or system. Fibre Channel products have

defined a new standard of performance, delivering a sustained bandwidth of over 97 MB/second for large file transfers, and tens of thousands of I/Os per second for business-critical database applications on a Gigabit link. The benefits that a SAN offers are:

- Higher availability: Clustered servers with alternate paths providing server redundancy
- Higher performance: Scalable storage capacity and fast throughput over great distances
- Centralized management: Ability to share, manage, and protect data centrally, eliminating individual maintenance
- Storage consolidation: A common resource pool that is platform independent
- Business recovery: Remote clusters and disk mirroring providing backup and journalizing

IBM Seascope Storage Area Networks

IBM Seascope SAN can interconnect multivendor, heterogeneous servers, and IBM Seascope Disk and Tape Storage Systems. It enables universal data access, centralized storage and data management as well as outboard data access, replication and migration. SAN also eases future enhancements. Seascope is a blueprint for comprehensive storage solutions optimized for a connected world. The Seascope architecture integrates leading technologies from IBM—including disk storage, tape storage, optical storage, powerful processors, and rich software function—to provide highly reliable, scalable, and versatile application-based storage solutions that span the range of servers from PCs to supercomputers.

The Seascope architecture is the first of its kind to address network storage challenges using powerful storage servers, a rich spectrum of software, and a vast array of "building blocks" that snap together to create total storage solutions.

In this environment, simply deploying more and faster storage devices is not enough; a new kind of infrastructure is needed, one that provides network availability, data accessibility, and system manageability. The SAN meets this challenge: it is a high-speed network that establishes a direct connection between storage resources and servers. The SAN liberates the storage device, so that it is not attached to a particular server, and is attached directly to the network. In other words, storage is externalized and functionally distributed to the organization. The SAN also enables the centralizing of storage devices and the clustering of servers, resulting in easier and less expensive administration.

IT organizations are facing increasing demands for network uptime, data accessibility, and system management—all with flat or declining budgets. The SAN addresses those needs by taking advantage of today's network and channel technologies to consolidate servers, centralize backup, and externalize storage.

Figure Figure 10 on page 22 shows a SAN architecture schematic.

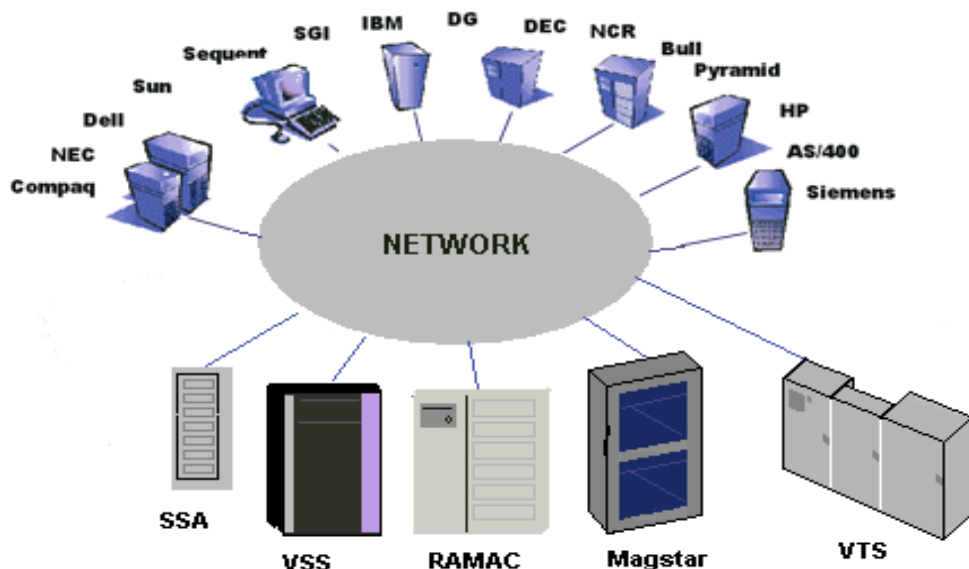


Figure 10. SAN Architecture

SAN uses industry standard components:

- Fibre optic cable SC connectors
- Fibre optic cable—multimode: short wave, up to 500 m; and singlemode: long wave, up to 10 Km.
- Gigabit Interface Converter (GBIC)—converts copper interface to optic interface.
- Host Bus Adapters (HBA)—FC PCI HBA: copper or optic media
- Hubs - Hot pluggable servers and storage—connect 7 to 10 nodes in a loop stacked to maximum of 127 ports.
- Fibre Channel to SCSI Bridge—connect SCSI-attached storage to Fibre Channel.
- Fibre Channel Switch—stackable to connect over 16 million nodes.

2.3 SCSI and SSA

This section provides a brief description of SCSI and SSA technologies. SCSI and SSA, as well as Fibre Channel are used on storage applications. All these technologies are available today. The FC-RSS takes full advantage of Fibre Channel as well of SCSI.

2.3.1 SCSI Technology

Small Computer System Interface (SCSI) is a bus-level interface. Computers can communicate with a variety of devices, such fixed disk, CD-ROM, and tape drives, in a daisy-chained bus. Four different types of buses are available: SCSI-1, SCSI-2, Ultra-SCSI and Ultra2-SCSI.

2.3.1.1 SCSI-1

SCSI-1 is an 8-bit interface with a burst transfer rate of 5 MB/s. SCSI is an electrical bus and must have some form of resistive termination on each end of the bus to prevent signal reflection. A device can be used as a terminator and must be placed on the last connector of the bus. SCSI-1 supports a maximum of eight devices (seven plus the host adapter) per bus.

2.3.1.2 SCSI-2

SCSI supports 8-bit, 16-bit and 32-bit devices. SCSI-2 is a faster bus, at 10 MB/s for 8-bit transfer and 20 MB/s for 16-bit transfer. SCSI-2 comes in many varieties: SCSI-2, SCSI-2 Fast, and SCSI-2 Fast/Wide. The maximum number of devices supported by one SCSI channel is 16. Other enhancements are: a Common Command Set (CCS) and a Tagged Command Queuing (TCQ). Currently, 32-bit bus is not implemented.

2.3.1.3 Ultra SCSI

Ultra-SCSI provides 16-bit transfer rates of 40 MB/s. It supports up to 15 devices. All types of SCSI devices can be mixed and coexist at the same bus with an Ultra-SCSI adapter. They will operate at the speed specified by the device. The pin and cable is compatible with SCSI-2 Fast and Fast/Wide (no change to physical connection). The higher transfer rate reduced the maximum cable length to 1.5 meters without using a SCSI repeater.

2.3.1.4 Ultra2 SCSI

The Ultra2-SCSI transfer rate goes up to 80 MB/s for 16-bit transfers. With the double transfer rate, cable lengths would be halved, and at 75 cm, would become unmanageable. To compensate for this, Ultra2 uses a new interface called *Low Voltage Differential Signalling* or LVD, which increases the length of cables. The maximum cable length is 25 meters. In addition, Ultra2 has two versions: Ultra2-SCSI (8-bit) and Wide Ultra2-SCSI (16-bit).

2.3.2 SSA

SSA is an innovative, Open Standards protocol that overcomes the limitations of parallel interfaces while providing the extensive benefits of serial technology at lower cost and higher performance. First developed by IBM, SSA is capable of supporting 128 devices at a maximum data transfer rate of 80 MB/s per loop using cables of 25 meters in length on a copper wire and up to 10 kms using fibre optics.

SSA is a serial link designed especially for low-cost, high-performance connections between I/O devices. SSA uses a cable containing two pairs of wires providing a two-signal connection. One pair is used for sending data and the other to receive data, making SSA a true full-duplex architecture.

SSA supports disks, CD-ROMs, tape drives, optical drives, printers, and scanners. The dual port card supports four simultaneous conversations at 80 MB/s. SSA provides automatic configuration and hot swapping. Multiple devices can transfer data simultaneously, which increases performance (unique to SSA). SSA controllers require SSA devices. Many features of SSA make it well suited for high-availability environments:

- Auto-configuration of devices, including a flexible addressing scheme that allows connections as strings or loops

- Hot swapping (the insertion or removal of nodes without switching the network)
- No single point of failure
- High level of error detection and error handling protocols

In comparison to SCSI, SSA provides the best trade-off between cost, availability, performance, and expandability.

Loop topology is the most common implementation. Devices form a series of point-to-point serial links. Up to 25 meters between devices, 2 data paths between any 2 disks can be configured with no single point of failure. Auto-selection of an alternate path in event of failure, and automatic loop rebuild after repair, are also provided.

Figure Figure 11 on page 24 shows a single loop with four disks attached to the adapter.

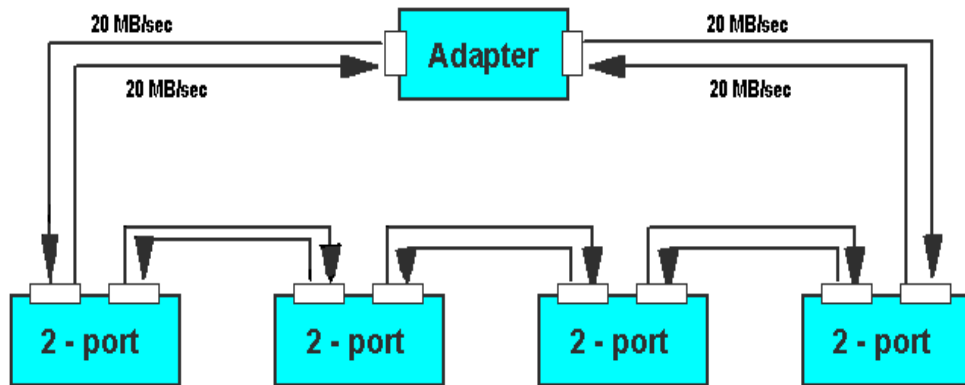


Figure 11. SSA Loop Configuration

One of the most impressive benefits SSA offers over SCSI is its high bandwidth, which yields excellent performance, even under heavy workloads. SSA addresses SCSI's throughput bottleneck with its fundamental building block—the SSA connection, or node.

Overall system availability is enhanced by the ability of SSA devices to act independently—peripherals can be attached in multiple paths, making a fault-tolerant system much easier to implement. In addition to enhancing fault tolerance, avoiding single points of failure in the path design also simplifies network maintenance.

SSA helps you to dynamically add subsystems and grow incrementally, easily, and inexpensively along with your business. SSA supports flexible configurations, enabling an assortment of system connection options that help you improve system throughput by isolating subtasks from production tasks.

2.3.3 RAID Technology

RAID stands for "Redundant Array of Inexpensive (Independent) Disks" and was developed by Berkeley University in 1987 in order to provide a method of classifying the different ways of using multiple disks to increase availability,

performance, and capacity. There are currently many different levels of RAID. Each definition performs optimally in a different environment. An array is nothing more than a collection or group of hard drives, and each RAID level distributes the data across the drives in the array in a slightly different manner.

The FC-RSS implements five levels of array: RAID-0, RAID-1, RAID-3, RAID-5 and RAID-0+1. This section covers these five levels.

2.3.3.1 RAID-0—Block Interleave Data Striping

Striping of data across multiple disk drives is a technique sometimes employed to maximize performance of a disk subsystem. An additional benefit of this data organization is drive spanning. Data striping improves performance with files, since reads and writes are overlapped across all disks. However, reliability is decreased, because the failure of one disk will result in a complete failure of the subsystem.

Figure Figure 12 on page 25 shows how RAID-0 is striped across all disks in the array.

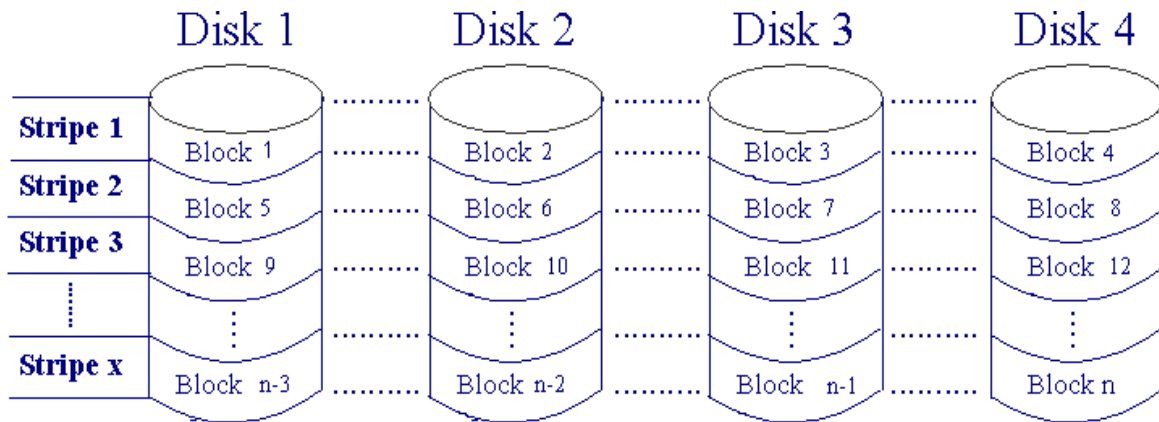


Figure 12. RAID-0

2.3.3.2 RAID-1—Mirroring

This approach retains two complete copies of all data. Whenever the system makes an update to a disk, it duplicates that update to a second disk, thus mirroring the original. Either disk can fail, and the data is still accessible. Additionally, because there are two disks, a read request can be satisfied from either device, thus leading to improved performance and throughput. Some implementations optimize this by keeping the two disks 180 degrees out of phase with each other, thus minimizing latency. However, mirroring is an expensive way of providing protection against data loss, because it doubles the amount of disk storage needed (as only 50 percent of the installed disk capacity is available for data storage).

Figure Figure 13 on page 26 shows how RAID-1 data is mirrored onto the second set of drives.

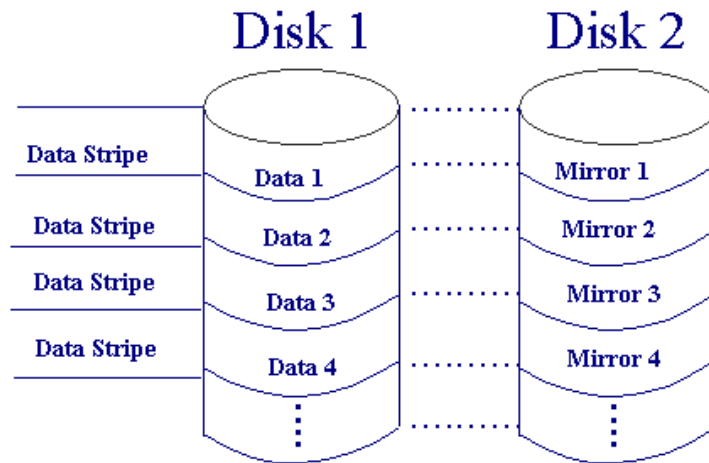


Figure 13. RAID-1

2.3.3.3 RAID-3—Bit Interleave Data Striping with Parity Disk

In this approach, there is data striping across all the disks, and all parity information is on a single drive. The data is split up on bits to fill the blocks on the disks. RAID-3 is best for large sequential I/O. RAID-3 offers fault tolerance and data integrity. All RAID-3 actuators move in tandem, offering high throughput for a single task; thus is suited for specialized applications but is inappropriate for multi-user systems.

Figure Figure 14 on page 26 shows RAID-3 data striped across disks in the array, with the parity fixed to one drive for all I/Os.

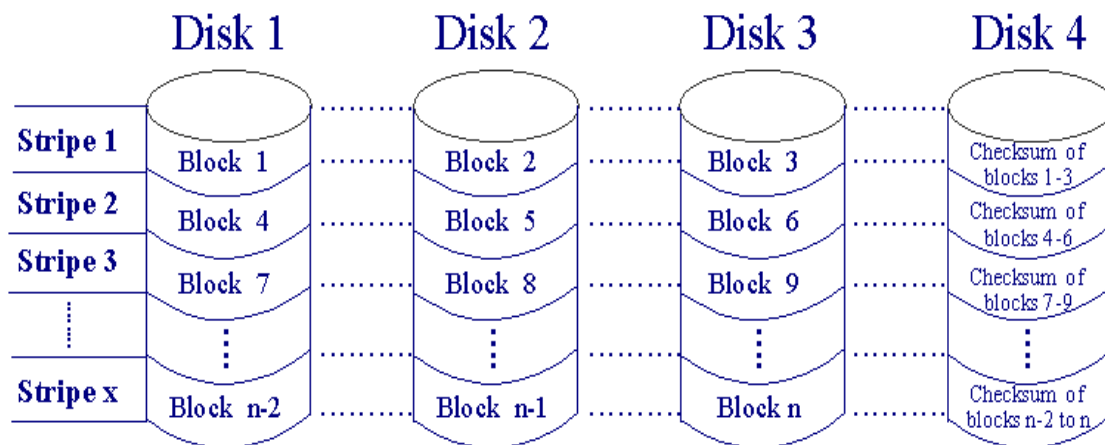


Figure 14. RAID-3

2.3.3.4 RAID-5—Block Interleave Data Striping with Skewed Parity

RAID-5 spreads both the data and data parity information across the disks one block (strip) at a time to ensure maximum read performance when accessing large files, and to improve array performance in a transaction-processing environment. Parity information is striped across the drives in order to balance the parity I/O. The equivalent of one drive per array is used for the parity data,

regardless of the size of array. Once again, the capacity left for data storage is always $n-1$. Writes suffer from the RAID-5 write penalty: for each write at the application level, four I/Os must occur in the subsystem. This is partially alleviated by full stripe or partial stripe writes.

Figure Figure 15 on page 27 shows how RAID-5 data and parity is striped across all disks in the array.

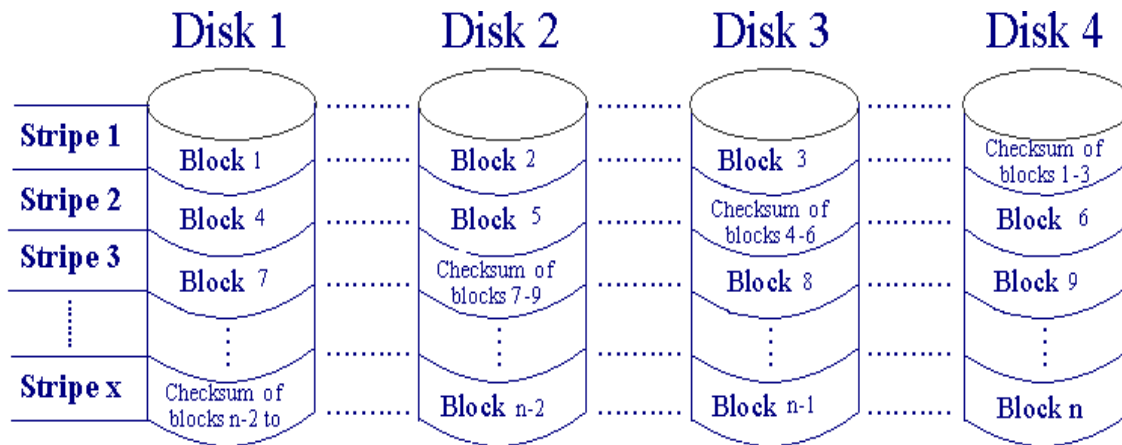


Figure 15. RAID-5

2.3.3.5 RAID-0+1—Block Interleave Data Striping and Mirroring

RAID-0+1 is also known in the industry as RAID-10. RAID-10 is not formally recognized by the RAID Advisory Board (RAB), but, as it is an industry standard term, we will use it here to refer to RAID-0+1. In RAID-10, data is striped across multiple disk drives, and then those drives are mirrored to another set of drives. The performance of RAID-10 is approximately the same as RAID-0 for sequential I/Os. RAID-10 provides an enhanced feature for disk mirroring that stripes data and copies the data across all the drives of the array. The first stripe is the data stripe; the second stripe is the mirror (copy) of the first data stripe, but it is shifted over one drive. Because the data is mirrored, the capacity of the logical drive is 50 percent of the physical capacity of the hard disk drives in the array.

Figure Figure 16 on page 28 shows RAID-0 striped data on disk 1 and disk 2, while the mirrored data is striped on disk 3 and disk 4.

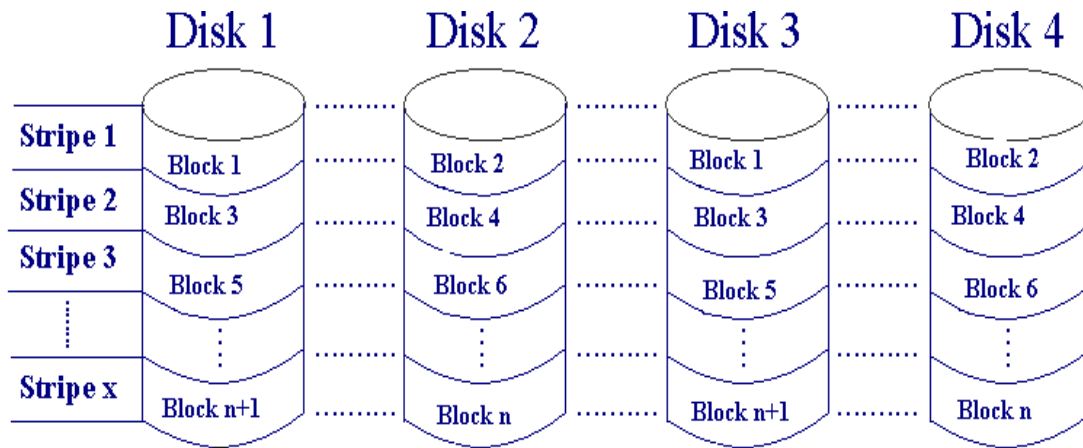


Figure 16. RAID-10

Chapter 3. Fibre Channel

The Fibre Channel communications protocol provides the means of implementing a high-speed, high-availability SAN. It is engineered to create an expandable, inexpensive, and practical means of transporting data between systems ranging from laptops to supercomputers. The FC standard is being developed by ANSI task group X3T11, Physical and Signaling Interface (FC-PH).

The following sections describe the FC layers, topologies, and equipment in detail.

3.1 Fibre Channel Layers

Fibre Channel (FC) is broken up into a series of five layers. The concept of layers, starting with the ISO/OSI seven-layer model, permits development of one layer to remain independent of the adjacent layers. Although FC contains five layers, those layers follow the general principles stated in the ISO/OSI model. The five layers are divided into two parts, Physical and Signaling layer and Upper layer. The five layers are illustrated in Figure 17:

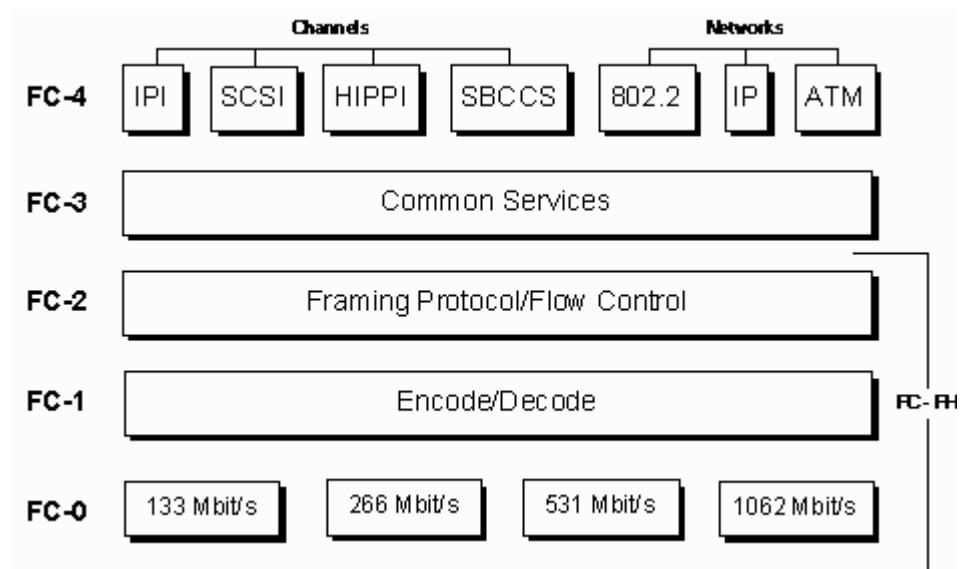


Figure 17. Fibre Channel Layers

3.1.1 Physical and Signaling Layers

The Physical and Signaling layers include the three lowest layers: FC-0, FC-1, and FC-2.

3.1.1.1 Physical Interface and Media: FC-0

The lowest layer (FC-0) defines the physical link in the system, including the cabling, connectors and electrical parameters for the system at a wide range of data rates. This level is designed for maximum flexibility, and allows the use of a large number of technologies to meet the desired configuration.

A communication route between two nodes may be made up of links of different technologies. For example, in reaching its destination, a signal may start out on copper wire and become converted to single mode fibre for longer distances. This flexibility allows for specialized configurations depending on requirements.

Fibre Channel often uses lasers to transmit data, and can thus present an optical health hazard. The FC-0 layer defines an Open Fibre Control (OFC) system, acts as a safety interlock for point-to-point fibre connections that use semiconductor laser diodes as the optical source. If the fibre connection is broken, the ports send a series of pulses until the physical connection is re-established and the necessary handshake procedures are followed.

3.1.1.2 Transmission Protocol: FC-1

The second layer (FC-1) provides the methods for adaptive 8B/10B encoding to bind the maximum length of the code, maintain DC-balance, and provide word alignment. This layer is used to integrate the data with the clock information required by serial transmission technologies.

3.1.1.3 Framing and Signaling Protocol: FC-2

Reliable communications result from Fibre Channel's FC-2 Framing and Signaling Protocol. FC-2 specifies a data transport mechanism that is independent of upper layer protocols. FC-2 is self-configuring and supports point-to-point, arbitrated loop and switched environments.

Together, FC-0, FC-1, and FC-2 form the Fibre Channel Physical and Signalling Interface (FC-PH). FC-2, which is the third layer of the FC-PH, provides the transport methods to determine:

- Topologies based on the presence or absence of a fabric
- Communication models
- Classes of service provided by the fabric and the nodes
- General fabric model
- Sequence and exchange identifiers
- Segmentation and reassembly

Data is transmitted in 4-byte Ordered Sets containing data and control characters. Ordered Sets provide the availability to obtain bit and word synchronization, which also establishes word boundary alignment.

3.1.2 Upper Layers

The Upper layer include two layers: FC-3 and FC-4.

3.1.2.1 Common Services: FC-3

The FC-3 defines functions that span multiple ports on a single node or fabric. Functions that are currently supported include:

- Hunt groups. A hunt group is a set of associated N_ports attached to a single node. This set is assigned an alias identifier that allows any frames containing the alias to be routed to any available N_port within the set. This decreases latency in waiting for an N_port to become available.
- Striping. Striping is used to multiply bandwidth, using multiple N_ports in parallel to transmit a single information unit across multiple links.

- Multicast. Multicast delivers a single transmission to multiple destination ports. This includes the ability to broadcast to all nodes or a subset of nodes.

3.1.2.2 Upper Layer Protocol Mapping (ULP): FC-4

The highest layer (FC-4) provides the application-specific protocols. Fibre Channel is equally adapt at transporting both network and channel information and allows both protocol types to be concurrently transported over the same physical interface.

Through mapping rules, a specific FC-4 describes how ULP processes of the same FC-4 type interoperate. A channel example would be sending SCSI commands to a disk drive, while a networking example would be sending IP (Internet Protocol) packets between nodes.

3.2 Fibre Channel Topologies

Fibre Channel relies on point-to-point communication, but this does not require two ports to be directly connected to each other. The three specifications are Point-to-Point, Fabric, and Loop, as shown in Figure 18 on page 31.

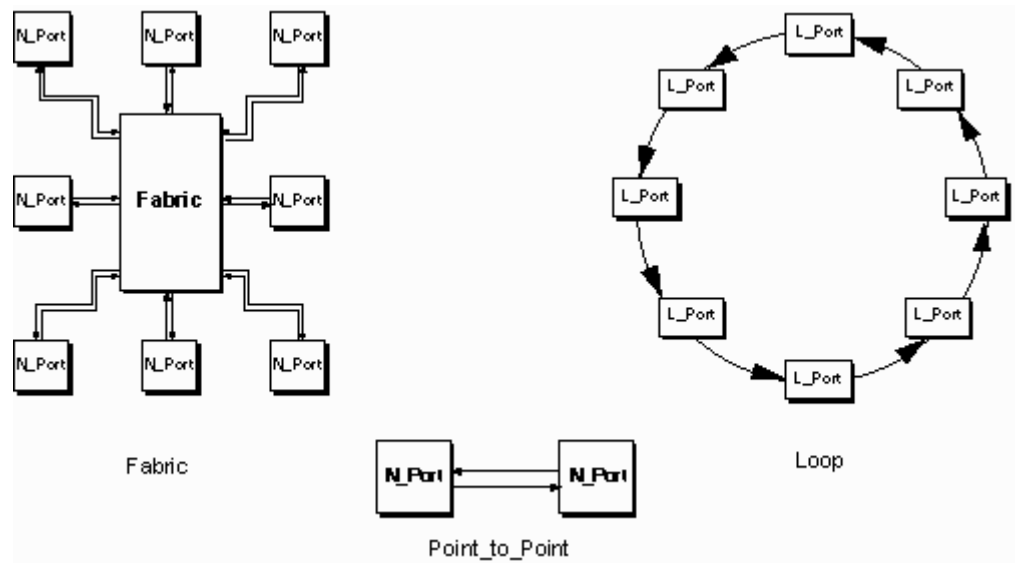


Figure 18. Fibre Channel Topologies

3.2.1 Point-to-Point

A *point-to-point* connection is the simplest topology. It is used when there are exactly two nodes, and future expansion is not needed. There is no sharing of the media, which allows the devices to use the total bandwidth of the link. A simple link initialization is needed before communications can begin.

3.2.2 Fabric

Expanding on the point-to-point topology is the *fabric* or *switched* topology. The fabric is an intelligent device which sets up point-to-point connections with itself and its attached nodes. Up to 2^{24} devices can be connected via a fabric. These nodes can be workstations, Fibre Channel Hubs or other fabrics. The fabric is

designed to handle multiple conversations at one time and route the appropriate frames to their specified destinations using the full bandwidth of the link.

3.2.3 Loop

The third topology, the *loop*, raises the level of complexity by introducing multiple nodes on a single connection. This is a cost-effective method for attaching up to 127 devices together. The loop uses a shared medium, thus arbitration is necessary to determine which node can transmit data.

If more than one node on the loop is arbitrating at the same time, the node with the lower Arbitrated Loop Physical Address gains control of the loop. Upon gaining control of the loop, the node can then establish a point-to-point transmission with another node using the full bandwidth of the media. When a node is finished transmitting its data, it is not required to give up control of the loop. This is a channel characteristic of FC. However, there is a "fairness algorithm", which states that a device cannot regain control of the loop until the other nodes have had a chance to control the loop.

3.3 Fibre Channel Hardware

The hardware necessary to implement a Fibre Channel configuration is very similar to that of implementing a token-ring or ethernet based network.

3.3.1 Connectors

The distance between nodes can determine the type of media that will be used. This in turn determines the type of connector that attaches to the cabling, or the type of modules that are used in the fabric. The flexibility of FC is evident in its wide variety of connectors and mediums, as specified by the FC-0 layer with distances supported up to 10km.

Fibre Channel uses a low-cost duplex SC connector for fibre optic interfaces. SC connectors ensure polarization and push-pull, insertion-removal ease of use for high reliability. The Galaxy connector reduces the connector footprint by 50%. A nine-pin D-type connector is used for copper interfaces. The HSSD connector is used with the copper GBIC. Media interface modules and connectors are shown in Figure 19.

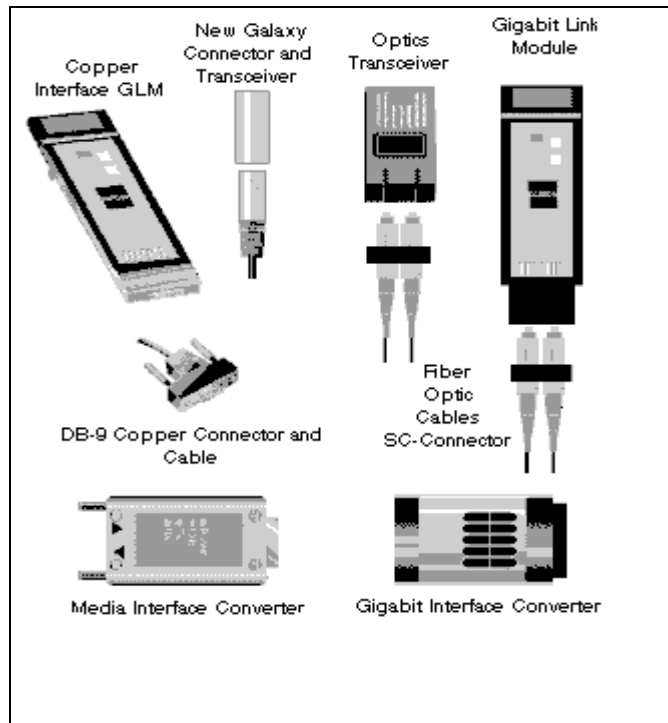


Figure 19. Flexible Use of Gigabit Copper or Fibre Optics

3.3.2 Concentrators

There are two types of wire concentrators that are used with Fibre Channel: hubs and switches.

The first type of connector is the hub, which provides a logical star topology while creating a physical loop topology. It is the central piece of hardware when implementing FC-AL. A hub, for example, could contain seven bays in which Gigabit Interface Converters (GBICs) of different types could be inserted. Thus, a short-wave fibre GBIC, a long-wave GBIC, and a copper GBIC could all be inserted into a hub providing different distance capabilities.

The second type of concentrator is the fabric switch. It is required to create the fabric topology, and implements a physical and logical star topology. Like the hub, the fabric has the ability to use different types of GBICs. However, the fabric's switching capabilities allow it to route the FC frame to the correct port and not simply route the frame to the next port on the loop.

Both hubs and fabrics can have the intelligence to bypass a malfunctioning port, thus preventing a defective node from bringing down the entire SAN.

Chapter 4. Performance Considerations

In this chapter, we provide some performance benchmark results, and discuss some considerations that need to be taken into account in order to obtain satisfactory throughput or I/O rates from your subsystem.

The subsystems used were not tuned using any of the methods discussed in Chapter 9, "Performance Tuning" on page 105. All tests were performed using the standard setup.

Note: Recommendations provided in this chapter have been based on measurements conducted under specific workload conditions. Performance data contained herein was obtained in a controlled environment. Measurements were obtained from a subsystem prior to general availability. Results from later levels may be different. Actual results that may be obtained in other operating environments may vary significantly. These values do not constitute a guarantee of performance.

4.1 Overview

The FC-RSS supports multiple RAID configurations: RAID-0, RAID-3, and RAID-5, as well as RAID-10, a variant of RAID-1 supporting mirroring of groups of drives. One advantage of RAID-10 is that it allows for striping data across disk drives, therefore balancing load across those drives. This can be done within the disk subsystem itself, independent of the host operating system. Another value of mirroring groups of drives as opposed to pairs of drives only is evident on the Windows NT file system (NTFS), which restricts a user to a maximum of 26 drive letters. RAID-10 clearly has important advantages over RAID-1; nevertheless, within this chapter we will refer to the RAID-10 implementation as RAID-1 to maintain consistency with existing literature regarding this product.

Internally, the storage subsystem is fully redundant, with two sets of power sources throughout, dual internal controllers, and up to 60 drives, which can be configured in fail-safe configurations, such as RAID-1, RAID-3, and RAID-5. Additionally, there are dual external Fibre Channel ports, as well as six internal low-voltage differential (LVD) busses. The fibre channel ports support up to 100 MB/s each, limiting external throughput to a total of 200 MB/s. In fact, the FC-RSS device has been shown to achieve rates quite close to this external throughput limit in performance tests conducted using Windows NT.

Within each of the internal controllers, records that are written are stored in nonvolatile cache memory, and when write cache mirroring is enabled (which is the default), they are additionally stored in the other internal controller's nonvolatile cache. Enabling write cache mirroring incurs some additional overhead costs but can be useful if the desire is to ensure 100% data availability, protecting against failure of a single controller. Even without write mirroring, in nearly all circumstances, data remains secure (but temporarily inaccessible) within the failed controller if one of the internal controllers fails. Your decision to enable write cache mirroring and a particular RAID level should be based on data availability and performance requirements.

4.1.1 Test Configuration

Following are measured rates for the FC-RSS I/O subsystem utilizing a variety of RAID levels, both with and without write cache mirroring. All of these measurements were conducted on a FC-RSS containing 48 disk drives of mixed capacity. The subsystem can support a maximum capacity of 60 disk drives, and such a configuration is likely to provide even higher performance results for some of the test cases that are limited in performance by the number of drives.

Two host systems, each running Windows NT with dual processors were used. One of the two hosts had dual PentiumPro 200 MHz processor, and the other had dual PentiumPro 300 MHz processors. Each host had between 64 and 128 MB of dynamic random access memory (DRAM). One Qlogic adapter was installed in each host (for a total of 2) and directly connected to one of the fibre channel ports in each of the internal controllers in the I/O Subsystem. Internally, the I/O subsystem was configured with 8 logical units names (LUNs) and 8 RAID groups, each containing 6 disk drives, one from each of the different internal LVD busses. A relatively small I/O queue depth was used—between 5 and 6 I/Os per LUN. Larger queue depths could conceivably result in even higher throughput and I/O rates, but at a cost of greater response times. A relatively small queue depth was selected in order to maintain I/O response times to a reasonable level.

4.1.2 Measurement Results

The following contains measurement results for a variety of workload and configuration cases. Measurements were conducted with the following variations: RAID options: RAID-0, RAID-1, RAID-3, and RAID-5 workload characteristics: Random and sequential access with 4K and 256K record sizes, and special tests that stress maximum cache performance with cache mirroring enabled or disabled

The following are some of the conclusions and observations from these results: The FC-RSS has demonstrated impressive throughput characteristics, up to 190 MB/sec and 16,000 IO/sec. Write throughput depends both on the RAID level selected and use of write mirroring. In general, the performance with write cache mirroring enabled is very good, and probably adequate for most applications. The use of write cache mirroring is recommended unless you have unusually high write performance requirements.

Note: All throughput rates in this report are reported either in Millions of Bytes per second (MB/s), where 1 MB = 1,000,000 bytes, or in I/O operations per second (IO/s). In all cases, local host caching was inhibited for the purposes of conducting these measurements.

4.1.2.1 Sequential and Cache Throughput Rates for 256 KB Records

Figure 20 on page 37 and Figure 21 on page 37 shows the maximum throughput rates for sequential access of 256 KB records and for a pure cache-hit workload, for various RAID levels, with and without use of write cache mirroring. In these cases, the subsystem delivers nearly 180 MB/s for sequential reads. Sequential writes deliver between 48 MB/s and 144 MB/s, depending on RAID level and use of write cache mirroring. Cache hit performance achieves 190 MB/s, near the maximum capable for two fibre channel ports.

Sequential Throughput 256 KB Records

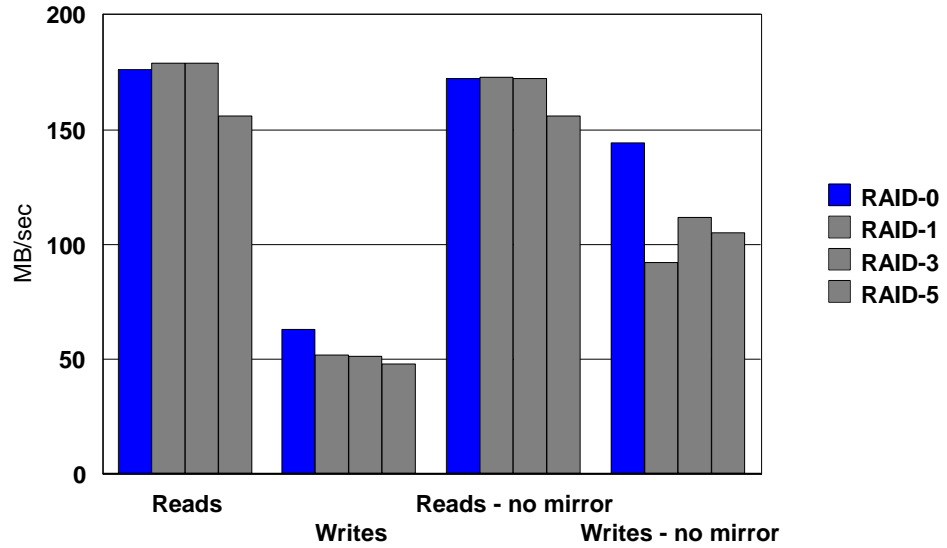


Figure 20. Sequential Throughput for 256 KB Records

Cache Hit Performance 256KB Records

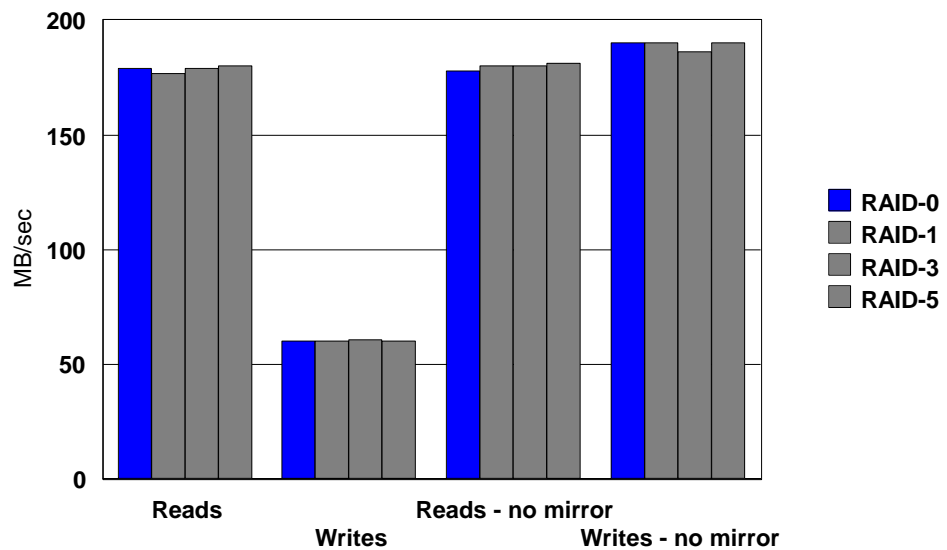


Figure 21. Cache Hit Performance for 256 KB Records

4.1.2.2 Random Workload Measurements for 256 KB Records

For the RAID-5 case with mirrored writes (only), we have some additional throughput measurements available for random 256 KB I/O requests. Random I/O request throughputs are highly dependent upon the physical device speeds, although write I/Os tend to garner some advantage from the write cache. The random I/O results are shown in Figure 22 on page 38.

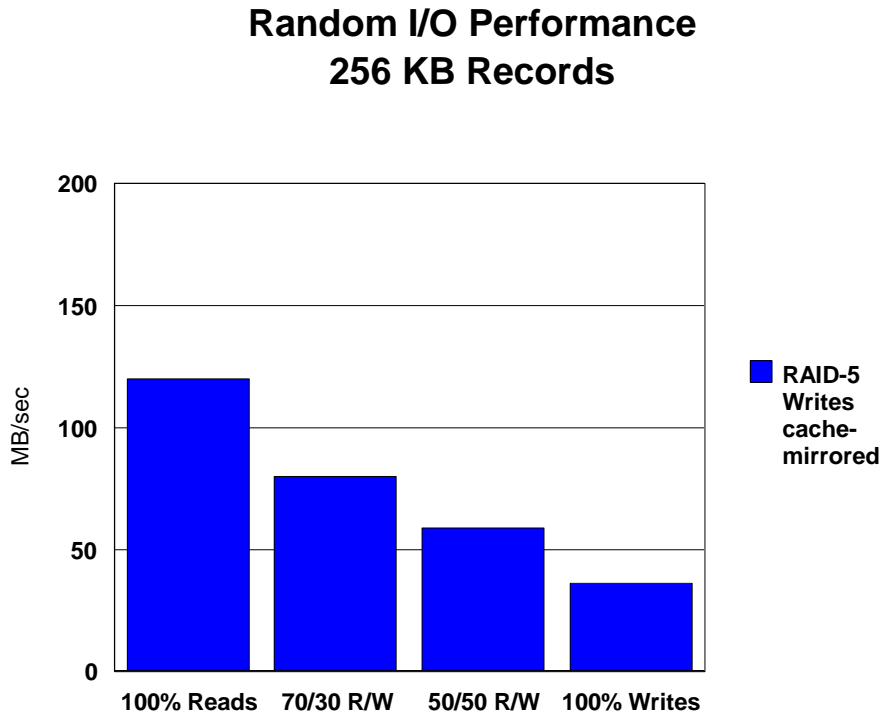


Figure 22. Random I/O, RAID-5 for 256 KB Records

4.1.2.3 Sequential and Cache Throughput Rates for 4-KB Records

Measuring the number of I/O operations per second is also a useful measure for evaluation. The primary value of such a measure is to allow an assessment to be made with regards to the overhead costs implicit within the storage subsystem controller. As before, in all cases local host caching was inhibited for the purposes of conducting these measurements. Figure 23 on page 39 and Figure 24 on page 39 shows these results.

Sequential Throughput 4 KB Records

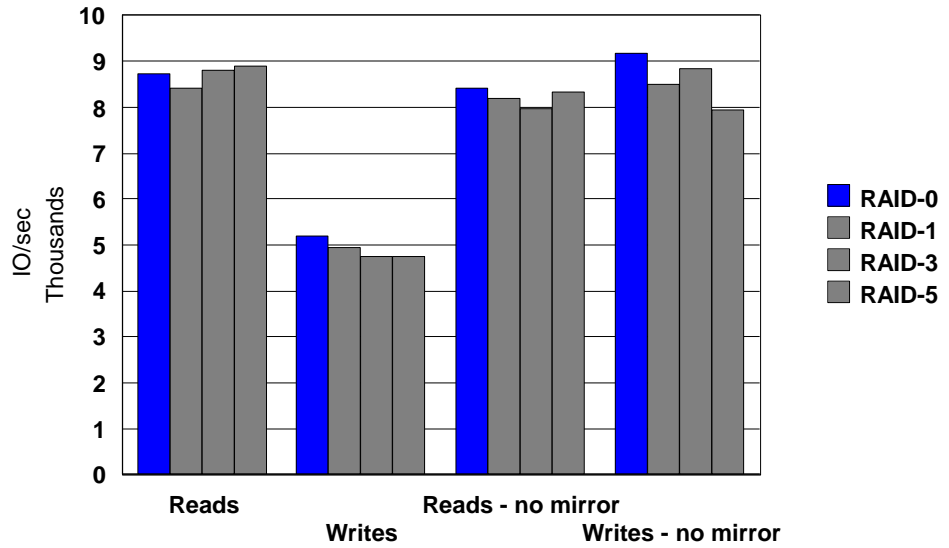


Figure 23. Sequential Throughput for 4 KB Records

Cache Hit Performance 4 KB Records

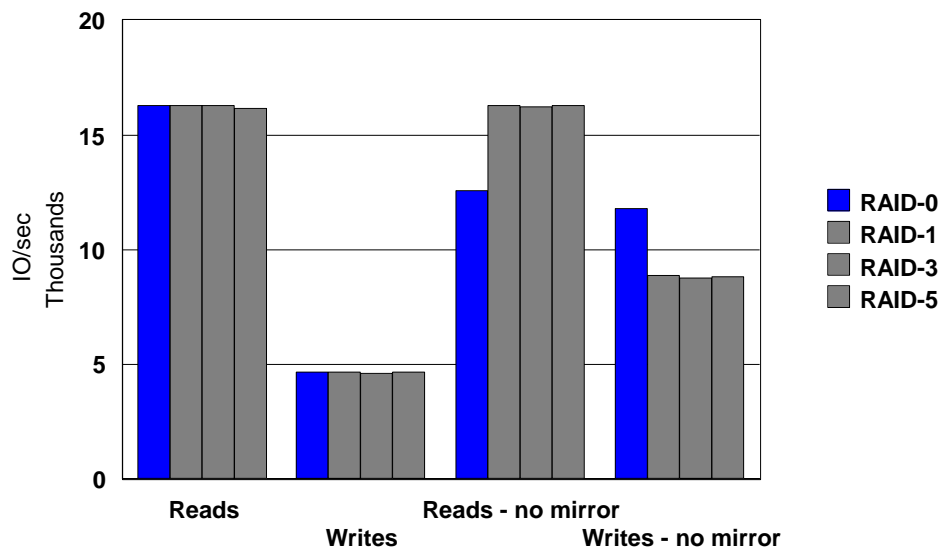


Figure 24. Cache Hit Performance for 4 KB Records

4.1.2.4 Random Workload Measurements for 4-KB Records

Random workloads tend to be more dependent on device speeds. Random reads require device accesses in order to complete. Write requests are able to complete immediately as long as adequate free space exists within the write cache in the storage subsystem. However, at some point, random writes also result in actual device activity for each request as data is destaged from the cache, and thus are also dependent upon physical device speeds.

Device dependent I/O activity, particularly with small block I/Os, is typically quite representative of actual workloads. Very often the host operating system absorbs the majority of "cacheable" accesses, leaving the disk subsystem with more random accesses. For some workloads, a substantial portion of the I/O activity in practice would result in both host and controller cache hits. It should be noted that all of these measurements are for 100% random workloads; thus, these values represent very conservative I/O rates with regard to actual field results for similar configurations. Typical workloads are expected to yield results between these numbers and those values presented in the previous two charts. Figure 25 on page 40 shows the results for 4 KB random I/Os:

Random Access Throughputs 4 KB Records

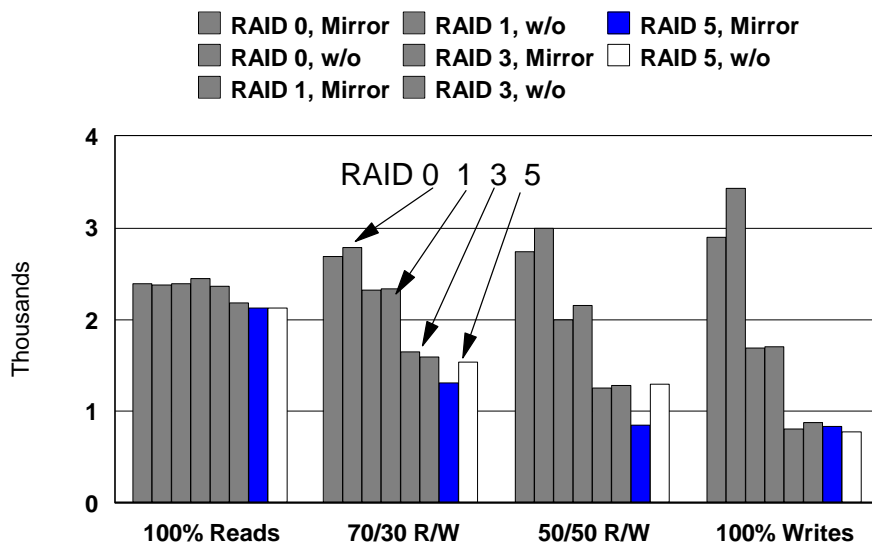


Figure 25. Random Access Throughputs for 4 KB Records

4.1.3 Additional Considerations

The FC-RSS is highly fault tolerant. A limited amount of performance testing was conducted during various failure conditions, as well as during RAID-5 drive rebuilding. The failure condition testing was for the RAID-5 with mirrored write configuration only. The results were very favorable. A down-one disk or down-one controller resulted in a performance degradation for a 70/30 random workload of much less than 10% (typically it was 5% or less). In fact, the only time that degradation ever exceeded 10% in these tests was when the power was shut off

on a entire drawer of drives, as well as during the subsequent rebuild of that row. In such a dramatic case, the degradation varied, but was around 20%. For an individual drive or controller (either down-one or drive rebuilding) the degradation was close to negligible.

In addition to the option for mirroring or not-mirroring of cached writes, there are other options available for tuning I/O performance. For example, the internal cache segment size can be changed from the default value of 4 KB to 16 KB, but no performance results are available. Changing the cache block size to a larger value is expected to even further improve throughput rates, particularly for mirrored writes, but is expected to come at a cost to the I/O rate. Of course, the actual change in performance will depend on the type of work actually being done. If most of the work involves large block I/O, then changing to a larger internal cache block size is expected to be beneficial all of the way around. Otherwise, better results are likely from using the default cache block size.

4.2 Conclusions

The FC-RSS is capable of high throughput rates, and very respectable I/O rates. It is highly configurable among RAID-0, RAID-1 (actually RAID-10), RAID-3, and RAID-5, as well as in regard to other optimization parameters. As a result, it is expected that the FC-RSS will be the I/O subsystem of choice for environments requiring high performance and reliability concerns.

Chapter 5. Application Considerations

In this chapter we discuss the issues that need to be considered for implementing the FC-RSS in different operating environments. The intent is to provide guidelines, considerations, and suggestions on what needs to be done, and not to give specific recommendations for a particular installation.

5.1 General Considerations

In installing a storage subsystem, you must consider a number of factors. Each of these factors will make different demands on the system. It is the responsibility of the systems administrator and/or the storage systems administrator to gather all the various requirements, prioritize these requirements, and select the settings, parameters and components that provide the best compromise of function, availability and performance. User requirements for access, availability, performance, and service level will determine the system configuration.

5.1.1 Availability

A system that is required to provide 100% availability, or fault tolerant, can be very costly. One hundred percent availability means that the system must be available 24 hours a day, seven days a week, year in and year out. To achieve 100% percent availability, systems have to have at least total hardware and software redundancy. In addition to having redundant components, 100% availability dictates that if and when a specific component fails, its backup will take over the failed component's function without impacting system or data accessibility. Furthermore, it requires that the components be so reliable that no other critical component will fail until the first failed component has been repaired or replaced in a fashion that is not disruptive to either the system or to the user(s). It also requires concurrent hardware and software maintenance capabilities.

Often, a fault tolerant system is not required. You can be of greatest value to your customer if you assist him/her in determining the level of availability that is desired, needed, or affordable, and in assisting the customer in implementing what is needed to deliver that level of availability.

5.1.2 Performance

Performance is another area in which there may be limits. Customer requirements may range from a system that delivers the best performance technology can provide, regardless of cost, to a system that will accomplish the task at hand without a specific turnaround or response time criteria. Systems will have differing performance requirements. A database system, an OLTP application, a data warehousing application, a customer service or online ordering system, a financial analysis, and a geophysical analysis for oil exploration will all have dissimilar performance characteristics. The response time requirements and the quantity of data to be retrieved and processed for each of these differ, and vary widely from application to application and from installation to installation.

System tuning should always be done from the top down. That is, host system tuning will yield a better return before attention is focused on I/O subsystems. As

a rule of thumb, in performance analysis you should start with application tuning, followed by subsystem tuning (DBMS and or transaction processing), then processor capacity tuning before turning your attention to the I/O subsystems. Please refer to Chapter 9, “Performance Tuning” on page 105 for performance information and tuning information.

5.1.3 Cost Benefit

Invariably, one of the ultimate determinants of any system is its cost, specifically the cost of the system relative to the benefit it brings its user or owner. Unless there is no limit to the monetary amount that can be spent on a system, including operating, support, maintenance, and personnel, cost is the third critical factor in systems design or acquisition. The point is that there will almost always have to be a balance between performance, availability, and cost.

5.2 Database Management System (DBMS) Specific Considerations

These are the issues and factors that should be taken into consideration when implementing the FC-RSS disk subsystem in a DBMS environment. You or your customer should consult with the DBMS application and technical support in setting up your disk subsystem.

5.2.1 Common Considerations

The following topics are common to almost all DBMS implementations. Each DBMS system has unique functions and requirements, but they all, in one way or another, will require selection of FC-RSS functions and parameters to address their unique requirements. An area often overlooked is the currency of the product or the maintenance level of the software.

In general, an applications relative use of system resources dictates where you invest in those resources. Applications with large numbers of I/O segments will need more disk actuators to satisfy those requests. Sometimes investment in one resource will reduce that needed for another, for example, the use of host memory cache to reduce the I/O load.

5.2.1.1 Database Size

The overall size of the database has to be considered. Is the database made up of one or multiple tables; are there indexes? If the database is a single table, will it fit on one LUN?

5.2.1.2 Operating Systems

For the various operating systems, if there is a choice, does the DBMS work with raw volumes or with journaled file systems (JFS). What are the trade-offs for using one versus the other? Typically, when a DBMS uses a JFS that is part of the operating system, it provides the extra transaction and data integrity that are inherent in a JFS, which provides filesystem integrity by keeping a log of updates to the meta data of the file system. Consider the following:

- Oracle - 7, 8, OPS (Oracle Parallel Server)—Database size, number of databases, parallel or single server, indices, raw volumes on a JFS, logging, indexing, data access, number of users, synchronous or asynchronous, contention, buffering, system memory area, random or sequential, single or

multithread/tasks, operating system environment, processor resources, LAN, tuning guides, Oracle TARs (Technical Assistance Requests)

- DB2 - UDB (Universal Data Base), Parallel—Database size, number of databases, parallel or single server, indices, raw volumes on a JFS, logging, indexing, data access, number of users, synchronous or asynchronous, contention, buffering, system memory area, random or sequential, single or multithread/tasks, operating systems environment, processor resources, LAN, tuning guides, IBM Redbooks, Consultline
- Sybase, Informix, and other database management systems—Database size, number of databases, parallel or single server, indices, raw volumes on a JFS, logging, indexing, data access, number of users, synchronous or asynchronous, contention, buffering, system memory area, random or sequential, single or multithread/tasks, operating systems environment, processor resources, LAN, tuning guides, technical support

5.2.2 DBMS Functional Characteristics

Additional DBMS functions that may affect performance or have critical performance criteria. Some of these are:

- Database logging
- Database reorganization
- Process serialization
- Backup and recovery

5.3 OLTP Specific Considerations

In most cases, online transaction processing (OLTP) applications can be characterized by many transactions with random arrival rates retrieving data from files or databases in a random fashion with a mixture of read and write I/O requests. The balance of the read-to-write ratio is often biased towards read activity; a 70:30 read-to-write ratio is often quoted. However, ratios of 50:50 or 30:70 may characterize your customer's environment more accurately. The characteristics of the whole application or a particularly critical subset may determine the settings of the performance tuning parameters for the system. As in previous examples, all of the factors need to be considered, including processor capacity, system memory, application buffering, LAN and/or workstation constraints, and number of users.

For an application with a 70:30 read-to-write ratio, availability of write cache may not be as important as would be the case for an application that had a 50:50 read-to-write ratio. Another factor determining the amount of subsystem cache required is the amount of I/O buffering done by either the operating subsystem or by the applications subsystem. As a rule of thumb, if host server memory is being used for I/O cache or buffering, you will need four times more disk subsystem cache for the disk's cache to be effective.

5.4 Data Warehousing Applications

Typically, data warehousing applications are characterized by fairly static data, retrievals of large quantities of information, often sequential in nature, and little or

no update or write activity. What determines the key system requirements for these applications may be questions like these:

- Is this a customer service application?
- Is this a marketing analysis application?
- Is this a financial analysis application?
- Is this a market behavior analysis application?

Some of these applications require real time online response times, while others may be more time consuming overnight batch jobs. The characteristics of the application will determine what needs to be done.

5.5 SAP, BAAN or Other General Business Applications

These applications in general will be a combination of all of the above, plus potentially, bill-of-materials or materials-requirement-planning functions. You may have to consider not only relational database management systems but also hierarchical database systems.

Part 2. Implementation Processes

Chapter 6. Installation

This chapter provides the procedures to install both the hardware and the software for the FC-RSS.

6.1 Components

Before the installation, ensure you have all the components of the IBM FC-RSS. Following is a list of the components.

6.1.1 2101-100 IBM Seascope Solutions Rack

The IBM Storage Rack may be used to hold various Seascope components such as storage servers, drives modules, or hubs. It provides a straightforward method of consolidating servers and associated accessories, saving space while permitting easy access to crucial components. The rack comes with two power distribution units (PDUs). Power is distributed to the devices and drawers through one of these PDUs.

6.1.2 2102-F00 FC-RSS—Control Unit

This is a fibre channel - arbitrated loop (FC-AL) front end with an Ultra SCSI-2 back end. It contains redundant, dual, 6 channel, Ultra2, Low Voltage Differential SCSI (LVDS), FC-AL to SCSI, RAID controllers. The subsystem can contain up to 60 drives (6 drawers of 10 drives each). A 256 MB capable (128 MB standard) fast-write caching system implements RAID levels 0, 1, 3, 5, and 0+1 to provide fault-tolerance redundancy. A maximum configuration contains two hosts and two FC-RSSs. Both FC-AL to SCSI controllers are connected to all of the six backend SCSI busses. Failure of one of the controllers is transparent to the system back end.

6.1.3 2102-D00 Fibre Channel Storage Expandable Unit

The 2102-D00 is a rack-drawer storage unit containing 10 hot-swap bays that support half-height and slim-line SCSI disk drivers. It also contains an electronics board (ESM or Environmental Services Monitor board) providing two main functions:

- Status reporting for the subsystem through the SCSI interface
- SCSI connection between the subsystem and the FC-RSS

The drawer has two hot-swappable, redundant 350W power supplies. Each power supply contains its own power cord. In addition, two hot-swappable cooling units containing separate dual fans provide cooling redundancy.

The storage unit has two SCSI connections, both using the VHDC I.8 mm 16-bit SCSI connector. There are two configurations possible with the drawer: two independent 5-drive buses and one 10-drive bus.

6.1.4 2103-H07 Fibre Channel Storage Hub

The hub is used to connect the hosts with the control unit. The following list summarizes the features of the hub:

- Ports—Four short wavelength GBICs come standard. Seven total configurable media interface optical ports for best performance and distance.
- Plug and Play—Fault tolerance for on-line configuration changes
- Modular—GBIC transceiver design allows you to utilize ports when needed.
- Management—Simplified cable management for centralized storage management applications.
- Figure 26 on page 50 shows a picture of the 2103 fibre hub.



Figure 26. Fibre Hub

6.1.5 FC GBICs

GBICs are converters that allow connection between fibre cables and a fibre hub. The GBIC housing has an integral guide key that is designed to prevent improper insertion. Use minimal pressure when inserting a GBIC into a port. Forcing a GBIC into a port might cause damage to the GBIC or the port. The GBICs are hot pluggable.

6.1.6 Fibre Cables

This is the physical media that provides transportation of data between host, control unit and storage server. Cable lengths available are either 5 m or 25 m. IBM supports multi-mode, 50 micrometer, shortwave fibre optic cable, and can transfer data at 1GB.

6.2 Configurations

Following are the most common configurations for the FC-RSS:

- Single Host—One host machine is connected by two fibre channel loops to each controller in a control unit. Figure 27 on page 51 shows a schematic of a single host configuration which provides availability in the case of one of the FC-AL loops failing. It also allows you to balance, to a certain degree, the I/O load across each loop.

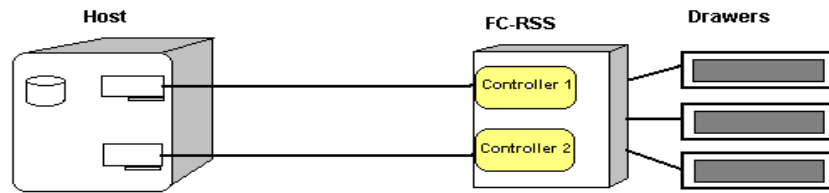


Figure 27. Single Host

- Independent Host—Two host machines are connected to a dual-controller unit. One host machine is connected by a Fibre Channel loop to one controller, and a second host machine is connected by another Fibre Channel loop to the other controller. Figure 28 on page 51 shows a schematic of an independent host configuration which does *not* provide availability in the case of a failing FC-AL loop, due to the failure of the FC-AL adapter, cable or controller.

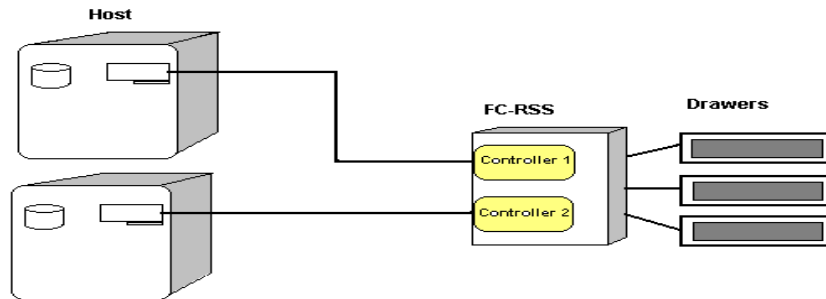


Figure 28. Independent Host

- Multi-host—A pair of hosts, each with host adapters connected to one or more RAID controller units. Windows NT requires MSCS and Windows NT 4.0 Server Enterprise. Use of a Fibre Channel Hub for each controller is recommended. Figure 29 on page 51 shows a schematic of a multi-host configuration which provides availability in case of failure of one of the FC-AL loops. In case one of the hosts fails, and with the appropriate software, the backup host can take over the failed host's disk.

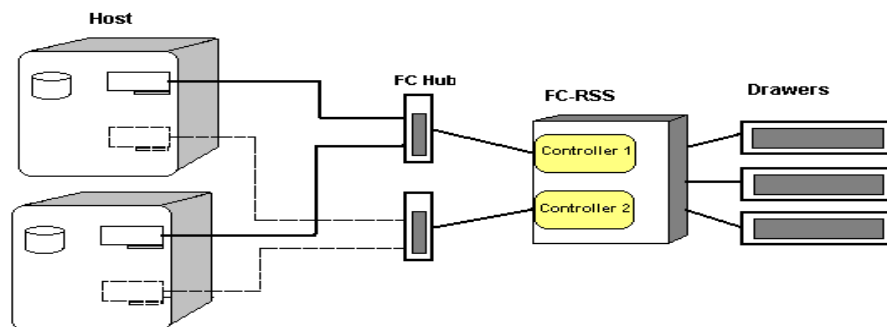


Figure 29. Multi Host (One FC-RSS)

6.3 Cabling

This section contains procedures for installing the IBM Fibre Channel RAID Storage Server.

6.3.1 Cabling the Rack

A rack may contain one or two power distribution units (PDUs). A PDU contains eight unswitched outlets.

Cabling of the rack is performed by the IBM Service Representative.

6.3.2 Cabling the Hub

The Fibre Channel Hub provides flexibility and is upgradable to available short-wave and long-wave optical fibre channel product port interface. Many configurations between host and hub depend on how many hosts and hubs you have. Initiators (servers) set up and initiate the transfer of data to or from the storage devices. The storage devices that receive the requests are the targets. Initiators and targets represent individual nodes that are linked by the shared FC-AL.

Installation tips:

- The GBIC housing has an integral guide key that is designed to prevent improper insertion.
- GBICs are hot pluggable. They can be inserted or removed while the hub is powered on.
- If a GBIC is inserted without media attached, it will remain in the *bypassed* state. When the media is attached, it detects a valid receive signal; the hub will then allow the GBIC and the attached devices to join the loop.

There are several types of GBICs available. You can insert any type of GBIC into any available hub port. The following list describes a typical installation:

1. Remove the plastic cover from the GBIC. All GBICs must have a cover to protect them from dust.
2. Insert the cables into the GBIC. Slide the GBIC into the port, then move the GBIC latch back to the locked position (flush with the rear of the GBIC). Connect the fibre optic cable to the installed GBIC.
3. Attach the cable assembly to the node that the particular port supports. The PCI Fibre Channel Adapter must be installed on one PCI slot inside the host.
4. Once the cabling is installed, power on the hub, then power on the host and the other devices to insure proper loop initialization.
5. Check the Device Active (green) LEDs on the Fibre Channel Storage Hub ports (Tablex).
6. Check the Port Bypass (amber) LEDs (see Table 5 on page 52).

Table 5. LEDs on HUB

Green LED	Amber LED	Port State
Off	Off	No GBIC installed
On	Off	Operational GBIC/Valid Signal

Green LED	Amber LED	Port State
Off	On	Fault GBIC/Port bypassed
On	On	Operational GBIC/No valid signal/Port bypassed *
* Change the position of the GBIC		

6.3.3 Cabling the Control Unit

For fibre optics connections, multi-mode, 50 micrometer, or 62.5 micrometer fibre optic cable must be used. To connect the cable to the Control Unit, a Media Interface Adapter (MIA) must be used, because the fibre optic connector in the Control Unit is a 9-pin connector for copper wire. Each controller on a Fibre Channel loop must have a unique *Fibre Host ID number*. The host ID numbers assigned to each controller are based on two elements:

- Host ID number set through hardware switches on the Control Unit. There are five Host ID switches that allow set ID numbers 0 through 127 for each controller.
- Software algorithms that calculate the actual Fibre Channel address, based on the Control Unit's hardware settings and the position on the loop or hub.

Figure 30 on page 53 shows the connectors of the Fibre Channel Storage Unit.

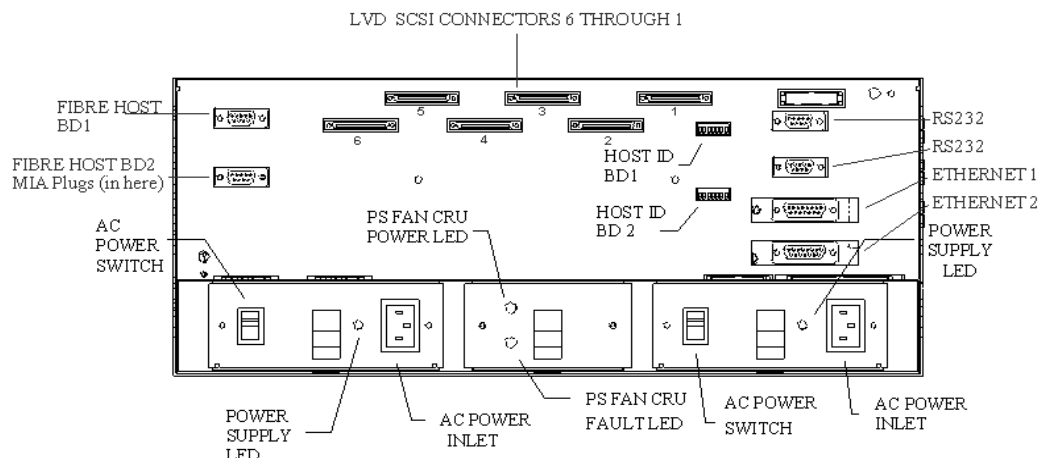


Figure 30. Control Unit (Rear View)

6.3.4 Cabling the FC Storage Unit

To connect the Control Unit to a Storage Unit, use the 68-pin, VHDCI LVD, Ultra-SCSI cables, provided by IBM. The Control Unit has six SCSI bus connectors and the Storage Unit has 2 connectors. Each connector represents a single SCSI bus that supports up to 10 drives per channel for a total a 60 drives.

6.3.4.1 Configuration for Two Independent SCSI Buses

To configure both SCSI buses of the drawer, independent of each other, connect one cable from a Control Unit's channel to the Bus 1 connector, and another channel to the Bus 2 connector. To separate the two buses, set switch 1 in the switch block to on (up). In this configuration, each bus contains 5 disks.

6.3.4.2 Configuration for One SCSI Bus

To configure the drawer as a single 10-disk SCSI bus, attach one SCSI cable from one controller's bus to the Bus 1 connector, then set switch 1 to off (down). There is no need to connect a terminator to the Bus 2 connector, as the drawer will automatically terminate the bus.

Figure 31 on page 54 shows the connections on the FC storage unit.

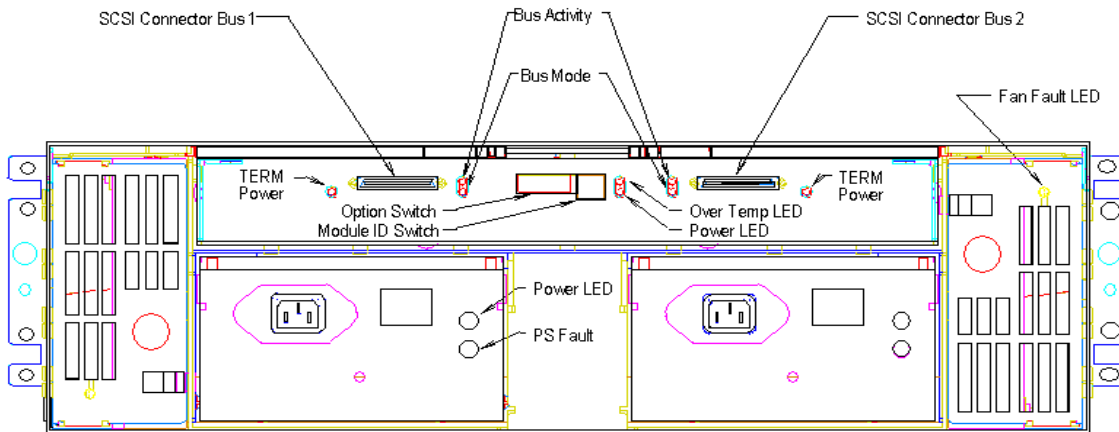


Figure 31. Fibre Channel Storage Unit (Rear view)

The Storage Unit supports:

- 16-bit, 4.5 GB disk drives
- 16-bit, 9 GB disk drives
- 16-bit, 18GB disk drives

The following figures (Figure 32 on page 55 through Figure 37 on page 60) show the different configurations that are available as more and more devices are added to the total subsystem. These figures also show the SCSI channel and ID that should be associated with each device, as well as the cabling connections from the control unit to the storage units.

Configuration 1: One IBM Fibre Channel Control Unit and one IBM Storage Unit.

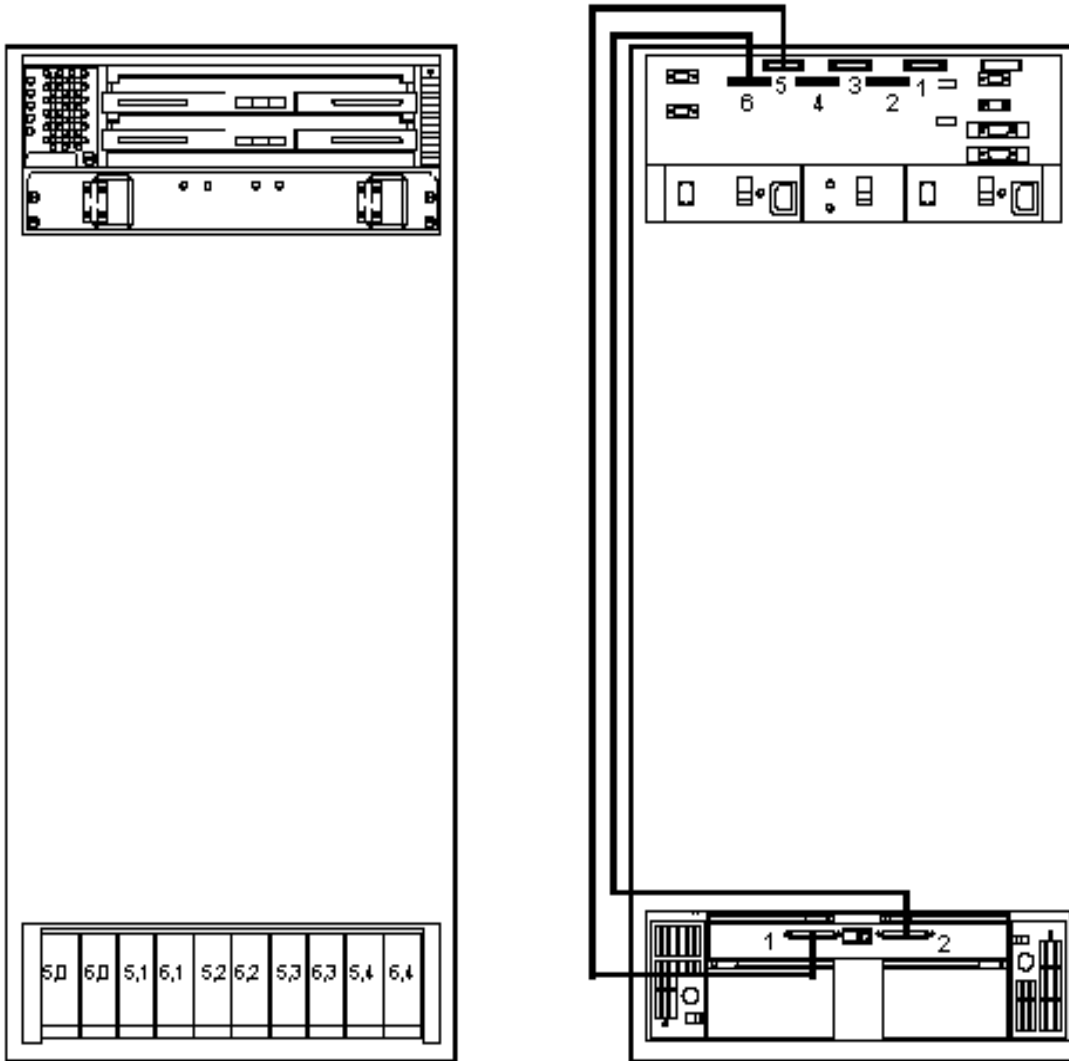


Figure 32. One Storage Drawer

Configuration 2: One IBM Fibre Channel Control Unit and two IBM Storage Units.

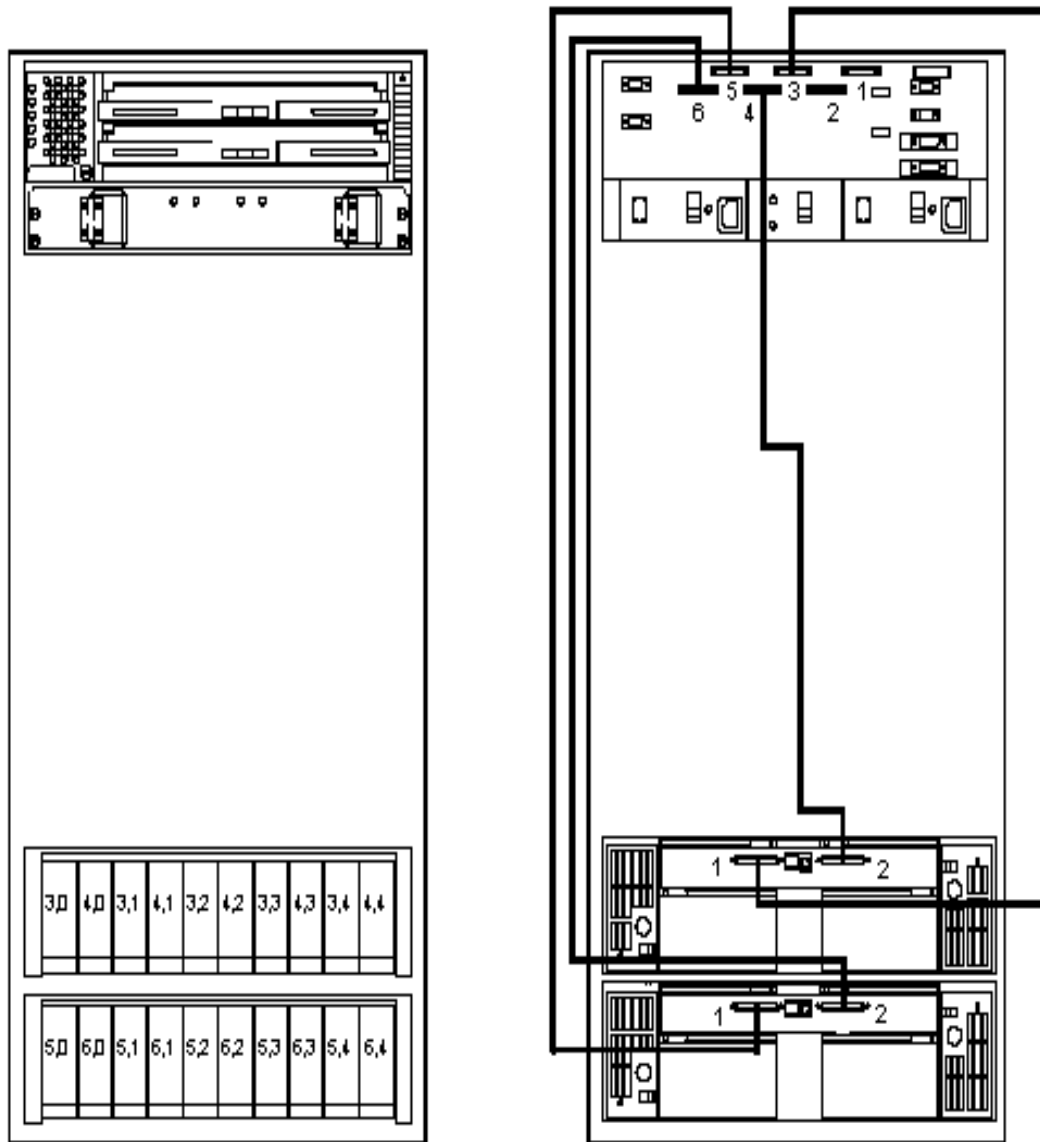


Figure 33. Two Storage Drawers

Configuration 3: One IBM Fibre Channel Control Unit and three IBM Storage Units.

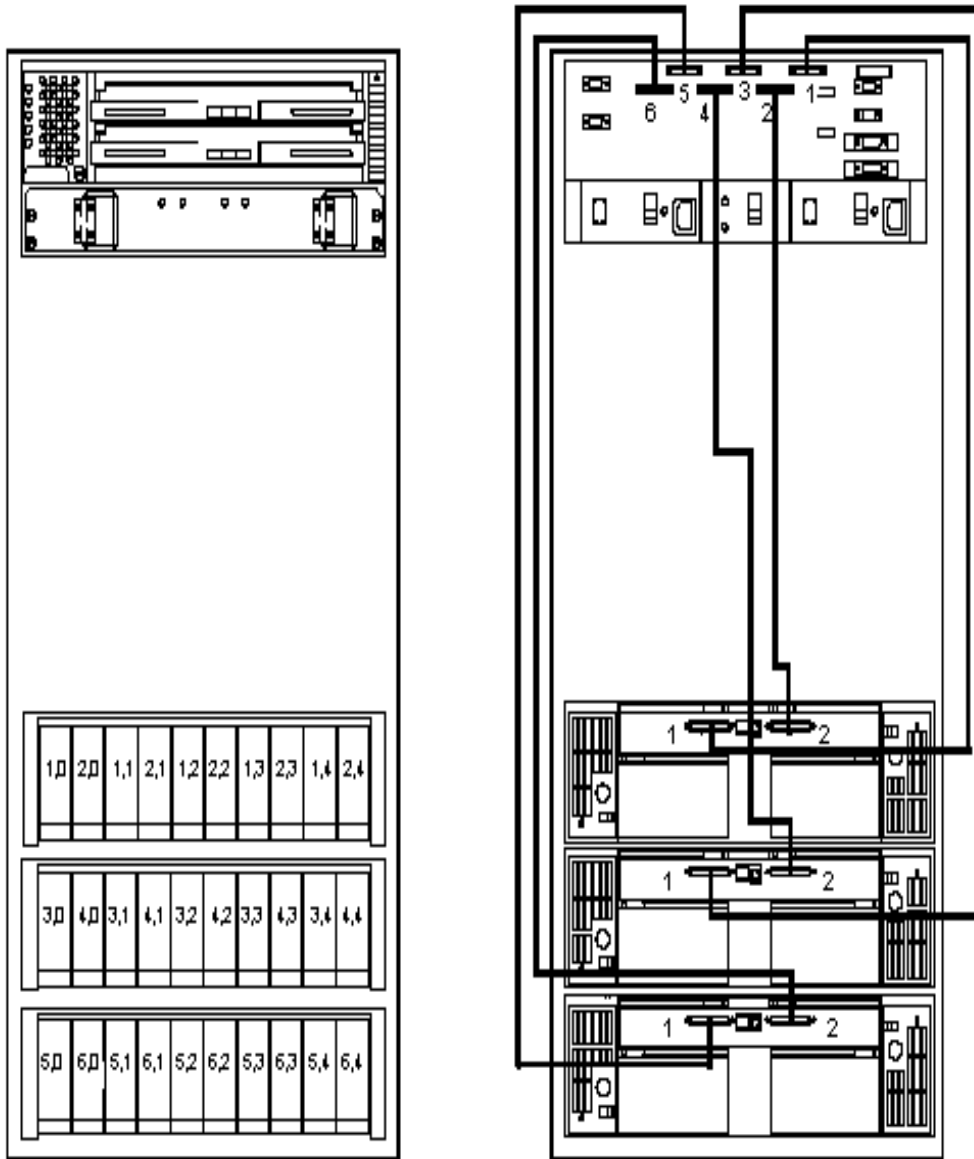


Figure 34. Three Storage Drawers

Configuration 4: One IBM Fibre Channel Control Unit and four IBM Storage Units.

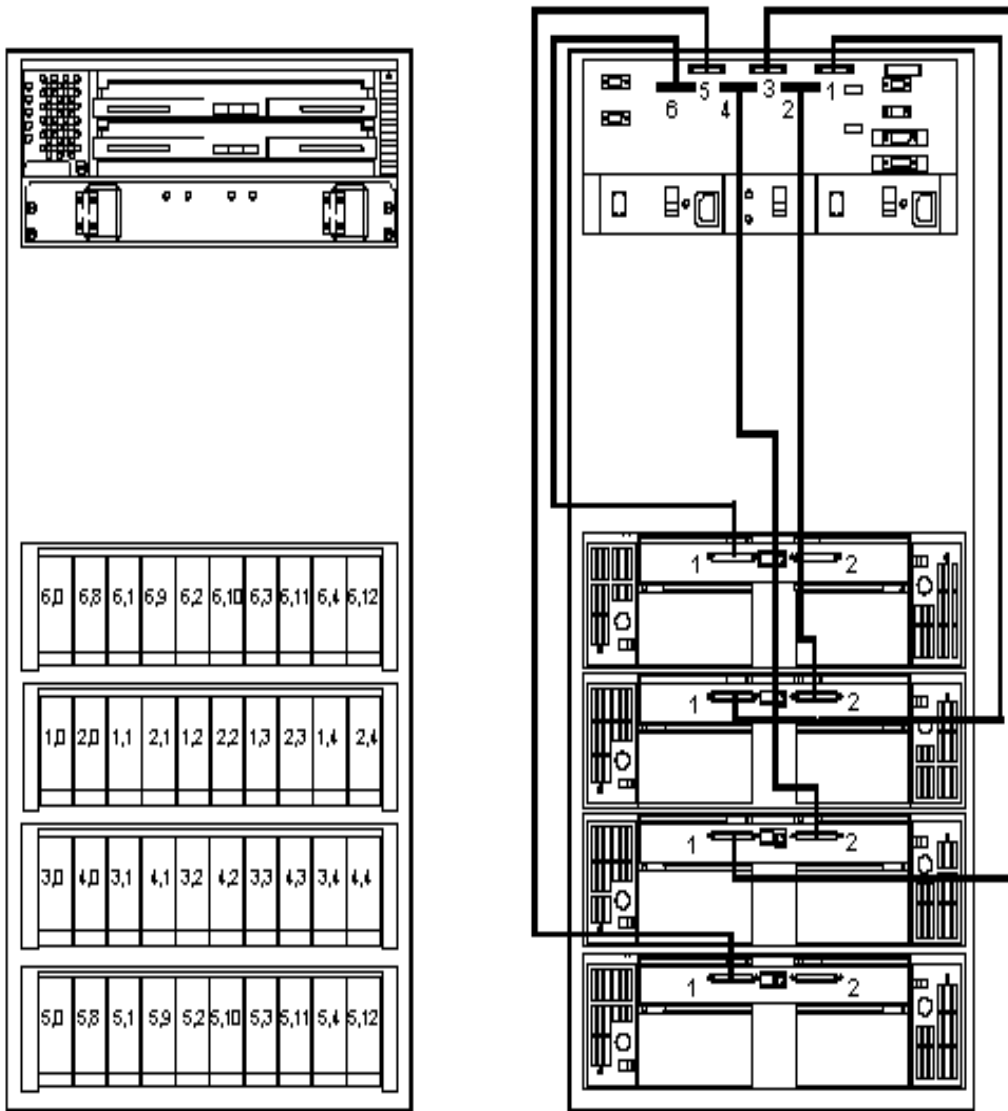


Figure 35. Four Storage Drawers

Configuration 5: One IBM Fibre Channel Control Unit and five IBM Storage Units.

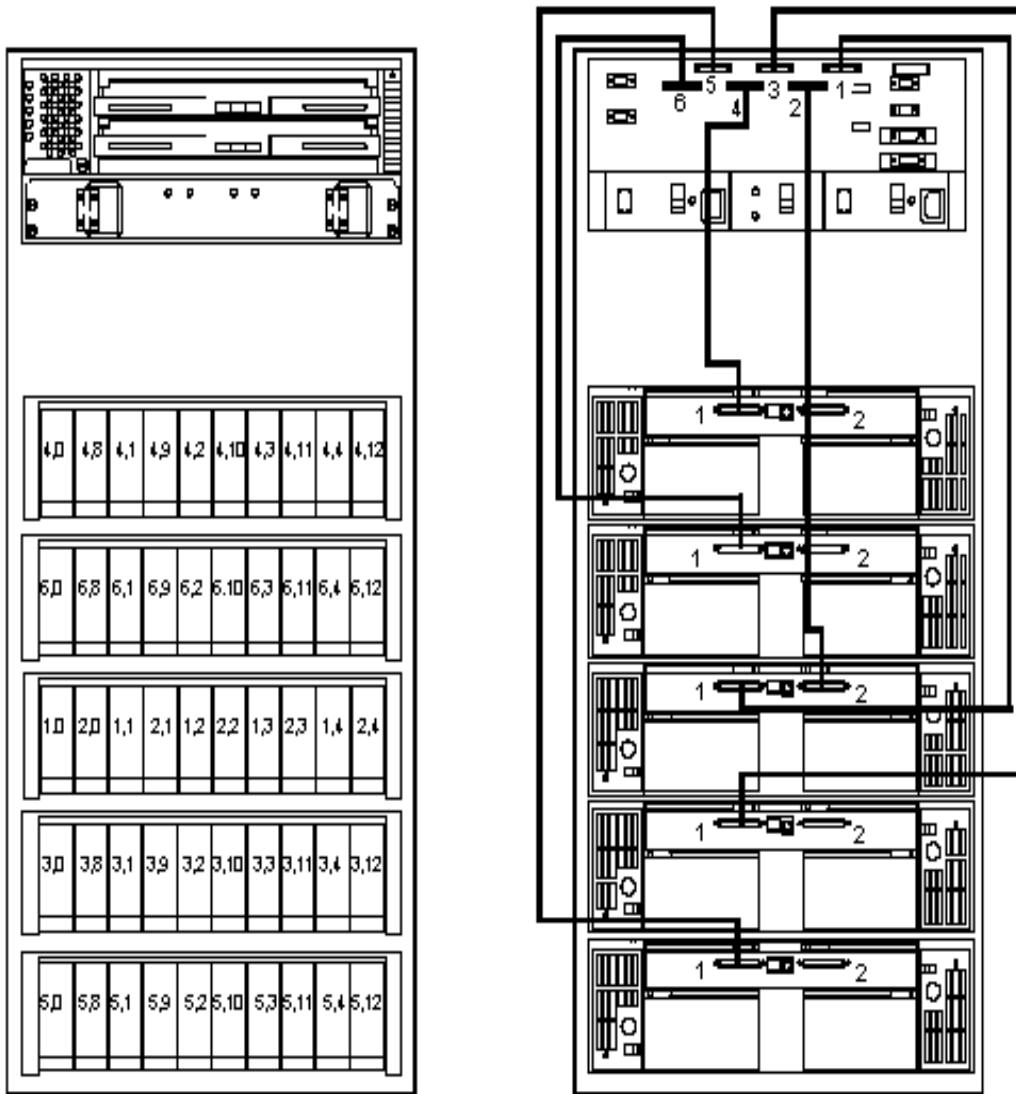


Figure 36. Five Storage Drawers

Configuration 6: One IBM Fibre Channel Control Unit and six IBM Storage Units.

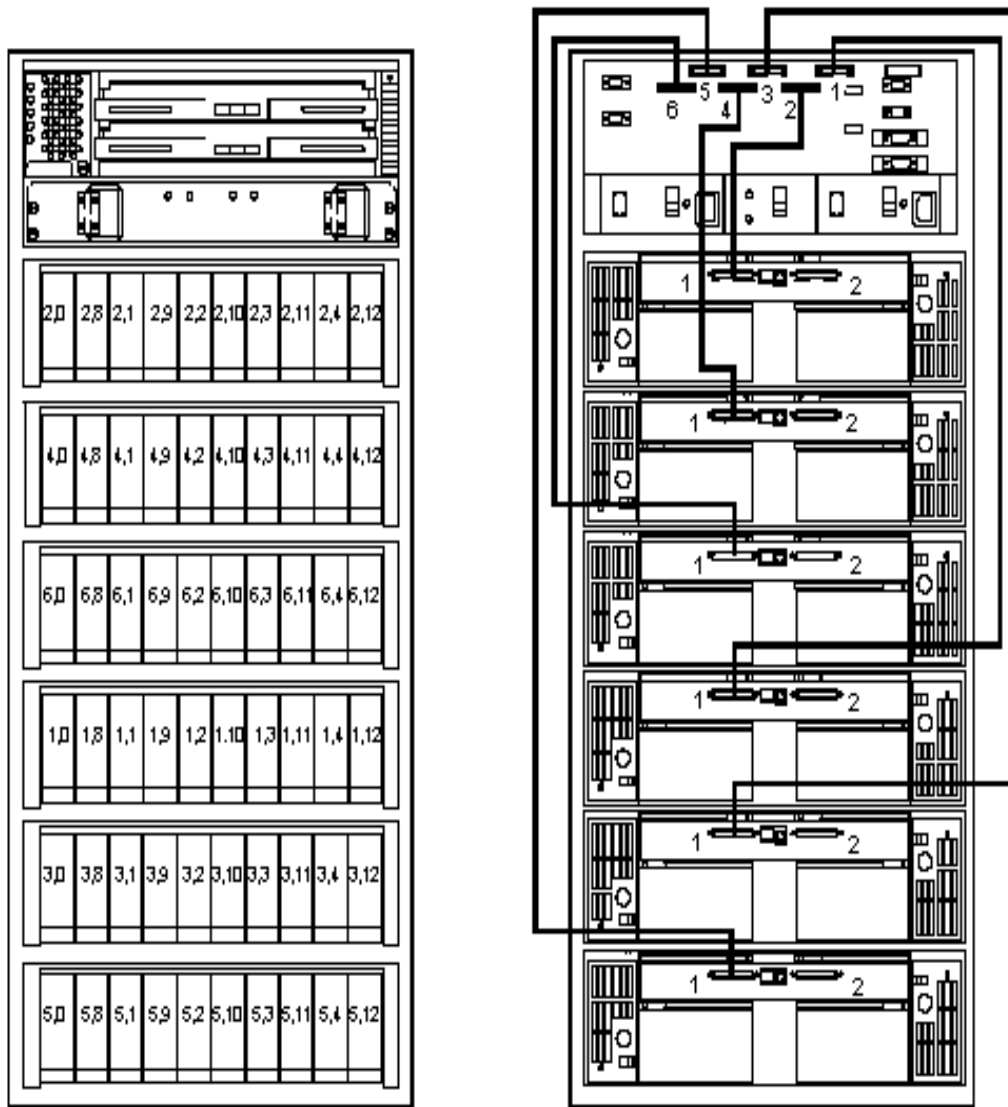


Figure 37. Six Storage Drawers

6.4 Host System Requirements and Installation Procedures

This section provides a description of the host system requirements and installation procedures needed to install the storage management software on your host system to manage and/or use the FC-RSS.

6.4.1 Windows NT System

The following hardware and software is required:

- At least an 800x600 resolution graphics card and a mouse
- A minimum adapter firmware version, depending on your adapter model
- One or two host machines with host module configurations supported by the software
- Windows NT Server 4.0 (for standard environment) or Windows NT 4.0 Enterprise Edition (for Cluster Server environment); this software was tested with Microsoft Service Pack 3
- A PCI Fibre Channel Adapter

IMPORTANT: For greatest level of I/O path protection, provide each controller in the Control Unit module with its own data path connection (that is, cable and host adapter) into the host system.

If fibre channel adapters are being used, you must set these up. Use the following procedure:

1. Go to the **HostAdapt** directory on the software's Installation CD.
2. Consult any readme file in that folder. Obtain specific instructions on what is needed to set up the PCI Fibre Channel Adapter.
3. Perform any required operation.

6.4.1.1 Installing the Software on a Standard Configuration

Assigning Drive Letters

To ensure SCSI adapter functionality, the installation procedure loads the Array driver before the native NT class driver. However, this means that the system will see the Control Unit's LUN first, before it sees any local SCSI drives, and will assign driver letters accordingly. Therefore, before you install this software, you must have "sticky" drive letters assigned to your existing local drives.

Note: The storage management software cannot be installed unless drive letters are assigned to existing drives.

Use the following procedure to assign "sticky" drive letters:

1. Start Disk Administrator (select **Start - Programs - Administrative Tools - Disk Administrator**), and view the disk configuration on your system. Check the number of partitions.
2. If you have at least two partitions (the boot partition and one other partition) and both of the following:
 - The last partition you created was created with Disk Administrator, *and*
 - The partition has a drive letter assigned and has a status of "Unknown" or has a file system on it,

Then you can continue with the installation procedure.

3. Create a new partition.

Note: If you do not have room for a second partition on your local system, you can use a Control Unit's LUN for the second partition. The Control Unit should have at least one LUN configured on it. You can access the Control Unit LUN without having installed the storage management software. Connect the Control Unit to the system and restart the system. The Control Unit LUNs appear in the Disk Administrator, and you will be able to create a second partition on a Control Unit LUN using the procedure below. Note that you can delete or modify the Control Unit LUN later without causing the boot partition drive letter to be reassigned.

1. Select a drive containing some free disk space.
2. Select **Partition - Create**. The partition is created (a drive letter is assigned, and the partition status is "Unformatted").
3. Select **Partition - Commit Changes Now**. The partition status changes to "Unknown", and sticky drive letters are assigned to all your existing partitions.

Installing the Software Manager

Use the following procedure to install the storage management software. You must have administrator privileges to access this software. If you have independent hosts, you must perform this procedure on both hosts. In the following examples, *c:* is designated as your boot drive and *d:* is designated as the CD-ROM device drive.

1. Insert the Installation CD into the *d:* drive.
2. Select **Start - Settings - Control Panel - Add/Remove Programs**.
3. When you see *d:\fcsm\setup.exe*, select **Finish**.
4. Follow the instructions on the screen to complete the installation procedure.

Note: Regarding `\program file\ibm\fcsm` on your boot drive: For the purpose of providing examples of the installation requirements, this will refer to the default directory on drive *c:*

5. Remove the CD and restart the system.
6. Verify that the storage management software is installed and the Disk Array Monitor service is running:
 - Select **Start - Programs. IBM Fibre Channel Storage Manager** should appear in the list.
 - Select **Start - Settings - Control Panel - Services**. The Disk Array Monitor should be running if this service's status is *Started*.

6.4.1.2 Installing the Software on a Cluster Server Environment

This section describes how to install the storage management software in the Multi-Host configuration, using Cluster Server software.

Installing the Hardware

Hardware installation involves installing host adapters and FC-RSS.

Installing the Host Adapters

The following considerations apply:

It is possible that you might be installing two adapters in each node to run two cables from each node to both controllers in a Control Unit. For example, if you are using single-channel host adapters and want the dual path configuration, you must install two host adapters in each node.

We recommend that you set that Host-Side ID on the host adapter in node A to 7 and the Host-Side ID on the host adapter in node B to 6.

If you are installing two host adapters in each node for the dual bus configuration, selecting the same Host-Side in each node may help prevent conflicts.

After that, install the correct driver for your host adapter.

Installing the FC-RSS

The FC-RSS must be installed by a trained IBM Service Representative.

Chapter 6.2, "Configurations" on page 50 shows a Fibre Channel connection using either a single or dual path configuration.

Important: Make sure you do not install the Cluster Server Software until told to do so in this chapter.

CAUTION:

Once you have connected the RAID Modules to both nodes, do not boot Windows NT on both nodes at the same time until you have installed the Cluster Server Software on at least one node. To suspend the booting, go to Windows NT; press the spacebar at the Operating System (OS) loader screen.

6.4.1.3 Installing the Storage Management Tools on Node A

Use the following procedure to install the storage management software. You must have administrator privileges to access this software. If you have independent hosts, you must perform this procedure on both hosts. In the following examples, *c:* is designated as your boot drive and *d:* is designated as the CD-ROM device drive.

Assigning Drive Letters

The installation procedures loads the Array driver before the native NT class driver. However, this means that the system will see the Control Unit's LUN first, before it sees any local SCSI drives, and will assign driver letters accordingly. Therefore, before you install this software, you must have "sticky" drive letters assigned to your existing local drives.

Note: The storage management software can not be installed unless drive letters are assigned to existing drives.

Use the following procedure to assign "sticky" drive letters:

1. Start Disk Administrator (select **Start - Programs - Administrative Tools - Disk Administrator**), and view the disk configuration on your system. Check the number of partitions.
2. If you have at least two partitions (the boot partition and one other partition) and both of the following:

- The last partition you created was created with Disk Administrator, *and*
- The partition has a drive letter assigned and has a status of "Unknown" or has a file system on it,

Then you can continue with the installation procedure.

3. Create a new partition.

Note: If you do not have room for a second partition on your local system, you can use a Control Unit LUN for the second partition. The Control Unit should have at least one LUN configured on it. You can access the Control Unit LUN without having installed the storage management software. Connect the FC-RSS to the system and restart the system. The Control Unit LUNs appear in the Disk Administrator, and you will be able to create a second partition on a Control Unit LUN using the procedure below. Note that you can delete or modify the Control Unit LUN later without causing the boot partition drive letter to be reassigned.

1. Select a drive containing some free disk space.
2. Select **Partition - Create**. The partition is created (a drive letter is assigned and the partition status is "Unformatted").
3. Select **Partition - Commit Changes Now**. The partition status changes to "Unknown", and sticky drive letters are assigned to all your existing partitions.

At this point, you must proceed with the installation of the software

Installing the Storage Management Tools

1. From Node A, insert the Installation CD into your CD-ROM drive.
2. Select **Start - Settings - Control Panel - Add/Remove Programs**.
3. Select **Install** and follow the instructions on the screen.
4. When you see d:\fscm\setp.exe select **Finish**.
5. Follow the instructions on the screen to complete the installation procedure, The default path is `|program file|ibm\fscm` on your boot drive.
6. When installation is completed, remove the CD and restart the system.
7. Verify that the storage management software is installed and the Disk Array Monitor service is running:
 - Select **Start - Programs. IBM Fibre Channel Storage Manager** should appear in the list.
 - Select **Start - Settings - Control Panel - Services**. The Disk Array Monitor should be running if this service's status is *Started*.

Setting Up The FC-RSS

After you have completed the procedures, shut down Windows NT on node A, but leave node A running (do *not* turn off the node).

6.4.1.4 Installing the Storage Management Tools on Node B

After you have verified the software installation on node A and configured logical units, install the software on node B using the same procedure used on node A. However, you do not need to configure any logical units.

Assigning Drive Letters

Before you install this software, you must have "sticky" drive letters assigned to your existing local drives.

Note: The storage management software can not be installed unless drive letters are assigned to existing drives.

Use the following procedure to assign "sticky" drive letters:

Boot Windows NT on node B.

1. Start Disk Administrator (select **Start - Programs - Administrative Tools - Disk Administrator**), and view the disk configuration on your system. Check the number of partitions.
2. If you have at least two partitions (the boot and one other partition) and both of the following:
 - The last partition you created was created with Disk Administrator, *and*
 - The partition has a drive letter assigned and has a status of "Unknown" or has a file system on it,

Then you can continue with the installation procedure.

3. Create a new partition.

Note: If you do not have room for a second partition on your local system, you can use a Control Unit LUN for the second partition. The Control Unit should have at least one LUN configured on it. You can access the Control Unit LUN without having installed the storage management software. Connect the FC-RSS to the system and restart the system. The Control Unit LUNs appear in the Disk Administrator, and you will be able to create a second partition on a Control Unit LUN using the procedure below. Note that you can delete or modify the Control Unit LUN later without causing the boot partition drive letter to be reassigned.

1. Select a drive containing some free disk space.
2. Select **Partition - Create**. The partition is created (a drive letter is assigned and the partition status is "Unformatted").
3. Select **Partition - Commit Changes Now**. The partition status changes to "Unknown", and sticky drive letters are assigned to all your existing partitions.

At this point, you must proceed with the installation of the software

Installing the Storage Management Software

1. From node B, insert the Installation CD into your CD-ROM drive.
2. Select **Start - Settings - Control Panel - Add/Remove Programs**.
3. Select **Install** and follow the instructions on the screen.
4. When you see d:\fcs\setp.exe select **Finish**.
5. Follow the instructions on the screen to complete the installation procedure, The default path is `program file\ibm\fcs` on your boot drive.
6. When installation is completed, remove the CD and restart the system.

7. Verify that the storage management software is installed and the Disk Array Monitor service is running:
 - Select **Start - Programs. IBM Fibre Channel Storage Manager** should appear in the list.
 - Select **Start - Settings - Control Panel - Services**. The Disk Array Monitor should be running if this service's status is *Started*.

Verify Drive Letters

Use the Windows NT Disk Administrator to verify that the drive letters assigned to the configured logical units are the same those assigned to node A.

Shut down Windows NT on node B, but leave node B running (do not turn off the node).

Installing the Cluster Server Software

1. Install the Cluster Server on node A.
 - Refer to *Microsoft Clustering for Windows NT Server Administrator's Guide* for the correct procedure to install the Cluster Server Software.
Note: During the installation, specify that you want to **Form a New Cluster**.
 - After the system restarts, leave node A up and running.
2. Install the Cluster Server software on node B.
 - Refer to *Microsoft Clustering for Windows NT Server Administrator's Guide* for the correct procedure to install the Cluster Server Software
Note: During the installation, specify that you want to **Join The Existing Cluster**.
 - After the system restarts, leave the node B up and running.

To verify the software installation, you must follow the next procedure:

1. On either (A or B), select Start - Programs - Administrative Tools (Common).
2. Click on Cluster Administrator.
3. In the "Cluster or Server Name" prompt, type either name of the cluster or the name/IP address of one of the nodes.
4. If the installation was successful, the computer names of both nodes appear on the left side of the Cluster Administrator window.

6.4.1.5 Starting the Software

You must have administrator privileges to access this software and follow the next instructions:

1. Start Windows NT if it is not already running. Select **Start - Programs - IBM Fibre Channel Storage Manager**.
2. The application should appear under **IBM Fibre Channel Storage Manager** (Configuration, Maintenance, Tuning, Recovery and Status).
3. If the FC-RSS is using an Independent Controller configuration, you need to perform the procedure below to indicate the configuration to the storage management software.
 - Start any storage management software application.

- When the Module Selections screen appears, highlight the module that has independent controllers, and select Exit.
- Click the box to **Independent Controllers?** and select **OK**.
- Return to the main Module Selection screen, and the **Indep. Cntrls?** column now says **Yes**.
- Assign names to the modules, if desired, and add any comments (you may want to add location details or unique information to further clarify the module name).

6.4.2 HP - UX System

The following hardware and software are required:

- HP hardware series K
- HP-UX 10.20 or HP-UX 11.0
- A minimum controller firmware version depending on your controller model
- An X-terminal with at least an 800x600 resolution graphics card and a mouse
- One or two host machines supported by the software
- An HP A3404A FC adapter

Use the *ioscan -f* command to scan the I/O data connection and list all the hardware paths and disk drives for the devices attached to the system. This process lists the targets (host-side IDs) of the devices connected to and recognized by the host adapter. There should be a target listed for each controller in the Control Unit. If not, check the cable connections.

6.4.2.1 Installing The Software on HP-UX

Use the following procedures to install the storage management software.

Attention: Make sure that all controllers for all control units are optimal before installing the storage management software. Installing with failed controllers can result in problems with RDAC protection, unrecognized controller pairs or modules, and even operating system crashes.

You must have superuser privileges to install and access this software. Be sure that other users are not on the system while you are installing this software.

The program, *swinstall*, may not check for some items when it determines space requirements. Therefore, make sure that there is enough space in the root disk or root volume group for the following:

- A new copy of an OS kernel (at least 7 to 9 MB in /stand)
- Temporary files that are created and removed during installation (at least 2 or 3 MB in /tmp is recommended).

1. Start HP-UX if it is not already running.
2. Insert the installation CD-ROM.
3. Type:

```
mount -o cdcase /dev/dsk/<cd-rom device node> /cdrom
```

If you do not know the device node for the system's CD-ROM, use the *ioscan -fkn* command to list your system's devices.

Type:

```
swinstall -s /cdrom/fcsm/fcsm
```

4. Highlight **FCSM**

5. From the top menu, select **Actions - Mark for Install**.

6. From the top menu, select **Actions - Install** (analysis).

You can select Log File to view a description of the analysis phase. You could also select **Actions - Show Description of Software** for details about this software.

7. When you see "The Analysis Phase succeeded." select **OK** to begin the installation.

8. Select **Yes** at the next two screens to continue with the installation.

9. Select **Done** when you see "The Execution Phase succeeded."

10. Select **OK** to confirm that you are ready to reboot the operating system. The reboot will be immediate.

11. When the reboot is completed, view the syslog.log file to verify that additional installation tasks were performed.

12. Manually reboot the system again to ensure that the NVSRAM settings were changed. You need to perform the reboot by typing:

```
/sbin/shutdown -ry
```

13. To verify the installation of this software, perform the following steps:

- at the command line, check the list of files to see that the following file sets are included.

- Type:

```
swlist -v FCSM
```

List of file sets:

- FCSM-EXE

- FCSM-MAN

- FCS-FW

- Type:

```
swlist -v FCSDVR
```

List of file sets:

- FCSDISC3

- At the command line,

- Type:

```
swverify -v FCSM
```

Verify that the end of the file contains "Verification succeeded", and that there are no error messages. If there are any error messages, reinstall this software by using:

```
swinstall -s /cdrom/fcsm/fcsm
```

- At the command line type:

```
ps -ef | grep arraymon
```

A confirmation line should show the array monitor (*arraymon*) as a working process.

- At the command line type:

```
ps -ef | grep rdaemon
```

A confirmation line should show two RDAC daemons as working processes.

6.4.3 Sun Solaris System

The following hardware and software are required:

- At least an 800x600 resolution graphics card and a mouse
- One or two host machines supported by the software
- A Jaycor FC64-1063-IBM
- Solaris 2.5.1, or 2.6 operating system installed on one of the following:
 - The host machines connected by fibre channel cable to the FC-RSS.
 - LUN 0 of a control unit, if you want to use the modules as a boot device. This involves preparing the host system to boot from the control unit and installing all your software.

The Solaris operating system has patches available to ensure optimum performance of the storage management software within its environment. Install the recommended patches before installing the storage management software. You must reboot the operating system, using *boot -r*, for the changes take effect. Therefore, make sure that other users are not on the system when installing the patches. Do the following before installing the software:

1. Determine the Solaris version by typing:

```
uname -a
```

The version is specified as: SunOS <hostname>

- 5.5.1 (for Solaris 2.5.1)

- 5.6 (for Solaris 2.6)

2. See Chapter 1 for instructions on how to install the correct patch on the host system.
3. After installing the appropriate patches, reboot the Solaris system for the changes to take effect.

6.4.3.1 Installing the Software on a Standard Configuration

Use the following to install the storage management software.

You must have superuser privileges to install and access this software. Be sure that other users are not on the system when installing this software.

Make sure that the CD-ROM device is attached when you boot the system.

1. Insert the Installation CD into the CD-ROM drive.

2. Type:

```
cd /cdrom/fcsm
```

3. At the system prompt, type

```
pkgadd -d FCSM
```

4. Follow the instructions on the screen to install the package provided on the CD-ROM.
You will see a message that the installation process includes executing certain scripts with administrator privileges.
5. Select **Yes** at the prompt "*Do you want to continue with the installation on <FCSM>?*"
It may take 1 or 2 minutes before you notice certain tasks occurring. The readme file is then available for you review.
6. Before reviewing the readme file, quit *pkgadd*.
7. Did the message say you needed to reboot the system?

No	Yes
Go to "Verifying The Installation"	Continue with Step 8.

8. Shutdown the system by typing:

```
shutdown -y -i0 -g0
```
9. Reboot the system by typing:

```
boot -r
```
10. When "The system is ready" appears, do you also see the message "NVSRAM settings are correct"?

No	Yes
Continue with Step 11	Go to "Verifying The Installation"

11. If the NVSRAM message says that settings were changed, shut down and reboot the system again by typing:

```
shutdown -y -i0 -g0
```



```
boot -r
```
12. After the system reboots, go to "Verifying the Installation"

6.4.3.2 Verifying the Installation

To verify the installation of this software, perform the following steps:

1. At the command line, type:

```
pkinfo -l FCSM
```


A list should appear, showing statistics (version information, installation date, file information, etc.) about the software. The status field reads "completely installed" if the installation was successful.
2. At the command line, type:

```
ps -ef | grep arraymon
```


A confirmation line should show the array monitor (*arraymon*) as a working process.
3. At the command line, type:

```
ps -ef | grep rdaemon
```

A confirmation line should show the RDAC resolution/restart daemons as two working processes.

6.4.4 RS/6000 System

The following hardware and software is required:

- A minimum controller firmware version, depending on your controller model
- One machine with host module configuration supported by the software
- AIX operating system Version 4.3.2 or higher

6.4.4.1 Installing The Software On AIX Version 4

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

Notes:

1. Do not run the diagnostics on the FC-RSS at any stage during the installation of the FC-RSS software. Wait until installation is complete before running diagnostics.
2. Controller microcode and EEPROM files are distributed on the FC-RSS microcode diskette included with the FC-RSS subsystems and with additional controllers.

Downloading of array controller microcode and EEPROM files is normally required to update microcode. Read and follow any instructions shipped with the FC-RSS 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 FC-RSS hardware and software for the first time on AIX Version 4 are:

- Ensure that there is at least 8 MB of free space in the **/tmp** directory.
- Ensure that the FC-RSS is not connected to the host system at any time before installing the storage management 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 off power to the FC-RSS after removing the drive definitions from the system, and leave it off until you have installed the storage management software.

- Ensure that the storage management software is not already installed by using the command:

```
lslpp -l "devices.fcp.disk.array.*"
```

Installation

The task can be accomplished through SMIT.

1. Insert the AIX Version 4 install media into the appropriate media drive.
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.
3. During the installation process, the screen will display messages while the storage management 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 FC-RSS to the host system and turn the power on to the FC-RSS.
6. Restart the host system.
7. 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 7. This step puts the array controllers into dual-active mode.
8. Ensure that all disk array controllers (DACs) are now active.

Note: If the instructions shipped with the FC-RSS Controller Microcode diskette require downloading of microcode and EEPROM files, those actions would begin at this point.
9. Run diagnostics to each DAC on the system to ensure all paths and hardware are functioning correctly.

6.4.4.2 Installing a Storage Management Software Update

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

Action

Attention: The download process requires that all activity on the FC-RSS be halted. Make sure that all file systems are unmounted, vary off the volume groups, and all processes which access the FC-RSS are halted on all machines attached to the FC-RSS before proceeding.

1. Using SMIT, remove all devices associated with the FC-RSS by entering the following SMIT fast path:

```
smit fraidmgr
```

2. Insert the new software media in the appropriate device driver.
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** 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. Reboot the RS/6000.

6.4.4.3 Verifying the Software Installation

Perform the following to verify the software installation.

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

For AIX Version 4, verify the 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 software is in the applied state.

2. Select the Disk Array Router Configuration option from the FC RAIDiant Disk Array Manager (storage management software) menu in SMIT.

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 Disk Array Device listed for each array subsystem you have installed. The number displayed depends on how many banks are in your subsystem. Also, each Control Unit 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 an FC-RSS or the dual active 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 software was installed, the disks will be identified as Other FC SCSI Disk.

Chapter 7. Configuration

This section covers the post-installation process, including allocating new storage, storage expansion, and modifying storage attributes.

7.1 Solaris, HP-UX and Windows NT

This section covers the post installation process, including allocating new storage, storage expansion, modifying logical unit number (LUN) and Drive Group Attributes, and networking for Solaris, HP-UX and Windows NT.

7.1.1 Storage Allocation

Prior to allocating disks for storage, determine how much current and future storage is needed. Also, decide how many, if any, hot spares will be required to increase the level of subsystem availability.

Important: When using the IBM Storwatch Fibre Channel RAID Specialist or a host-RAID configuration with more than one host, it is possible to open up more than one configuration session. The management software does not provide resource locking; thus, you must be sure that you are using only one configuration session at a time, or else the drive/LUN could be in use and could cause problems for one of the hosts.

7.1.1.1 Drive Groups

The drive group is an associated set of physical disks that can have storage allocated to create LUNs. There are three types of drive groups:

- Unassigned drive group. These are the disks in the subsystem which have not been allocated to a specific drive group for allocation into a LUN or hot spare.
- Assigned drive group. This group contains disks that are ready to be allocated into LUNs, or that already contain configured LUNs.
- Hot Spare drive group. These disks make up the group containing only hot spares. Note that the maximum number of hot spare disks per subsystem is equal to the number of SCSI channels in use.

Drive groups can be dynamically increased in capacity, while IO operations are taking place to the drive group. However, the LUNs contained within the drive group *cannot* be increased in size.

7.1.1.2 LUN Creation

A LUN represents an allocated and configured amount of storage from an Assigned drive group. It is in turn presented to the host by the RAID module as a physical disk.

LUNs can be created using two methods:

- Creating a new LUN from disks belonging to an Unassigned drive group. This mode creates a new drive group and allows the specification of a RAID level and the number of drives within that drive group.
- Allocating unused space of an existing Drive Group. This mode does not allow the specification of the number of drives within the drive group or within the RAID level. The RAID level is determined by the drive group and is not

specific to each LUN. Chapter 7.1.2, “Modifying LUN and Drive Group Attributes” on page 80 describes how to setup and change LUN sizes.

LUN Creation from Unassigned Drive Group

To create a new LUN from unassigned drives, you must have drives remaining in the Unassigned drive group; otherwise, this option is grayed out.

The following procedure will create a LUN from the unassigned drives group.

1. Launch the Configuration application. Notice that both controllers are listed. If this is the first time the Configuration application has been run, a 10 MB LUN is created. This 10 MB LUN is the default LUN that is automatically created when during initial installation or after the configuration is reset.
2. Select the **Create LUN** button, or right-click on the **Unassigned Group**, and select **Create LUN**.
3. Select the desired RAID level depending on the availability and performance requirements. The RAID level setting might change the Number of Drives setting, or may be limited by the number of drives available. For example, you must have an even number of drives for RAID 1, while you need at least 3 unassigned drives for RAID 3 or 5 LUN.
4. Select the number of drives that are required for the new LUN. The drives are automatically assigned unless the value is changed by selecting **Options**.
5. Set the number of LUNs to be initially created. The default settings will result in LUNs which will occupy all of the available space in the drive group.
6. Select **Options**, and a screen like that shown in Figure 38 will appear.

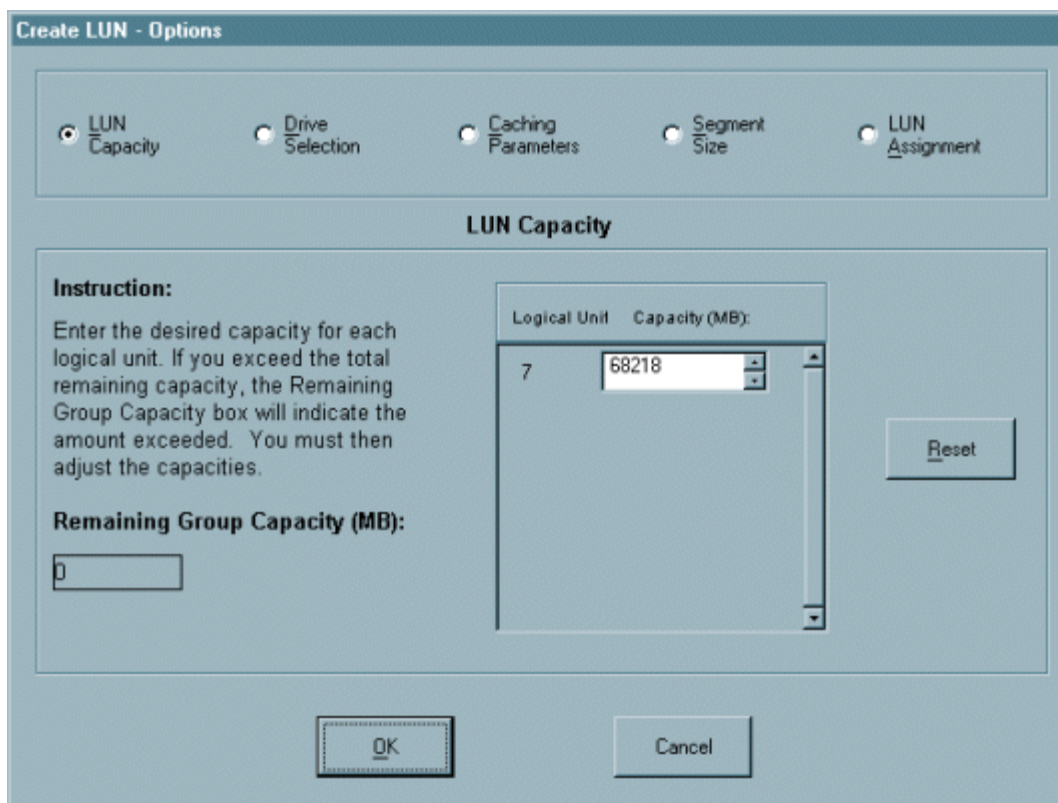


Figure 38. Create LUN, Options

7. Set the capacity of the initial LUNs in MB.
8. Select **Drive Selection**. Choose the number of drives that are equal to the number of drives specified in step 4 which results in Figure 39. For performance and availability choose drives that are on different channels.

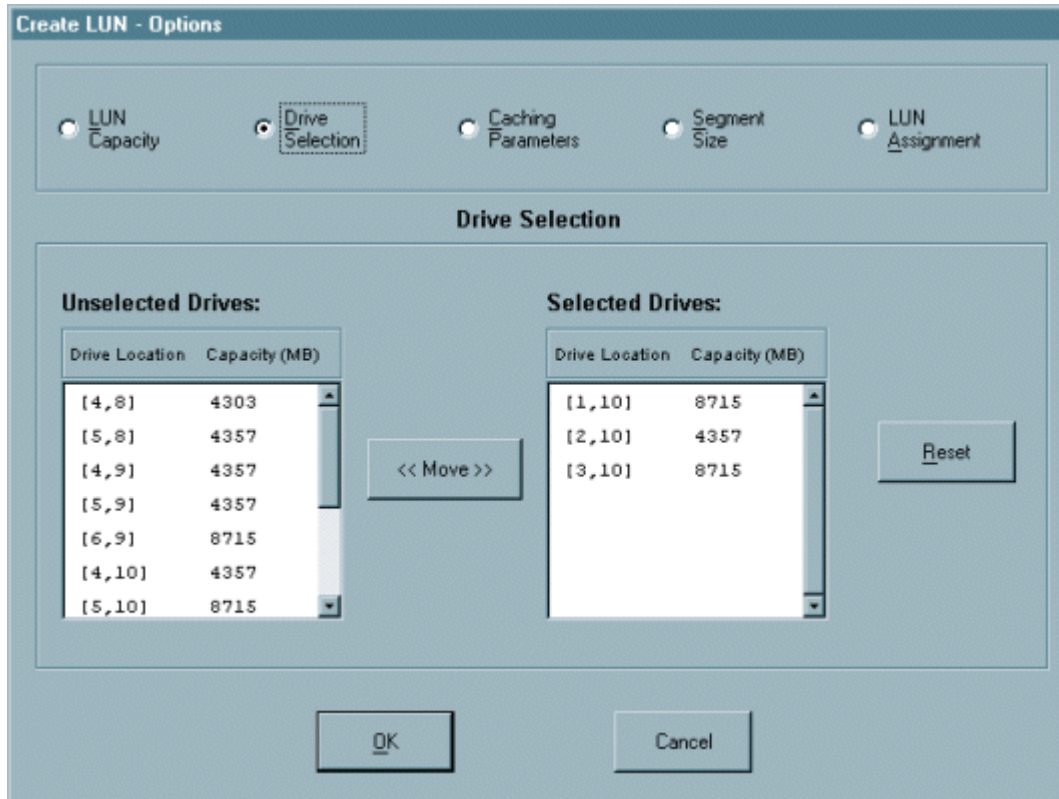


Figure 39. Creating a LUN, Drive Selection

9. Select **Segment Size**. Specify the segment size depending on the type of data and I/O transactions that are required.
10. Select **LUN Assignment**. This option will allow the specification of which controller will have access to the LUNs.
11. Select **Caching Parameters**. The first two choices, *write caching* and *write cache mirroring*, determine whether or not to cache write transactions and whether or not to mirror the other controller's write cache. The third option, *cache without batteries* determines if the controller will use cache while the battery is either not installed or not charged.
12. Select **OK** to return to the main LUN creation screen. An information box will pop up if more than six drives are configured in the same drive group. Select **OK**.
13. Click the **Create** button, and select **OK** to confirm.

At this point, the newly created LUNs are displayed in the main Configuration window, and the status will be Formatting as shown in Figure 40 on page 78.

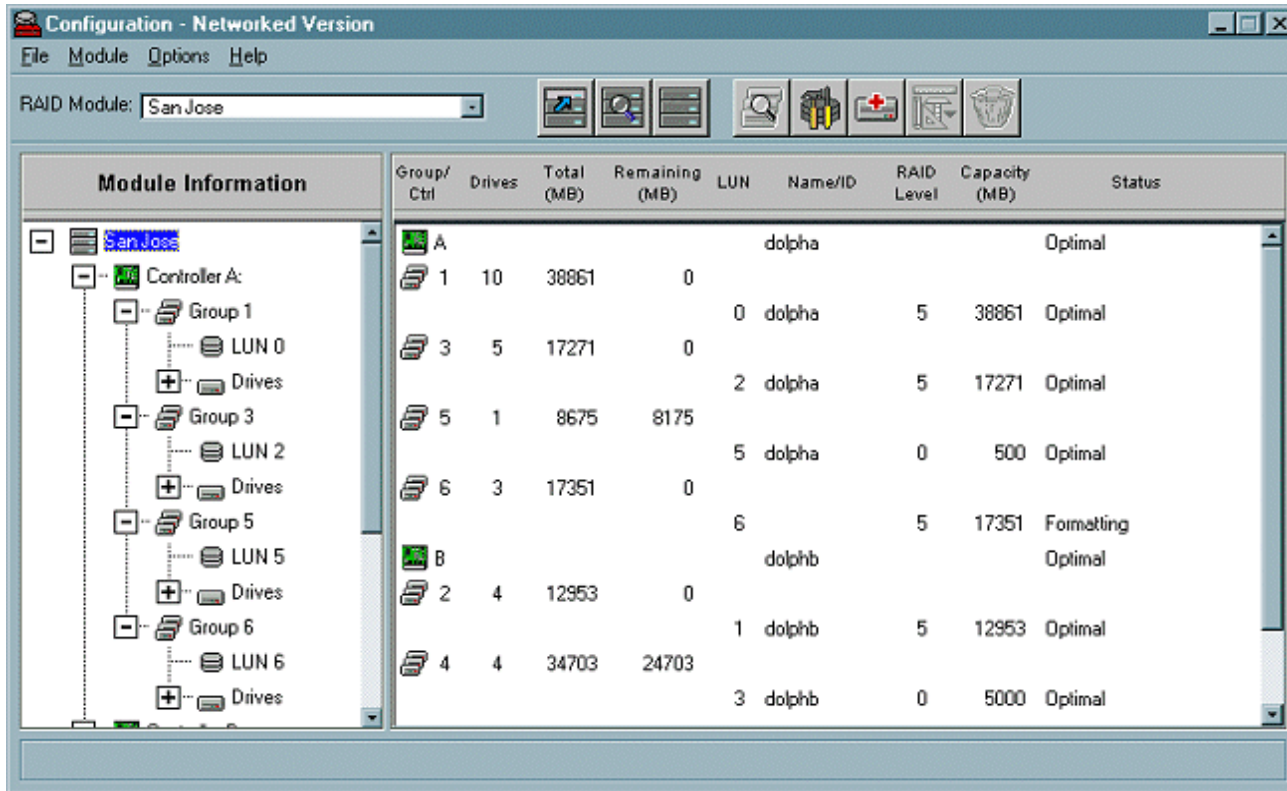


Figure 40. Configuration, One LUN Created

Creating a LUN from an Existing Drive Group

This procedure provides the steps necessary to create a new LUN from a previously created drive group.

1. Right-click on the drive group to which you want to add the LUN.
2. Select **Create LUN** which will result in Figure 41 on page 79.

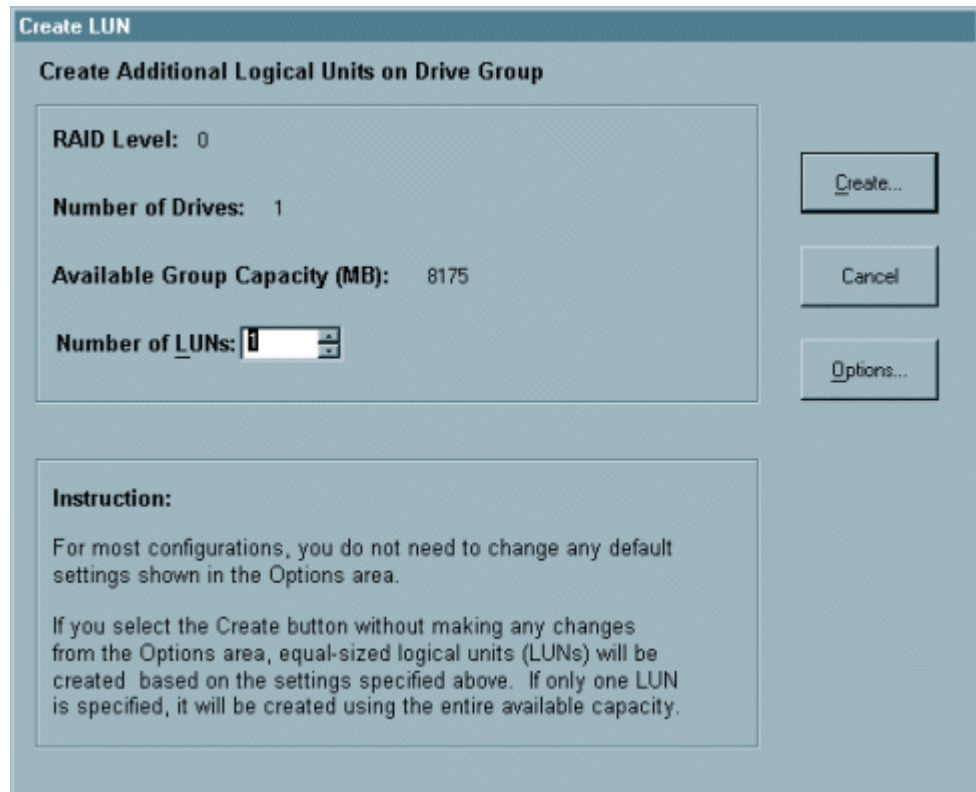


Figure 41. Creating a LUN from an Existing Drive Group

3. Proceed with the steps listed under section Chapter , “LUN Creation from Unassigned Drive Group” on page 76. The options, **Drive Selection** and **LUN Assignment**, are grayed out. The new LUN will have the same RAID level, drives, and controller assignment as the selected drive group.

7.1.1.3 Hot-Spares

Hot spares add another layer of protection to the data on your RAID modules. Hot spares automatically replace failed drives in your LUNs. You can only create hot spares if you have unassigned drives.

Attention: A hot spare can only replace a drive of equal or lesser capacity. If your subsystem contains drives of multiple sizes, always choose a drive of the larger capacity to become a hot spare.

To allocate a hot spare, perform the following steps:

1. Right-click on the **Unassigned** drive group and select **Create Hot Spare**, or click on the **Create Hot Spare** button.
2. Select the number of hot spares to be created. You can create as many hot spares as your configuration has channels. This can also be limited by the number of unassigned drives in the system.
3. Select **Options**, which will result in the display shown in Figure 42 on page 80 from which you can choose which drive(s) will become hot spares.

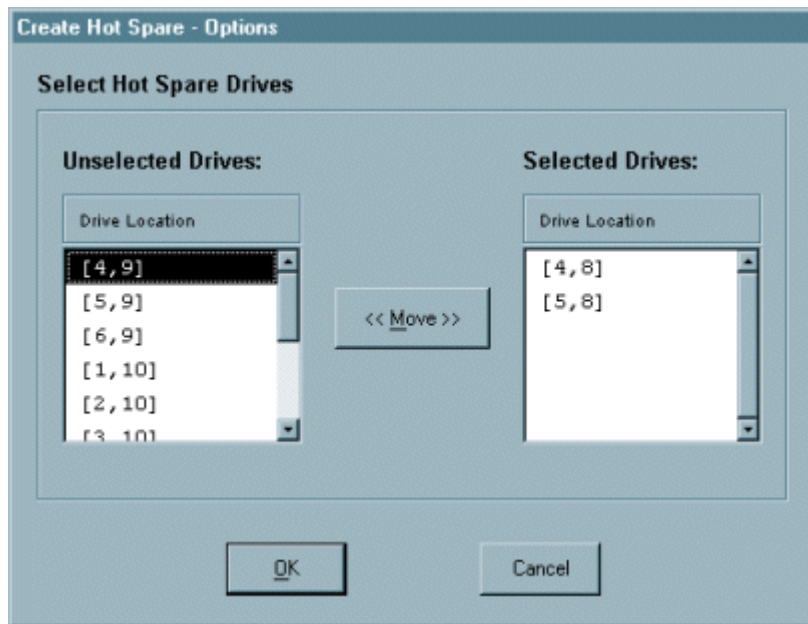


Figure 42. Creating a Hot Spare - Options

4. Select the drives to become hot spares.
5. Select **OK**, and then **Create**.

7.1.2 Modifying LUN and Drive Group Attributes

For an existing LUN or drive group, it is possible to change its characteristics. LUNs and drive groups can be modified in the following ways:

- Change the RAID level. The RAID level of the drive group can be dynamically changed. All LUNs within the drive group will all have the same RAID level. You may be limited in the new RAID level by the number of drives in the drive group.
- Add capacity. You can increase the number of disks in the *drive group*. This does not increase the capacity of the LUNs contained within the drive group. Future LUNs can take advantage of the additional capacity of the drive group.
- Modify segment size. You can change the segment size of an individual LUN. When changing the segment size of a LUN, you can only increase or decrease the size by a factor of 2 per conversion. Therefore, if you wanted to increase from a segment size of 32 KB to 128 KB, you would first have to finish converting to 64 KB and then increase to 128 KB.
- Defragment the LUNs of the drive group. This option may increase the total available capacity.
- Change the cache parameters. You can change the basic write-cache policies: write caching, write cache mirroring, and cache without batteries.
- Change the reconstruction rate. You can change the reconstruction rate assigned to individual LUNs.

You *cannot* make the following changes without deleting a LUN and recreating it:

- Change the LUN capacity

- Decrease the number of drives in the drive group

7.1.2.1 Changing RAID Level

Use this procedure to change the RAID level of all the LUNs in a drive group.

1. From within the configuration application, right-click the Drive Group you want to modify.
2. Select the **Modify Group/LUNs** option, then select **Modify RAID Level**.
3. Choose the RAID level you want for the drive group.
4. Click OK to make the change.

As the controller changes the RAID level, the configuration display will show a status of "Modifying" for the drive group and for the individual LUNs. The status will give you values indicating both the percentage of the individual LUN modified, and the percentage of the overall drive group modified. When the status of all elements changes to Optimal, the operation is complete.

7.1.2.2 Adding Drives to a Drive Group

Use this procedure to increase the capacity of a drive group by adding disks to the group.

Note: Adding drives does *not* increase the capacity of the LUNs in the drive group. The LUNs are spread out over the disks within the drive group. However, after adding disks, the drive group will have capacity for additional LUNs.

Perform the following procedure:

1. Start the configuration application.
2. Right-click the Drive Group you want to modify.
3. Select **Modify Group/LUNs**, and then select **Group Capacity**.
4. Select the drives you want to add to the drive group. You should pick drives of the same size as the current drives in the group.

As the controller adds the disks, the configuration display will show a status of "Modifying" for the drive group and for the individual LUNs. The status will give you values indicating both the percentage of the individual LUN modified, and the percentage of the overall drive group modified. When the status of all elements changes to Optimal, the operation is complete.

7.1.2.3 Modifying Segment Size

Use this procedure to modify the segment size of individual LUNs. You do not have to change all the LUNs within a drive group. Follow these steps:

1. Start the configuration application.
2. Right-click the Drive Group or LUN that you want to modify. Select **Modify Group/LUNs**, then select **Modify Segment Size**.
3. One at a time, select the LUNs you want to modify, then pick the new segment size. Note that you can only increase or decrease the segment size by a factor of two.

As the controller changes the segment size, the configuration display will show a status of "Modifying" for the drive group and for the individual LUNs. The status

will give you values indicating both the percentage of the individual LUN modified, and the percentage of the overall drive group modified. When the status of all elements changes to Optimal, the operation is complete.

7.1.2.4 Defragmenting the Drive Group

Defragmenting the drive group can free up space after the deletion of a logical unit from the drive group. To defragment the drive group, perform the following procedure:

1. Start the configuration application.
2. Right-click the drive group you want to defragment. Select **Modify Group/LUNs** then select **Defragment**.
3. A message is displayed stating one of the following possibilities:
 - That there is no space to recover in the drive group you selected. Click **OK** to return to the Main Configuration screen. You are finished with this procedure.
 - That there is space to be recovered on the selected drive group. The message indicates how much space can be recovered. Click **OK** to recover the space or click **Cancel** to return to the Main Configuration screen.

If there is space to recover and you clicked **OK**, the software begins to defragment the drive group. As the controller defragments the drive group, the configuration display will show a status of "Modifying" for the drive group and for the individual LUNs. The status will give you values indicating both the percentage of the individual LUN modified, and the percentage of the overall drive group modified. When the status of all elements changes to Optimal, the operation is complete.

7.1.2.5 Deleting LUNs

When you delete a LUN, any data in the LUN is lost. If you delete all the LUNs in a drive group, the drives in the group return to the unassigned drive group.

Attention: Do not delete all the LUNs in the RAID module (FC_RSS) and exit configuration without creating at least one LUN. Failure to do so can result in losing communication with the RAID module. Before you delete a LUN, check your *Installation and Support Guide* for special restrictions on LUN deletion. This includes deleting partitions and unmounting volume groups or file systems.

To delete a LUN, perform the following procedure:

1. Perform the necessary operating specific steps to prepare for LUN deletion. See your *Installation and Support Guide* for information.
2. Launch the Configuration application.
3. Select the drive group containing the LUN you want to delete in the Module directory tree.
4. Press the **Delete** key, select **Options** then **Delete** from the drop-down menus, or right click on the drive group containing the LUN, then select **Delete**.
5. Select the LUNs you want to delete, or click **Select All** to delete all the LUNs in the drive group.
6. After the LUNs are deleted, you need to perform the necessary operating specific steps to delete a drive. This may involve rebooting the system, or

rerunning configuration utilities. See your operating system documentation and the *Installation and Support Guide* for further information.

7.1.2.6 Resetting the Configuration

Use this option only as a last resort if your configuration is totally inaccessible, or if you want to completely redo your configuration.

When you reset the configuration, the selected RAID Module, as well as *all* drive groups and their LUNs are reset back to a default configuration, based on NVSRAM settings specified in the controller. The default is a six-disk drive group with a 10 MB LUN.

- Because deleting LUNs causes data loss, you should back up data on *all* the drive group/LUNs in the RAID Module. Deleting LUNs also deletes any filesystems mounted on the LUNs.
- First you must stop all I/Os to the affected RAID Module. You should also ensure that no other users are on the system, including other instances of the *Configuration* application.
- You cannot reset the configuration on a RAID Module with an independent controller configuration.
- This operation could also fail if the storage management software cannot gain exclusive access to the drive groups/LUNs, such as when filesystems are mounted.

To reset the configuration on the RAID Module, perform the following procedure:

1. From the configuration application, select **File**, then **Reset Configuration**.
2. Select **OK** to confirm that you want to reset your configuration.

Attention: Resetting the configuration will destroy all data on the LUNs in your RAID module, this is your last chance to Cancel!

3. Select **OK**, to confirm again that you want to reset your configuration. A default configuration appears in the main Configuration screen.
4. You will have to redefine all your LUNs and drive groups using Create LUN.

Note: Your operating system may have additional requirements to complete the configuration process so that it can recognize the new LUNs, including adding drives and possibly restarting the operating system. Refer to the *Installation and Support Guide* specific to your operating system and to the appropriate system documentation for specific details.

7.1.2.7 Windows NT Drive Letter to LUN Mapping

Under Windows NT, it can become confusing determining which drive resides on a specific FC-RSS LUN. In order to solve this problem, use the *lad* command.

Usage: *lad -a*

The program is run from the installation directory and the *-a* parameter can only be used under windows NT. The results of running *lad -a* are listed in Figure 43.

```

*****
RAID1 (2 Controller(s))

Series3 Controller 1T81916722, SystemName=Drive0 - (port=2,path=4,SCSI Id=27)

Disk[lun0]=0
Disk[lun1]=1
Disk[lun2]=2
Disk[lun3]=3
Disk[lun4]=4
Disk[lun5]=5
Disk[lun6]=6
Disk[lun7]=7

2 LUN(s)      PhysicalName      SystemName      Drive Mapping
LUN 1         \\.\SymDrive1     Drive1          [Disk1->E:\]
LUN 5         \\.\SymDrive5     Drive5          [Disk5->I:\]
Series3 Controller 1T82119046, SystemName=Drive8 - (port=2,path=4,SCSI Id=28)

Disk[lun0]=8
Disk[lun1]=9
Disk[lun2]=a
Disk[lun3]=b
Disk[lun4]=c
Disk[lun5]=d
Disk[lun6]=e
Disk[lun7]=f

4 LUN(s)      PhysicalName      SystemName      Drive Mapping
LUN 0         \\.\SymDrive8     Drive8          [Disk8]
LUN 2         \\.\SymDrive10    Drive10         [Disk10->F:\]
LUN 3         \\.\SymDrive11    Drive11         [Disk11->G:\]
LUN 4         \\.\SymDrive12    Drive12         [Disk12->H:\]

```

Figure 43. *Lad -a* Output Under Windows NT

7.1.3 Networking

This section covers the tasks involved with setting up the disk subsystem for remote network management. Setting up and installing the Fibre Channel RAID Specialist and configuring the controllers for network attachment will be covered.

7.1.3.1 Summary of Preparation Tasks

Before you install the storage management software, you must perform the tasks described in Table 6 to ensure that the components on the network are set up and operating properly, and that you have all the controller information needed for the software to operate correctly.

Perform the tasks in Table 6 on page 85, and use Table 7 on page 85 to record the information such as RAID Module and controller names, hardware ethernet and IP addresses, and host machine configuration type.

Table 6. Summary of Tasks for Configuring Controllers on the Network

Task to Perform	Why Perform This Task?	For More Details See:
1. Install all hardware components, RAID Modules, cables etc.) that you want connected to the network. Note: You need the proper type of transceiver when connecting your RAID Modules to the network.	To ensure that the network hardware is preset.	The <i>Installation Guide</i> specific to your hardware components.
2. Determine the hardware ethernet address for each controller in all the RAID Modules connected to the network.	To configure the BOOTP server on the network.	Section "Determining the Hardware Ethernet Address" on page 86.
3. Establish and record a naming scheme for each RAID Module and for each controller in the RAID Modules connected to the network.	To help when assigning names and addresses to your controllers and host machines.	Section "Determining Names for RAID Modules and Controllers" on page 86
4. Obtain IP addresses from your network administrator for each controller in all the RAID Modules connected to the network.	To configure the BOOTP server on the network and to set up the host (or DNS) table.	Section "Obtaining IP Addresses" on page 87
5. Configure the hardware and IP addresses on the BOOTP server on the network.	To set up the network address information for the BOOTP server.	Section "Setting up BOOTP and DHCP Servers" on page 87
6. Verify that the TCP/IP Protocol is installed and set up the host (or DNS) table.	To verify your Windows machine is configured to reach the controllers over the network.	
7. Power up the components connected to the network.	To ensure all devices and links are operational.	The <i>Installation Guide</i> specific to your hardware components.

Samples for both single-controller and dual-controller RAID modules are listed on Table 7 on page 85.

Table 7. Sample Table for Recording Module and Controller Information

RAID Module Name	Hardware Ethernet Address	IP Address	Controller Name	Host Machine Configuration Type
Example: (The first RAID Module on the host machine mason.) RAID Module Name: musso_hi_001	00a0b802042c	8.54.19.15	mas_7	Single-host
Example: (The first RAID Module on the host machine auburn.) RAID Module Name: baskin_hi_001	00a0b808065a	8.54.19.28	aub_1	Single-host
	00a0b807f92b	8.54.19.29	aub_2	

Table 8 on page 86 provides a place for you to record module and controller information. Keeping this record will help you properly set up the BOOTP table for the network server and host or DNS table to use the storage management software. It will also help you correctly add additional RAID Modules after initial installation.

Table 8. Table for Recording Module and Controller Information

RAID Module Name	Hardware Ethernet Address	IP Address	Controller Name	Host Machine Configuration Type

7.1.3.2 Determining the Hardware Ethernet Address

Use the following procedure to record the correct hardware ethernet address for each controller in all RAID Modules connected to the network.

1. Turn the power off to all the RAID Modules on the network.

Attention: Consult the manual shipped with your RAID Module for specific procedures on removing/replacing controllers.

2. Remove the controller (one at a time for redundant controllers) from the RAID Module.
3. Record the ethernet address as it appears on the label on the controller back on top of the RAID Module/controller connector. Use Table 8 on page 86 to record the IDs.
4. Replace the controller in the RAID Module.
5. Repeat step two through four for *each* controller in *every* RAID Module.
6. Continue with Section “Determining Names for RAID Modules and Controllers” on page 86.

7.1.3.3 Determining Names for RAID Modules and Controllers

When you open any storage management application for the first time after installation, you will be required to add information about the controllers in each RAID Module to allow the software to recognize them. You will also be able to provide a unique name for each RAID Module and controller. Use the information below, and use Table 8 on page 86 to record the information you will need to add the RAID Modules.

RAID Module Names

The RAID Module name can be anything you want; however, the software allows only 16 characters. You can provide a unique name for each module (see

example names in Table 7 on page 85), or you can accept the default name which will be Net Module xxx (where xxx is a sequential number). If you choose to rename the modules, the new name might include the name of the host machine, an indication of its location, and a unique number.

Controller Names

The controller names can also be anything you want; however, the name should be meaningful to you in the first eight characters, because the software only displays the first eight characters in its device-name display.

Furthermore, you need to provide a *unique* name for each controller in all RAID Modules (see Example Names in Table 7 on page 85). Therefore, we suggest that the name identifies the name of the host machine (connected to your RAID Modules by I/O data connection), and a number representing the controller or its host-side ID. However, keep in mind the eight-character display limit. Because names for host machines could be much longer than the character limit allows, you may need to be creative when creating or abbreviating them.

7.1.3.4 Obtaining IP Addresses

Each controller in every RAID Module on the network needs a unique IP address. Contact your network administrator to get the IP addresses you need and record them in Table 8).

7.1.3.5 Setting up BOOTP and DHCP Servers

This section covers the required and optional settings for configuring a BOOTP (UNIX) or DHCP (Windows NT) server. Review Table 9 on page 87 and then consult your operating system documentation for setting up the appropriate server.

Table 9. Required And Optional Entries for Setting up the BOOTP or DHCP Server

Entry	Description	Format (Example) In BOOTP Server
Required Entries		
subnet mask	Mask used to route packets to defined subnets.	dot notation (sm = 255.255.248.0)
router	IP address of machine that routes packets to other locations or networks.	dot notation (gw = 8.54.19.2)
controller name	Controller name you created for the controller (refer to Figure 8)	controller name mas_7
IP address	IP address of controller (refer to Figure 8)	dot notation (ip = 8.54.19.15)
Ethernet Address	Hardware ethernet address of controller (refer to Figure 8)	hexadecimal notation (ha = 00a0b802042c)

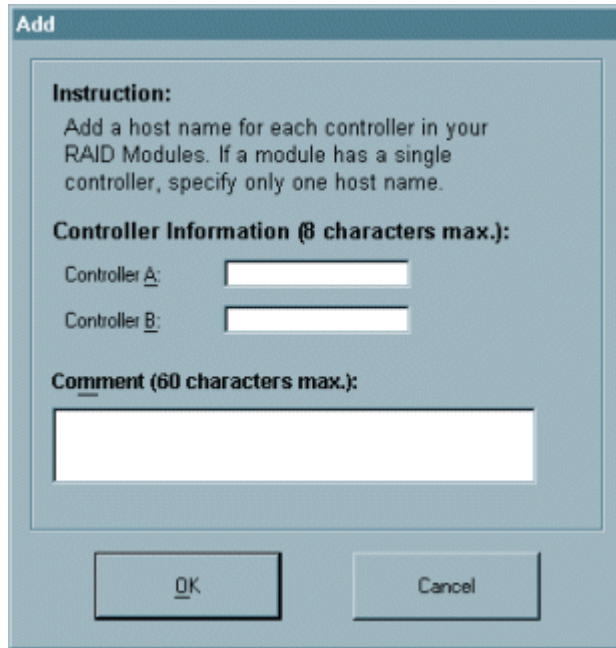
Entry	Description	Format (Example) In BOOTP Server
tag 164	<p>The remote management station (RMS) tag; includes a tag and IP address for each Windows station where you are installing the storage management software.</p> <p>Attention: You can have a maximum of <i>four</i> remote management and SNMP network management stations (NMS); that is the combined total of tags 164 (164 =) and 165 (165 =).</p> <p>The IP address must be converted to hexadecimal notations. Example conversion: 8.54.19.7 becomes 0x08361307</p>	hexadecimal notation (T164= 0x08361307)
Optional Entries		
vm=rfc 1084	Some BOOTP servers require this vendor specific tag in order to return vendor-specific configuration information for controller operations.	vm= rft 1084
tag 163	The network time-out tag sets a value on the controller for the amount of time (in seconds) it will wait for a reply from this software. The default is 30.	T163=30
tag 165	<p>The SNMP tag; includes a tag and the IP address of the NMS. Set this tag if you want the controller to send traps directly to the designated NMS. You can also enable the storage management software to send traps to the designated NMS. See Chapter 7.1.3.7, "Simple Network Management Protocol" on page 89.)</p> <p>Enter the hexadecimal notation of the IP address into the BOOTP table.</p>	hexadecimal notation (T165=0x994F8004)

7.1.3.6 Verifying LAN Connectivity

After you have configured the BOOTP or DHCP server, you will need to perform the following steps to verify that the FC-RSS is on the network:

1. Install the storage management software onto your Windows 95 or Windows NT machine. For more information, see the *Storwatch Fibre Channel RAID Specialist Windows NT and Windows 95 Installation Guide*.
2. Verify that the Disk Array Monitor Service is running.
 - Windows NT: Select **Start -> Settings -> Control Panel -> Services**. The Disk Array Monitor is running if its status is *Started*.
 - Windows95: An icon is the service tray in the lower right corner of your screen. Move your cursor over the icon. A status message pops up indicating that the Disk Array Monitor is running.

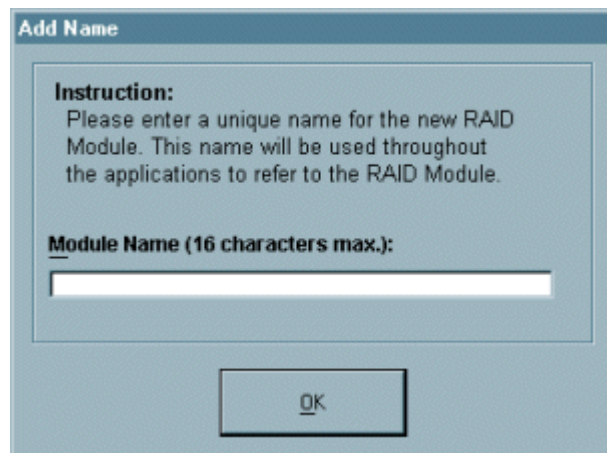
3. Run the Configuration Application. Select **OK** to add RAID modules, or select **Module** then **Select...** This will result in a display like that shown in Figure 44.



The 'Add' dialog box has a title bar with the text 'Add'. Inside, there is an 'Instruction:' section with the text: 'Add a host name for each controller in your RAID Modules. If a module has a single controller, specify only one host name.' Below this is a section titled 'Controller Information (8 characters max.):' containing two text input fields labeled 'Controller A:' and 'Controller B:'. Underneath is a section titled 'Comment (60 characters max.):' with a larger text input field. At the bottom of the dialog are two buttons: 'OK' and 'Cancel'.

Figure 44. Add Module Screen

4. Enter the controller names for the first RAID Module you define in Table 8 on page 86. If the controller names are different from the IP hostnames for the controllers, use the IP hostnames. If your machine is on a different IP subnet from the FC-RSS, then you may need to modify your machine's IP settings to search multiple subnets. Select **OK**.
5. The Add Name screen appears, as shown in Figure 45 on page 89.



The 'Add Name' dialog box has a title bar with the text 'Add Name'. Inside, there is an 'Instruction:' section with the text: 'Please enter a unique name for the new RAID Module. This name will be used throughout the applications to refer to the RAID Module.' Below this is a section titled 'Module Name (16 characters max.):' with a single text input field. At the bottom of the dialog is one button: 'OK'.

Figure 45. Add Name for RAID Module

7.1.3.7 Simple Network Management Protocol

This section covers the setting up and configuration of Simple Network Management Protocol (SNMP) of the subsystem.

Enabling SNMP Notification

Note: In order to enable or disable the (SNMP) notification option, you must edit the `\program files\ibm\fcsm\rmparams` file. See Chapter , “Changing the `rmparams` File” on page 91 for instructions on changing the `rmparams` file.

This software can provide remote notification of RAID events to a designated Network Management Station (NMS) using Simple Network Management Protocol (SNMP) traps.

Use the following procedure to set up the NMS, to enable or disable this notification option for each host, and to understand the contents of an SNMP trap message.

1. Set up the NMS. You only need to set up your designated NMS *once*. Use the following procedure to compile this software’s Management Information Base (MIB) into your NMS.

Note: The MIB was written to standard version 1 SNMP syntax. It was designed specifically for the storage management software. It is not applicable to other vendors’ RAID products.

- a. Copy the `\program files\ibm\fcsm\rm6traps.mib` file to the NMS.
- b. Follow the steps required by your specific NMS to compile the MIB.

Note: For details on the required steps, consult your network administrator or the documentation specific to the NMS product you are using.

2. Make sure the host system is connected to the network with the NMS station.
3. Enable this notification on your host. You must edit three parameters in the `\program files\ibm\fcsm\rmparams` file so that the software will know where to direct the SNMP trap messages when a RAID event occurs.

Contents of an SNMP Trap Message

If the trap is anything other than informational, you should use the Status or Recovery applications to get more information or specific instructions on how to remedy the problem. Do *not* remove the SNMP message until the problem has been corrected since most trap notifications associated with this software are not persistent. Table 10 on page 90 explains the syntax of the SNMP messages.

Table 10. Contents of an SNMP Trap Message

Item	Description
Host IP Address	The standard IP address dot notation for the host where the RAID Module is attached.
Host Name	The text name assigned to the host.

Item	Description
Trap Type	<p>There are four trap types:</p> <ul style="list-style-type: none"> • Informational - No failure has occurred. Information is provided that an event, such as a LUN creation, has occurred. • Degraded Array - A subsystem component failure has occurred. The module is still operational, but now lacks redundancy in some area. (Some examples are: a drive failure resulting in a degraded LUN, a power supply failure, or a failure of one controller in a controller pair.) • Failed Array - A failure has occurred that leaves a module or LUN unable to communicate. (Some examples are: a failure of the controller in a single-controller RAID Module, or a multiple drive failure resulting in a dead LUN.) • Critical - This is a small category of failures that may require timely intervention to correct. (Some examples are: the two-minute warning from an uninterruptible power supply, or a warning that write-back cache has been disabled due to a low battery.)
Affected RAID Module	Indicates which RAID module had the event (for example, wareagle_001).
Condition Description	A brief description of the event.

Disabling SNMP

Perform the following steps to disable this notification on your host and stop this software from sending SNMP traps:

1. In the `\program files\ibm\fcsm\rmparms` file, change the value of the `SNMP_ACTION_NUMBER` line to 0.
2. Save the `rmparms` file and copy it to the backup directory. It is critical that you have a backup copy of this file.

Changing the rmparms File

The `\program files\ibm\fcsm\rmparms` file is used to store several configurable options available through this software. Table 11 on page 92 shows a list of the parameters in the file that you may need to change during initial installation.

Note: Before you edit the `rmparms` file, make sure you read Chapter , “Enabling SNMP Notification” on page 90.

If you installed an upgrade of the storage management software, you may want to refer to the `rmparms.ori` file to determine if you want to make additional customized changes to the new file. Remember, the new file contains new parameters for this version, so do *not* replace it with the old file. Most of the options in the `rmparms` file (other than the ones shown in Table 11 on page 92) should *not* be changed except through the graphical user interface (GUI).

For more information about the `rmparms` file, read the text file about `rmparms`. From the installation directory, type `help rmparms`. Because of the critical nature of the `rmparms` file, you should make a backup copy of the file in a directory other than the installation directory before you modify the current file. If this file becomes corrupted or missing, you should copy the backup `rmparms` file

to the installation directory to resume normal operations. Furthermore, if you make any changes to this file directly, or if you use the options in the GUI that write to this file (such as log settings), always copy the new version of the file to the backup directory once you are sure the changes work.

Table 11. Changeable Parameters in *rmparams* File

Description	When to Change	How to Change
SNMP_TARGET_IP_Address=xxx.xxx.xxx.xxx		
Provides the SNMP IP Address for remote notification.	If you want to enable SNMP	Change this value to the dotted notation IP address of the network management system where the MIB has been compiled. On most systems, this IP address can also be a dotted notation multicast address or broadcast address.
SNMP_Community_Name=NONE		
Provides the community name to be used for SNMP remote notification.	If you want to enable SNMP.	Add the community name. The most commonly used community name is public ; however, your network environment may require another name.
SNMP_Action_Number=0		
Indicates the SNMP trap notification option you want to use.	If you want to enable SNMP.	Change this value to either 1 or 2. The value 1 will send SNMP traps to the designated NMS station. The value 2 means that all the relevant data is written to the file <i>trap.dat</i> . Note: Setting this value to 0 will completely disable the SNMP option.
System_DefaultModuleSelect=TRUE		
Determines whether or not the Module Selection screen is displayed each time you open an application.	If you want to disable the Module Selection screen	Change TRUE to FALSE
System_MaxControlers=16		
Determines the maximum number of RAID controllers supported by this software.	If you have more than 16 RAID controllers.	Change the 16 to the appropriate number.

Setting up Notification Options

The `rmscript.bat` file is accessed by the status monitor and automatic parity check whenever these utilities find an exception. By default, no action is taken. You can edit this file to specify a set of actions to take when an error is found.

For example, if you want to install third-party notification software to perform various events (such as paging or emailing), you would add the appropriate script for each event near the end of this file (making sure that `del %1` remains the last line in the file). Consult the third-party's software documentation for specific details.

Note: For more information, read the text file about `rmscript.bat`. From this software installation directory, type:

```
help rmscript
```

7.1.4 Expansion

This section covers the addition of drawers and disks to the subsystem.

7.1.4.1 Adding Disks to Existing Drawers

This procedure covers adding new disks to an existing drawer and should be performed with the FC-RSS powered on. Adding a disk to a drawer is a non-disruptive procedure.

1. Remove the dummy drive module(s) from the drawer.
2. Insert new drive modules into the drawer.

Within about one to two minutes, the new drive will be listed in the Configuration Application under the *Unassigned* group. It will be labeled with the corresponding (channel,id) pair.

You should now proceed with Chapter 7.1.1.2, “LUN Creation” on page 75 to create LUNs.

7.1.4.2 Adding Drawers to the FC-RSS

Refer to Chapter 6.3, “Cabling” on page 52 for cabling up the new drawer. If the new drawer is the fourth or greater drawer in the rack, or a drawer will change from containing two SCSI channels to only one channel, then steps can be taken to avoid a full restoration of data. If the new drawer is the second or third drawer in the rack, then there is no need for recabling; you can attach an available cable as outlined in Chapter 6.3.4, “Cabling the FC Storage Unit” on page 53.

The FC-RSS addresses disks using a (channel,id) format. This is independent of the disk’s physical location in the rack, the placement of the drawer within the rack, or the drawer id, but it is related to the location of the disk in the drawer.

Warning: Information about the individual disk’s (channel,id) is stored on the disk itself. When moving disks from one drawer to another, it is important to move these disks according to the following procedure. Failure to do this properly, could result in the loss of data on those disks. Backups of the data on the disks that are going to be moved should always be taken first.

For example, after powering down the system, the disk at (2,0) is moved from the top drawer to the bottom drawer, which is then recabled as channel 2. When we power the system back up, the RAID module would find the same physical disk at location (2,0), but it would now be in the bottom drawer. Thus, data would be preserved.

In the following sections, an example of how to migrate from a two-channel or split-drawer configuration to a single-channel drawer will be demonstrated.

Preparation

Prior to powering down the disk subsystem, perform the following steps:

1. Determine which disks in the subsystem will be moved to the new drawer. This can be done by identifying which channels in the drawer will be moved. In general, if a drawer is split between two channels, the second channel should be moved to the new drawer. This will limit the maximum number of moved disks to five.
2. **Backup** all data on any LUNs which could be affected by moving drives.

- Using either tape or small stickers, mark those drives that will be affected with the corresponding (channel,id). Use the List/Locate Drives function to identify the (channel,id) pair for each drive. In Figure 46, the drives for both channel one and channel two have been marked. In this example, we will be migrating the drives from channel two, to a single drawer.

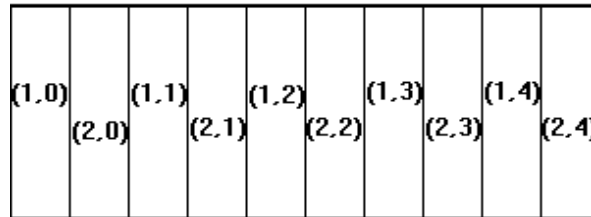


Figure 46. Two-Channel Drawer

- After labeling the drives to be moved with (channel,id). Power off the RAID controllers and disk drawers.

Recabling

- Move the SCSI cable representing the second channel (in this example, channel two) on the original drawer, to Bus 1 Connector on the back of the new drawer as shown in Figure 47.

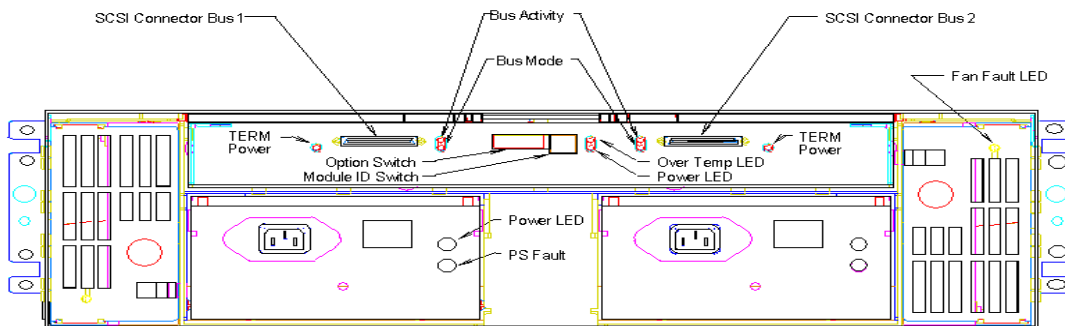


Figure 47. Drawer (Rear View)

- On the original drawer, set Option switch 1 to the off position. At this point all of the Option dip switches should be in the **off** position. Only those drawers which are split between two channels should have Option switch one in the **on** position.
- On the new drawer, verify that the Option switch is in the **off** position.

Migrating Disks

In this section we will physically move the drives to the new drawer, while preserving the (channel,id) pair.

- Figure 46 is an illustration of how the disks are laid out when there are two channels, while Figure 48 illustrates the placement of the moved disks.

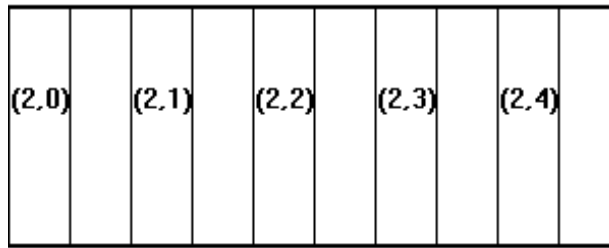


Figure 48. Single-Channel Drawer

If you compare Figure 46 with Figure 48, you will notice that the disks have shifted one slot to the *left*.

2. Move the drives from the original drawer to the new drawer while preserving the (channel,id) pair.
3. Power the drawers and then the controllers back on.
4. Verify that your host can still access the LUNs and data.
5. The remaining slots in both the original and new drawer can be filled with additional disks. These new drives will become ids 5 through 9. You can now proceed with creating additional LUNs or hotspares.

7.2 AIX

The following sections cover the post installation configuration process under AIX.

Note: The AIX software uses different terminology than the software used to configure the FC-RSS under Solaris, HP-UX and Windows NT as displayed in Table 12 on page 95.

For a map of the SMIT menus, see Appendix A, “SMIT Menus” on page 119.

AIX	HP-UX, Solaris, Windows NT
FC RAIDiant Disk Array Manager	IBM FC Storage Manager
Array	Drive Group
Spare	Unassigned Drive Group
Dac (Disk Array Controller)	FC-RSS Controller
Dar (Disk Array Router)	not-applicable
Fast Load Caching	Write Cache mirroring

Table 12. AIX and HP-UX/Solaris/Windows NT Terminology

7.2.1 Preparation

Prior to allocating any storage, you should verify that the DARs and DACs are recognized by AIX. To verify, perform the following procedure:

1. type: `lsdev -C |grep 2102`

This example results in one DAC, one DAR and the default 10 MB LUN:

```
dac0      Available 04-03-01      2102-F10 Disk Array Controller
dar0      Available                2102-F10 Disk Array Router
hdisk10   Available 04-03-01      2102-F10 Disk Array Device
```

2. Place the DAC in dual active mode type: `chdev -l dar#` (where # is the number of the DAR).

7.2.2 Storage Allocation

This section covers array and hot-spare creation.

7.2.2.1 Creating an Array

Creating arrays and thus allocating storage follows the same basic theories as the other platforms; however, the interface for performing these functions is done through `smitty`.

To create a new array, perform the following procedure:

1. Type **smitty fraidmgr**.
2. Select **RAIDiant Disk Array Manager**.
3. Select **Create a SCSI RAID Array**.
4. If you plan on creating a LUN on a new array select **Create a SCSI RAID Array**.
5. Select the DAC that the array will be attached to.
6. Select the RAID level. **Note:** AIX only supports RAID levels 0, 1 and 5.
7. Use the **F7** key to choose the disks that will become members of the array, and then press **enter**. The candidate disks should be chosen so that the best performance can be achieved throughout the array. An example would be to choose disks that are attached on different SCSI busses for the array.
8. The **Size of Array** field determines the size of the *hdisk* to be created. Use the F1 key, to receive help on any field. When you are done modifying any of the fields, press **enter** to create the *hdisk*. For performance, it is recommended to choose the maximum size available.
9. After the *hdisk* has finished formatting, follow your AIX documentation for using *hdisks* to create volume groups using LVM. **Note:** When using the `smitty` interface to set the characteristics of the logical volume, the **Enable BAD BLOCK relocation** must be set to **no**.

If you are creating a LUN from an existing array, perform the following:

1. Type **smitty fraidmgr**.
2. Select **RAIDiant Disk Array Manager**.
3. Select **Create a SCSI Sub RAID Array** (if the maximum size was not chosen in step 8 above).
4. Select the *hdisk* that you want to create the sub array on.
5. The **Size of Array** field determines the size of the *hdisk* to be created. Use the F1 key to receive help on any field. When you are done modifying any of the fields, press **enter** to create the *hdisk*.

6. After the *hdisk* has finished formatting, follow your AIX documentation for using *hdisks* to create volume groups, or adding the *hdisk* to an existing volume group, using LVM. **Note:** When using the smitty interface to set the characteristics of the logical volume, the **Enable BAD BLOCK relocation** must be set to **no**.

7.2.2.2 Creating a Hot Spare

Hot spares add another layer of protection to the data on your RAID modules. Hot spares automatically replace failed drives in your LUNs. You can only create hot spares if you have unused drives.

Attention: A hot spare can only replace a drive of equal or lesser capacity. If your subsystem contains drives of multiple sizes, always choose a drive of the larger capacity to become a hot spare.

To create a hot spare, perform the following procedure:

1. Type **smitty fibmraid**.
2. Select **Change/Show Drive Status**.
3. Select the DAC that this hot spare will be created by and press **enter**.
4. Choose a drive of type *Spare Drive* and press **enter**.
5. Select **Drive Status** and press **F4**, then select **Mark Drive Hot Spare** and press **enter**.
6. Confirm the creation by pressing **enter**.

7.2.3 Modifying Array Attributes

1. Type **smitty fibmraid**
2. Select **Change/Show SCSI RAID Array**
3. Select the *hdisk* you want to modify.
4. Modify the values to be changed and press **enter** when you are done.

7.2.4 Expansion

This section covers the addition of drawers and disks to the subsystem.

Follow the instructions contained within Part 7.1.4.2, "Adding Drawers to the FC-RSS" on page 93, however, to identify a drive perform the following procedure:

1. Type **smitty fibmraid**
2. Select **Change/Show Drive Status**
3. Select the *hdisk* or DAC that contains the disks that you want to identify
4. Select the disk to be identified
5. Select Drive Status and press **F4**. Then select **Turn on drive visual identifier**. You may want to put a sticky label on the drive.

Note: AIX will flash the amber check light, not the activity light.

6. To turn off the identification, perform the same procedure, but change the Drive Status to **Turn off drive visual identifier**.

Chapter 8. Management

In this chapter we will overview disk subsystem management topics, and expand on topics such as migration from other disk subsystems and common post install issues.

8.1 Overview and Philosophy

The management software that comes with the FC-RSS provide many powerful and versatile management functions. For Windows NT, Sun Solaris and HP-UX attached systems, the IBM Fibre Channel Storage Manager and the StorWatch Fibre Channel RAID Specialist provide functions to configure the FC-RSS subsystem, perform problem determination and recovery actions, and dynamically modify disk groups and LUNs through either a graphical or command-line interface. The graphical interface can be invoked either from the host server or from the Storwatch FC Specialist running on a Windows workstation.

Both the IBM Fibre Channel Storage Manager and the StorWatch Fibre Channel RAID Specialist provide the following functions:

- Redundant dual controller support with automatic failover
- RAID storage configuration and management through an easy-to-understand GUI interface
- Recovery Guru, an online technical consultant that will help you diagnose and recover from component failures
- Built in "health check" that monitors the condition of RAID modules and provides detailed reports as needed
- Dynamic expansion capabilities that allow you to change the RAID level, add more capacity to drive groups, restore data storage space, and change LUNs' segment size concurrent with I/O activity
- Performance Monitor, which allows you to monitor I/O activity on selected controllers and LUNs for a selected RAID Module check
- Configurable automatic parity check
- Auto Code Synchronization that automatically updates downlevel firmware when a controller is replaced
- Error notification via SNMP or configurable scripts

The only departure from this structure is the Fibre Channel Storage Manager for the IBM AIX operating system. Under this operating system, configuration and management tasks are performed via the familiar SMIT interface. On the RS/6000 AIX platform, the storage subsystem management software is very similar to the 7135 RAIDiant Array software management functions. The FC SCSI RAIDiant Disk Array Manager provides the following capabilities within the SMIT menus:

- Create, configure, reconfigure or delete RAID arrays, and subarrays
- Change RAID levels, drive selection or array size, write caching, read prefetch, and command queue depth, and modify array ownership

- Display and monitor DAR (disk array router) configuration attributes, array status or drive status
- Perform dynamic load balancing
- Modify health check frequency
- Restore or repair LUNs after parity errors or drive failure

8.2 StorWatch Fibre Channel RAID Specialist

The StorWatch Fibre Channel RAID Specialist allows you to manage one or more FC-RSS subsystems from a Windows workstation. This provides the flexibility of centrally managed storage administration for multiple system platforms and the consistency of a single GUI interface.

This interface is not supported for IBM RS/6000 attachments.

8.3 Command Line

All the management functions provided in the IBM Fibre Channel Storage Manager and StorWatch Fibre Channel RAID Specialist GUI interface can also be performed through the command line interface. Again, the sole exception to this is the IBM AIX Fibre Channel Storage Manager, which, as we have noted before, has its own SMIT interface.

The command line interface provides additional tuning capability for cache management through the *raidutil* command. In addition, on Windows NT server implementations, the *lad* command provides additional disk mapping information.

8.4 General Functions

The IBM Fibre Channel Storage Manager and StorWatch Fibre Channel RAID Specialist provide functions to configure, tune, check subsystem status, and perform firmware maintenance on the IBM Fibre Channel RAID Storage Server. Figure 49 below are the icons that represent the functional categories.

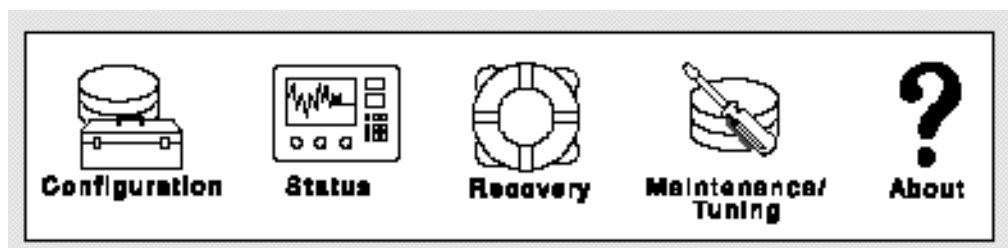


Figure 49. GUI Function Selection Icons

In the AIX environment, all storage management functions are performed through the SMIT (Systems Management Interface Tool) interface.

8.4.1 Initial Setup

8.4.1.1 Serial Port Setup

The FC-RSS provides a serial interface for IBM service personnel to use in performing diagnostics or emergency configuration procedures. Normally this interface is not even required or used by service personnel. In some situations, under the direction and guidance of the product support organization for the FC-RSS, there may be a need for a low level interface to the controller. This is the only time that this interface should be utilized. Use of this interface requires a special password and knowledge of internal command language.

8.4.1.2 Configuration

On all server platforms except AIX, once you select the configuration function through either the Fibre Channel Storage Manager or through the FC Specialist, the GUI presents the selections as illustrated in Figure 40 on page 78. Refer to Chapter 7, "Configuration" on page 75 for details on configuration using either IBM Fibre Channel Storage Manager or StorWatch Fibre Channel RAID Specialist and to the specific platform *Fibre Channel Storage Manager and StorWatch Fibre Channel RAID Specialist User's Guide(SC26-7285)* and the *Fibre Channel Installation Guides*. For each of the supported platforms, there is a operating system unique guide, for Windows NT it is *SC26-7283*, for Sun Solaris it is *SC26-7286*, for Hewlett Packard HP-UX it is *SC26-7287*, and for IBM AIX it is *SC26-7290*.

Figure 50 shows the icon for selecting the configuration function.

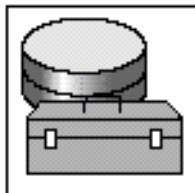


Figure 50. Configuration Icon

For AIX, you will use the standard AIX SMIT interface for configuration. See *Fibre Channel Storage Manager for AIX SC26-7290* for interface description and commands. You configure the FC-RSS controller via the DAC (Disk Array Controller) SMIT option and you configure RAID arrays via the FCRDAM(FC RAIDiant Disk Array Manager) SMIT option.

If you are also installing the 2103-H07 IBM Fibre Channel Storage Hub, you should also consult the *Fibre Channel Storage Hub Service and User's Guide (SC26-7289)* for configuration and cabling options.

8.4.2 Operations

In the following sections, we list the functions that you would use, as systems administrator, to configure, manage and monitor your FC-RSS in a production environment. Each section describes a set of functions you will likely use in the day-to-day operation of the FC-RSS. These include monitoring, modification, and maintenance tasks.

8.4.3 Monitoring

Through either the FC Storage Manager or the StorWatch FC Specialist, the functions that are used for monitoring system status include:

- Reviewing message/error logs
- Monitoring performance
- Checking array rebuild status and rebuild rate

As with the configuration function, the status monitoring function is accessed through the GUI interface by clicking on (selecting) the following icon as shown in Figure 51.

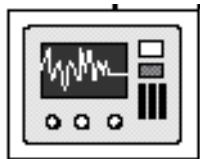


Figure 51. Status Monitoring Icon

In an RS/6000 environment, you monitor DAR (Disk Array Router) configuration and healthcheck notification frequency, AEN (automatic error notification), and load balancing via the DAR configuration SMIT option, and you list RAID arrays or view disk drive status via the FCRDAM SMIT option. For viewing disk array ownership or controller status, you use the DAC configuration SMIT option.

8.4.4 Modifications

These are the functions that are used for tuning your IBM Fibre Channel RAID Storage Server and for maintaining the firmware and microcode on the subsystem.

8.4.4.1 Tuning

In this section, we describe the functions that you use to tune your FC RAID subsystem. They include functions such as:

- RAID selection, disk group mapping, LUN assignment and reassignment
- Defining LUN balancing
- Setting cache parameters
- Setting array reconstruction ratio

Selection of the icon as shown in Figure 52 will take you to the tuning function.

For RS/6000 environments, disk subsystem tuning is done via SMIT options.

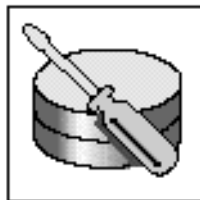


Figure 52. Maintenance and Tuning Icon

8.4.4.2 Changing RAID Groups

Once RAID groups are defined, the IBM Fibre Channel RAID Storage Server software gives you the capability to redefine and reassign RAID groups. You can change from one type of RAID to another. Refer to Chapter 7, “Configuration” on page 75 and to Chapter 9, “Performance Tuning” on page 105 for more complete information on those functions.

8.4.5 Maintenance

These are the tasks that are required to maintain and upgrade the firmware and software on your IBM Fibre Channel RAID Storage Server controllers and the microcode on your disk drives. Selection of the icon as shown in Figure 53 will take you to the maintenance function.



Figure 53. Maintenance and Tuning Icon

8.4.5.1 Fixes

Controller firmware patches and fixes can be applied through the GUI Maintenance and Tuning function by selecting the firmware upgrade option from the menu bar.

8.4.5.2 Upgrade

Applying new levels of controller firmware or disk microcode can also be done from the firmware upgrade option under the maintenance and tuning function.

8.4.5.3 RS/6000 and AIX Maintenance

Maintenance and microcode update for controller or disks is performed through standard SMIT menus (through the SMIT fastpath `smit update_all`).

8.4.6 Migration Considerations

In general, migration from internal disk or any disk subsystem requires planning. You should validate that the FC-RSS will meet the customer’s capacity, availability and performance needs. You will need to put in place a plan for moving existing data or files to the new subsystem. You will need to determine how much time and processing to allow for the migration. There should be a plan to test the migration process and procedures prior to the installation of the new subsystem in production. Another thing that needs to be checked is that the server system has the adapter capacity to enable migration. Adapter capacity and placement may dictate the overall migration plan.

Other considerations that you should review are application, DBMS, or storage management software (for example, Veritas) dependencies or idiosyncrasies.

8.4.6.1 From SCSI RAID5

When migrating from RAID5 SCSI subsystems like the 7135 or any other external RAID5 disk subsystem you need to consider the following:

- Capacity required on the FC-RSS
- Amount of read write cache needed
- LUN, file system(s), and data organization

8.4.6.2 From SSA

When migrating from 3527, 7131, or 7133 SSA subsystems, JBOD, or RAID1, 3, or 5, you should consider the following:

- Capacity required on the FC-RSS
- Amount of read/write cache
- Shared access, homogeneous and heterogeneous(via 7190 or VICOM)
- SSA disks attached via the native SBUS adapter or via 7190/VICOM

8.4.6.3 From SCSI JBOD or RAID1

Migration from internal or external SCSI disks should be straightforward. Physical disks can be mapped one-for-one to LUNs defined in the FC-RSS. Mirrored disks or RAID1 arrays can also be mapped one-for-one up to the largest physical disk supported by the IBM Fibre Channel RAID Storage Server. If larger disks are needed, then a RAID 0+1 array can be defined to provide a LUN of the required capacity.

8.4.7 Clustering Considerations

There are no unique considerations for installing the FC-RSS in a clustering environment beyond what is documented previously in this redbook or in the product manuals.

8.4.7.1 MSCS

See Microsoft Cluster Support (Wolfpack) considerations in Chapter 6.4.1.2, "Installing the Software on a Cluster Server Environment" on page 62, and the *Fibre Channel Storage Manager Microsoft Windows NT Installation Guide* SC26-7283, for additional information.

Chapter 9. Performance Tuning

This chapter will focus on the tuning of the FC-RSS. The various options and methods of tuning the three areas of the system are covered, including: Fibre Channel, controller cache, disk groups and LUNs.

9.1 Fibre Channel

In configuring the cabling layout for your subsystem to host connections, you must be aware of possible cable failures and the performance implications of a single loop or connection. FCs inherent nonredundancy requires you to take precautions when implementing your cabling. However, when you configure your installation for maximum performance you also gain the benefit of maximum performance, an attribute not often found with other storage front-ends.

9.1.1 Basic Configurations

Basic configurations consist of a single host attached to a single FC-RSS.

FC frames can only travel in one direction and in one path from the host to the subsystem. Therefore, we recommend that you implement redundant paths from the host to the FC-RSS controllers. This would require a second set of cables, hubs and host adapters, because if any one component fails in the path from host to FC-RSS controller, the *entire* path has failed.

For example, your configuration might consist of a host connected to a hub and the hub connected to both controllers. Then you would only be protected against two types of failures: FC-RSS controller failure and *one* hub-to-controller cable failure. If either the host-based FC adapter, the hub, or the host-to-hub cable fails, you will have lost access to the *entire* FC-RSS.

The solution? Implement a point-to-point topology between the host and the FC-RSS. In this case there is no need for a hub, but instead, a second host-based FC adapter is required. Not only do you have the full redundancy of two independent paths, but now you have the doubled the bandwidth to communicate with the FC-RSS.

9.1.2 Advanced Configurations

If your installation includes multiple hosts or multiple subsystems you will have to plan your cabling structure carefully in order to preserve redundant paths to the subsystems.

When an advanced configuration is used, hubs must be introduced to allow for connecting of as many as four devices together. The steps for cabling up the fibre should be as follows:

1. Label the hubs A and B. The first hub will be cabled up to controller A in the FC-RSS, or if two FC-RSSs are used, the hub will be cabled up to both controller As. The second hub will be cabled up to controller B, or if two FC-RSSs are used, both controller Bs.

The cabling will allow you to survive hub failure, hub to controller cable failure, or controller failure, because we can fail the LUNs over to the other controller and use the second fully redundant path to connect to the LUNs. If you have

two or more FC-RSSs, then continue to use two hubs, the first for the controller As and the second for the controller Bs.

2. Install two FC adapters per host, and connect each adapter in the host to a different hub. You can simplify the configuration by always labeling the first FC adapter "channel A" and the second FC adapter "channel B", and then connecting them to their respective hubs.

This setup will allow for the failure of either a host adapter or the host-to-hub cable. Also, by using two host-based FC adapters, the hosts will be able to communicate at speeds up to 200MB/s.

Figure 54 shows an example of a configuration with two hosts, each with two FC adapters, connected through two hubs to two FC-RSS, providing full availability.

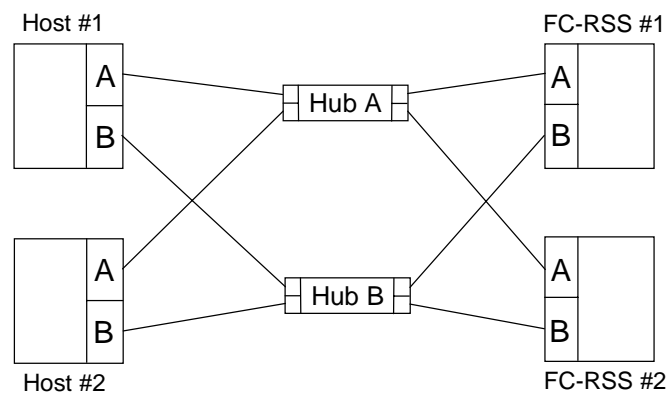


Figure 54. Example of an Advanced Configuration

9.2 Cache

When tuning cache parameters, you must thoroughly understand the types and size of data that you are working with, because incorrect tuning of the FC-RSS can lead to decreased performance. This includes determining whether you perform mostly sequential or random accesses. These features can *only* be tuned using the `raidutil` command or `smitty`. For usage information on this or any other command line utility type: `help <command>` for windows or `man <command>` for unix.

9.2.1 Read Ahead Prefetching

Read ahead prefetching is best suited for large sequential reads. When a read operation is issued by the host, the controller performs additional reads to be stored into the cache. If sequential reads are being issued, the performance of the FC-RSS can be increased, because the amount of cache hits increases. However, if random reads are used, then using this option can actually hurt performance, because the controller spends time performing reads that are not needed. This can be enabled and configured on a per-LUN basis.

Performance can be greatly affected by enabling the controller's prefetching algorithms under certain circumstances. Prefetching is useful if the I/O profile contains:

- Dominant reads

The majority of the I/Os issued by the host system must be read commands.

- Sequential reads

The read commands issued by the host system must be sequential in nature, in *ascending* order. This means that subsequent read commands must access higher logical block addresses on a logical unit. Note that the accesses need not be consecutive.

- Small number of processes

There should be only a small number of processes. Too many I/O processes, even if sequential in nature, may appear to the controller as random I/O.

One rule of thumb used to determine whether there are "too many" processes is to attempt to determine the amount of cache that can be used for read requests (i.e. the total cache on the RAID controller minus that used for write operations and that used for mirroring the alternate controller's cache). If there is adequate read cache available to fulfill all of the data plus the prefetch requirements, prefetching will be beneficial. If the various processes begin to compete for the same cache on the controller, prefetching will be detrimental to performance.

- Large I/O requests

If the host system requests a large number of blocks to be transferred in a read request, the user will typically see greater benefit from prefetching, than if the I/O requests are for smaller number of blocks.

Note that if prefetching is enabled, and the I/O profile is not like that listed above, prefetching can actually be detrimental to I/O performance.

You can specify the number of blocks to prefetch in one of two ways:

- The controller can be directed to always prefetch a specific number of blocks.
- The controller can be directed to prefetch some scalar number multiplied by the number of blocks requested by the host in its last read request.

Thus, the number of blocks to prefetch will be variable, based on the size of the host I/O request.

For *AIX*, follow the procedure outlined in Part 7.2.3, "Modifying Array Attributes" on page 97, and select **Size of Read Prefetch (blocks)**.

For *HP-UX*, *Solaris*, and *Windows NT*, use the `raidutil` command, to set a minimum and maximum prefetch amount depending on the number of blocks requested. An example of the command follows:

```
raidutil -c <device name> -f "1 4,5" -k "2 4,5" -b "256 4,5" -t "1024 4,5"
```

- `-c device name` would have a device name in the form `cXtXdXsX` (UNIX), `driveX` (NT), or host name of the controller, for example, `mas_7` (networked version).
- `-f "1 4,5"` would enable prefetching on LUNs 4 and 5.
- `-k "2 4,5"` indicates that the prefetch amount is 2 times the read request length on LUNs 4 and 5.
- `-b "256 4,5"` results in up to 256 blocks being prefetched on LUNs 4 and 5.

- `-t "1024 4,5"` says that any request over 1024 blocks will not be prefetched on LUNs 4 and 5.

Note: The `-b`, `-k` and `-t` must all be *non-zero* for pre-fetch caching to be enabled.

9.2.2 Cache Flushing

Note: These parameters are not configurable under AIX.

The cache flushing attributes are used to specify the threshold at which point write cache is flushed to disk and at what point it stops flushing the cache. There are two command line options to the `raidutil` command that are used:

- Demand flush threshold
- Demand flush amount

9.2.2.1 Demand Flush Threshold

Usage: `raidutil -c <device name> -T #`

This parameter is set by using `#` as a decimal value (0-100). Use this value to specify the percentage of cache that must be filled before the controller begins to flush the unwritten cached data to the media. For example, a value of 80 would indicate that 80 percent of the allowable cache must be filled with unwritten data before flushing occurs.

If this value is set high, a large part of the cache will be consumed by write requests before the controller starts flushing write data to disk. This increases the chances for more cache hits to occur on the dirty data, thereby lessening the number of drive I/Os the controller will have to issue. However, the cache available to satisfy host read requests will be relatively small.

If this value is set low, only a small part of the cache will be consumed by write requests before the controller starts flushing write data to disk. Correspondingly, a relatively large part of the cache will be available to satisfy host read requests, thereby increasing the chance of more cache hits on read data.

9.2.2.2 Demand Flush Amount

Usage: `raidutil -c <device name> -A #`

This parameter is set by using `#` as a decimal value (0-100). Use this value to specify the percentage of the demand flush threshold that will remain unwritten after the demand flush has been completed. For example, if the demand flush threshold was set at 80 percent, a value of 20 for the demand flush amount would indicate that 16 percent of the cache will remain unwritten to disk.

The default setting for these parameters is 80,20 which means that when the cache is 80% full, flush the cache to disk until the cache is 16% full.

If the Demand Flush Amount is set very high, the controller will flush very little write data each time it flushes cache. This means that most write data will spend a relatively long time in cache, thus making it possible for more cache hits on the write data. However, the controller will be flushing data more times, thus creating greater disk activity, but of short duration.

If the value is set very low, the controller will flush large amounts of data each time it flushes cache. The controller will flush data relatively fewer times, thus creating few disk I/Os, but of longer duration.

We have noticed performance improvements when the parameters are set to 50,100 which means that when the cache is 50% full, flush the cache until it is 50% available. When the overall cache is more than 50% full, it will flush the cache until it is 50% full. Thus, the cache will always be 50% full. We are trying to keep as much data in cache as possible. In comparison, in the default (80,20) setting, there is a forced flushing down to 16% when the cache reaches its 80% capacity.

For example, to set the cache flushing parameters to 50, 100, type:

```
raidutil -c <device name> -T 50 -A 100
```

9.2.3 Cache Block Size

Note: This value is not configurable under AIX.

The RAID controllers manage the cache in blocks. A block consists of 512 bytes of data. By adjusting the cache block size, you are able to tune the cache to your specific data transfer size.

Cache block size is the number of blocks transferred into cache by the controller at one time.

A larger cache block size increases the probability of data near the accessed block also being available in cache. However, a larger cache block size also fills up the cache sooner.

A small cache block size should be used for systems that require transaction processing requests, or I/O streams that are typically small and random. While larger cache blocks should be used for large I/O, sequential, high bandwidth applications.

Using the `raidutil` command you are able to modify how large a cache block is. A cache block can be either 8 or 32 blocks in size, which translates into 4KB and 16KB cache blocks. For example, to set the cache block size to 32blocks, perform the following command:

```
raidutil -c <device name> -S 32
```

9.3 Disk Groups and LUNs

This section will focus on tuning and optimizing the back-end of the FC-RSS, which consists of the disk drive portion of the subsystem.

9.3.1 Choosing Disks for a Drive Group

Within the SCSI subsystem there are two independent PCI busses. Contained within these busses are the six SCSI channels that in turn connect to the drives. PCI bus 1 contains SCSI channels 1, 2 and 3, while PCI bus 2 contains SCSI channels 4, 5 and 6.

Note: Although you are not able to alternate between two controllers when accessing a LUN, you *are* able to alternate between both PCI busses.

For example, when creating a LUN, and you are selecting drives, the order in which you pick the drives becomes the order in which they are striped. Thus, you should pick a drive from PCI bus 1, followed by a drive from PCI bus 2, followed by a drive from PCI bus 1, and so on. Therefore, to create a three-drive LUN, for example, pick (1,0), (4,0) and (2,0).

9.3.2 Balancing LUNs

If your system is not configured for independent controllers, then you have the ability to balance the LUNs across the two controllers. In reality, you are not balancing LUNs but are instead balancing *Drive Groups*. Only one controller can own a drive group, but we are not limited to which controller owns a specific drive group. Thus, if one drive group contained a LUN which was receiving heavy I/O, then you would want to put that drive group on one controller and the rest of the LUNs on the other controller.

9.3.2.1 AIX

To automatically balance Arrays under AIX, perform the following:

1. Type **smitty fdar_menu**
2. Select the appropriate dar
3. Change Load Balancing to **yes**. By default, the DAR will balance the FC-RSS LUNs every 600 seconds as specified by the Load Balancing Frequency.

To manually balance arrays perform the following:

1. Type **smitty fdacc_menu**
2. Select **Change/Show Disk Array Ownership**
3. Select the DAC that you want to move the Array *from*.
4. Press **F4** under **Select Drive Group to Switch** to list the available Arrays.
5. Select the drive group and press **enter**.

9.3.2.2 HP-UX, Solaris and Windows NT

To optimize your configuration by balancing LUNs perform the following steps:

1. Determine which drive groups receive the most I/O. The best way to determine this is to use the Status Application's *Performance Monitor* tool. Select the LUNs and controllers and let the application run for a few days to determine which LUNs receive the most I/O.
2. Use the *LUN Balancing* tool under the Maintenance and Tuning application to balance out the activity between the two controllers.
3. Repeat step 1, to confirm that the balancing actions you performed have balanced out the I/O to the two controllers. Ideally you would want to have 50% of the total operations being handled by each controller.

A second form of balancing LUNs is to *create high I/O LUNs within different drive groups*. For example, if your FC-RSS is going to support both a heavily loaded Web server and database, do *not* create the LUNs for the Web server and database within the same drive group. You will have better performance by spreading them out over two drive groups.

Important: Two (or more) LUNs within the same drive group share the same physical drives, controllers and the same FC interface. Therefore, to maximize performance between multiple LUNs, balance I/Os across controllers, drives, PCI busses and FC interfaces.

9.3.3 LUN Segment Sizes

LUN Segment Size determines the number of contiguous blocks of data the controller will write to a single drive before switching to the next drive in the same LUN. It provides the host system a method of tuning the striping of data relative to the typical I/O request size.

For HP-UX, Solaris, and Window NT, see Chapter 7.1.2.3, “Modifying Segment Size” on page 81 for instructions on modifying the segment size.

For AIX, see Chapter 7.2.3, “Modifying Array Attributes” on page 97 and select **Segment Size (blocks)**.

A large LUN segment size is good for read performance in most RAID levels, since the controller will have to access fewer drives to obtain the data (which means that the other unused drives may be used for other IO operations). Large LUN segment sizes are typically useful for applications that require high I/O throughput. Smaller LUN segment sizes allow for more parallelism.

Small LUN segment sizes are typically useful for applications that require large numbers of small I/Os to be processed quickly.

Segment Size is defined as the amount of data that is written to a drive before advancing to the next drive in the drive group. This parameter, which is initially set during LUN creation, can be later modified on a per -UN basis.

It is a general rule that larger segment sizes are appropriate for sequential I/O, while a smaller segment size is designed for random I/Os. Secondly, the segment size should be greater than or equal to the size of your typical I/O size.

For random I/Os, the appropriate segment size is greater than the typical I/O size, but small enough to spread the LUN across all disks.

For sequential I/Os, it depends on the number of processes doing sequential I/O. For a single sequential I/O process, a smaller segment size will enable the controller to quickly get I/Os going to all the drives in the LUN. For multiple sequential I/O processes to the same drive group, larger segment sizes might prevent disk contention by reducing the chance that two processes will be sent to the same disk.

Note that as I/Os are processed at the application layer, the I/Os may be coalesced or split up as they move to the filesystem, LVM, adapter device driver, controller, then disk layers, further complicating the issue. Sometimes, one of these layers will wait for one I/O to complete before initiating the next I/O, thus hampering throughput.

Before and after changing segment sizes, use the Status Applications Performance Monitor tool to see if your changes have had any effect on the performance of your LUNs.

Chapter 10. Backup, Recovery and Availability

In this chapter, we discuss some backup, recovery and availability considerations to provide ongoing business access to data stored on the FC-RSS.

10.1 Backup

It is critically important to provide recovery information in the event of a disaster or loss of access to data. This data could be the operating system or application data. Different forms of data may require different backup requirements, and this should be allowed for.

In this chapter we explain how to back up application and system data using different methods to provide backup and restore services, and data management for your environment.

This can be done by using mirroring, RAID-5, a remote backup center, or software products such as ADSTAR Distributed Storage Manager (ADSM). You can also delegate this responsibility to another company, such as IBM Global Services (IGS). We discuss mirroring and RAID-5 in Chapter 2.3.3, "RAID Technology" on page 24, so they are not covered in this chapter.

It should be noted here that the backing up of FC-RSS data and its recovery are no different from those for data stored on other forms of disk, such as direct SCSI.

For more information on ADSM, please refer to the redbooks, *ADSM Concepts* (SG24-4877) and *Using ADSM Hierarchical Storage Management* (SG24-4631), and the ADSM Web site, www.storage.ibm.com/adsm.

10.2 Types of Failures

The FC-RSS is designed to provide data integrity in every failure situation, and data availability in most cases where a single failure occurs. Several components are redundant and provide predictive failure detection, which allows planned maintenance actions. These failures may or may not lead to some problems in data access.

In the FC-RSS, there are five main categories of possible subsystem component failure:

- Fibre connection failures
- Controller failures
- Cache failures
- Disk drive failures
- Power failures

10.2.1 Fibre Connection Failures

FC-RSS host interface adapters have two fibre ports that can be connected to either the same host or different hosts. Each port is connected to a separate

controller, and the fibre port-to-controller connection is defined at installation time.

A failure in the host fibre interface, the FC-RSS interface port, or the fibre channel between them will interrupt access to data from that interface. The ongoing I/O operations in the adapter during the failure will not complete. The adapter must be replaced before normal operation is resumed.

You can use disk mirroring across fibre ports on different interface adapters to obtain alternative paths to data. Disk mirroring provides two (or more) connections from a given host to copies of the same data inside FC-RSS. Should one interface fail, a mirrored copy of the data can be accessed from another fibre port.

10.2.2 Controller Failures

FC-RSS can handle controller failures. They do not cause loss of data access or data integrity.

Each controller contains the microcode needed to run the FC-RSS operations, along with diagnostics and configuration information.

Host interfaces and disk interfaces are connected to both controllers through PCI bridges, so all the interface connections can be accessed by one controller, in a failover scenario. In normal operation, the fibre ports of the host interfaces can be logically connected to any controller, even though they are physically installed on a specific port.

The two controllers perform a heartbeat operation on each other, to ensure that both are available. They send the heartbeat messages through the PCI bus.

In the unlikely event of a controller failure (for example, *Controller A*), the heartbeat operation will report the error to *Controller B*, or the heartbeat operation detects the failure (if the entire controller is inactive and cannot report errors) and starts a failover.

The I/O operations ongoing in the failed controller will not have completed. They will be timed out by the host device driver and redriven, at which point the available controller will process them.

All data flow is routed through *Controller B*, and the disks connected to the *Controller A* will be accessible after the failover. Normal data access will be resumed.

Once the failed controller is operational again, a *fallback* is performed, to put the controller back to operation. The FC-RSS returns to its normal state.

10.2.3 Cache Failures

By default, all cache data is mirrored between the two cache controllers. In the event of a cache failure in one of the controllers, the data is still accessible through the alternate control unit's cache. Cache data integrity is maintained through battery backup in the event of a cache power failure, as well as the mirroring performed by the alternate controller in the event of a cache card failure.

10.2.4 Disk Drive Failures

Single disk drive failures are transparent to the user, and data access to the application is maintained through the use of different RAID configurations and the availability of automatic hot spare processing. Please see Chapter 2.3.3, “RAID Technology” on page 24 for a description of the redundancy capabilities of the different RAID configurations that are available.

10.2.5 Power Failures

FC-RSS has redundant power supplies for both the processing controller and the disk drawers, and two separate power cords. In case of a single power supply failure, the remaining power supply can guarantee the normal operation of the system. Cooling fans are also redundant. Both can be replaced without shutting down the system, and without affecting data integrity.

Depending on the location, power supply can be either single-phase, two-phase, or three-phase.

10.3 Data Integrity

FC-RSS provides full data integrity, regardless of any single failure that may occur. Once data is written to FC-RSS, it is mirrored between the two controllers for further protection against loss of data availability. The two controllers are backed by independent batteries. It is then destaged to disk, and data integrity is maintained.

Finally, a single disk failure does not affect data integrity in RAID-1, RAID-3, RAID-5, or RAID-10 configurations; the failed disk will be replaced by a hot spare in the storage unit.

10.4 System Availability

Adapter failures, or the actions required to repair the system, can disrupt the environment. You should configure the system to eliminate this disruption. Consider all options when planning for system availability.

10.4.1 Planning for Maximum Availability

The FC-RSS has two host interface ports (one per controller) and six SCSI ports. The configuration of these ports effects performance and availability.

10.4.1.1 Minimal Disruption in Case of an Adapter Failure

When a host adapter fails, data flow through it ceases. The hosts connected to it cannot access the data until the card is replaced, or the logical volumes are reassigned. Logical volumes can be reassigned to another port already connected to the host, or to a new port can be assigned to the host.

Several vendors have developed software to provide multiple paths to a disk. This software can be used with FC-RSS to provide transparent failover support in case of a fibre adapter failure.

10.4.1.2 Minimal Disruption Caused by Maintenance Actions

To create minimal disruption when performing maintenance, the FC-RSS has been designed with redundant components so that they can be replaced or maintained without the loss of access to data.

Part 3. Appendices

Appendix A. SMIT Menus

This appendix contains a map of the SMIT interface displaying the options listed when configuring the FC-RSS under AIX. Figure 55 shows the SMIT menu logic path.

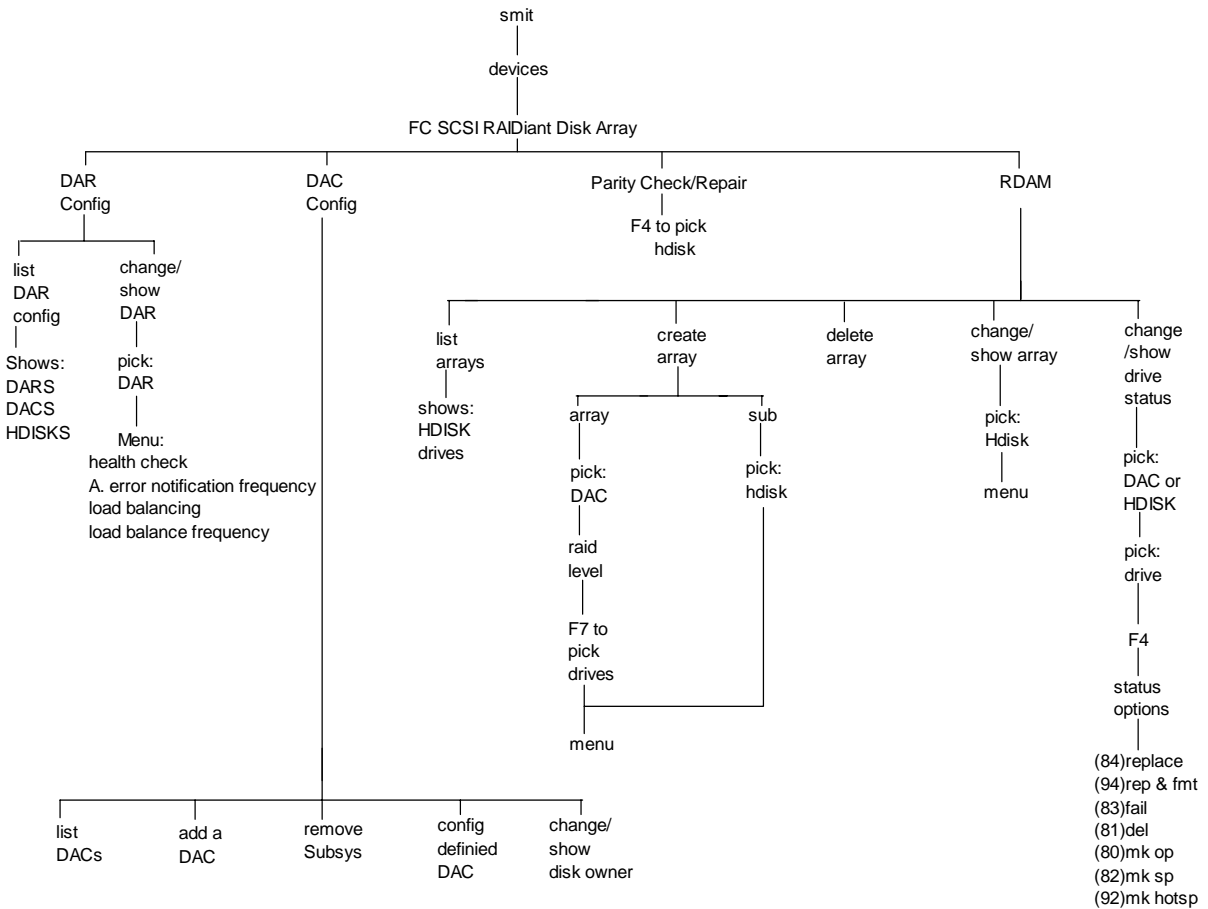


Figure 55. SMIT Menus

Appendix B. Support

B.1 Installation Support

Techline provides pre-sales technical support. Pre-sales and post-sales installation and non-defect technical support is provided by the Advanced Technical Support - Storage Systems Center in San Jose, California.

B.2 Field Support

Contact your local Storage Brand specialist or Techline for technical marketing assistance.

Installation and technical support is available through ViewBlue and WWQ&A (Worldwide Question & Answer) formerly known as ASKQ. Questions should be researched in publications and ViewBlue/WWQ&A using the search word "2102" before submitting a ViewBlue/WWQ&A question.

IBM Business Partner - Sales Agents are provided installation and technical support through IBMLink/PartnerLine. Questions should be researched in ServiceInformation--Search using the search words "2102", and in publications before submitting an ASK question.

Appendix C. Special Notices

This publication is intended to help customer management, storage administrators, and technical support staff understand, plan, implement and administer the new IBM Fibre Channel RAID Storage Server. The information in this publication is not intended as the specification of any programming interfaces that are provided by IBM Fibre Channel RAID Storage Server. See the PUBLICATIONS section of the IBM Programming Announcement for the IBM Fibre Channel RAID Storage Server for more information about what publications are considered to be product documentation.

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 IBM's product, program, or service may be used. Any functionally equivalent program that does not infringe any of IBM's intellectual property rights may be used instead of the IBM product, program or service.

Information in this book was developed in conjunction with use of the equipment specified, and is limited in application to those specific hardware and software products and levels.

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 the IBM Director of Licensing, IBM Corporation, 500 Columbus Avenue, Thornwood, NY 10594 USA.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact IBM Corporation, Dept. 600A, Mail Drop 1329, Somers, NY 10589 USA.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The information contained in this document has not been submitted to any formal IBM test and is distributed AS IS. The information about non-IBM ("vendor") products in this manual has been supplied by the vendor and IBM assumes no responsibility for its accuracy or completeness. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

Any pointers in this publication to external Web sites are provided for convenience only and do not in any manner serve as an endorsement of these Web sites.

Any performance data contained in this document was determined in a controlled environment, and therefore, the results that may be obtained in other operating

environments may vary significantly. Users of this document should verify the applicable data for their specific environment.

Reference to PTF numbers that have not been released through the normal distribution process does not imply general availability. The purpose of including these reference numbers is to alert IBM customers to specific information relative to the implementation of the PTF when it becomes available to each customer according to the normal IBM PTF distribution process.

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

ADSTAR	AIX
AS/400	IBM
RS/6000	S/390
SP2	

The following terms are trademarks of other companies:

C-bus is a trademark of Corollary, Inc. in the United States and/or other countries.

PentiumPro is a trademark of the Intel Corporation in the United States and/or other countries.

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

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

PC Direct is a trademark of Ziff Communications Company in the United States and/or other countries and is used by IBM Corporation under license.

ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States and/or other countries. (For a complete list of Intel trademarks see www.intel.com/tradmarx.htm)

UNIX is a registered trademark in the United States and/or other countries licensed exclusively through X/Open Company Limited.

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

Appendix D. Related Publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this redbook.

D.1 International Technical Support Organization Publications

For information on ordering these ITSO publications see “How to Get ITSO Redbooks” on page 127.

- *IBM Versatile Storage Server*, SG24-2221
- *Monitoring and Managing IBM SSA Disk Subsystems*, SG24-5251
- *Configuring the IBM VSS for Performance and Availability*, SG24-5279
- *ADSM Concepts*, SG24-4877
- *Using ADSM Hierarchical Storage Management*, SG24-4631
- *Implementing Netfinity Disk Subsystems: ServeRAID SCSI, Fibre Channel and SSA*, SG24-2098

D.2 Redbooks on CD-ROMs

Redbooks are also available on CD-ROMs. **Order a subscription** and receive updates 2-4 times a year at significant savings.

CD-ROM Title	Subscription Number	Collection Kit Number
System/390 Redbooks Collection	SBOF-7201	SK2T-2177
Networking and Systems Management Redbooks Collection	SBOF-7370	SK2T-6022
Transaction Processing and Data Management Redbook	SBOF-7240	SK2T-8038
Lotus Redbooks Collection	SBOF-6899	SK2T-8039
Tivoli Redbooks Collection	SBOF-6898	SK2T-8044
AS/400 Redbooks Collection	SBOF-7270	SK2T-2849
RS/6000 Redbooks Collection (HTML, BkMgr)	SBOF-7230	SK2T-8040
RS/6000 Redbooks Collection (PostScript)	SBOF-7205	SK2T-8041
RS/6000 Redbooks Collection (PDF Format)	SBOF-8700	SK2T-8043
Application Development Redbooks Collection	SBOF-7290	SK2T-8037

D.3 Other Publications

These publications are also relevant as further information sources:

- *Fibre Channel RAID Storage Server, Model 2102, Introduction and Planning Guide*, GC26-7281
- *Fibre Channel Storage Manager, Installation Guide for Windows NT*, SC26-7283
- *StorWatch Fibre Channel RAID Specialist for Windows NT & Windows 95*, SC26-7284
- *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 HP-UX Operating System, SC26-7287*
- *Fibre Channel Storage Hub Installation, Service and User's guide, SC26-7289*
- *Fibre Channel RAID Storage Server, Fibre Channel Storage Unit Service Guide, SY27-7604*
- *Seascope Solution Rack, Installation and Service Guide, SY27-7606*

How to Get ITSO Redbooks

This section explains how both customers and IBM employees can find out about ITSO redbooks, CD-ROMs, workshops, and residencies. A form for ordering books and CD-ROMs is also provided.

This information was current at the time of publication, but is continually subject to change. The latest information may be found at <http://www.redbooks.ibm.com/>.

How IBM Employees Can Get ITSO Redbooks

Employees may request ITSO deliverables (redbooks, BookManager BOOKs, and CD-ROMs) and information about redbooks, workshops, and residencies in the following ways:

- **Redbooks Web Site on the World Wide Web**

<http://w3.itso.ibm.com/>

- **PUBORDER** – to order hardcopies in the United States

- **Tools Disks**

To get LIST3820s of redbooks, type one of the following commands:

```
TOOLCAT REDPRINT
TOOLS SENDTO EHONE4 TOOLS2 REDPRINT GET SG24xxxx PACKAGE
TOOLS SENDTO CANVM2 TOOLS REDPRINT GET SG24xxxx PACKAGE (Canadian users only)
```

To get BookManager BOOKs of redbooks, type the following command:

```
TOOLCAT REDBOOKS
```

To get lists of redbooks, type the following command:

```
TOOLS SENDTO USDIST MKTTOOLS MKTTOOLS GET ITSOCAT TXT
```

To register for information on workshops, residencies, and redbooks, type the following command:

```
TOOLS SENDTO WTSCPOK TOOLS ZDISK GET ITSOREGI 1998
```

- **REDBOOKS Category on INEWS**

- **Online** – send orders to: USIB6FPL at IBMMAIL or DKIBMBSH at IBMMAIL

Redpieces

For information so current it is still in the process of being written, look at "Redpieces" on the Redbooks Web Site (<http://www.redbooks.ibm.com/redpieces.html>). Redpieces are redbooks in progress; not all redbooks become redpieces, and sometimes just a few chapters will be published this way. The intent is to get the information out much quicker than the formal publishing process allows.

How Customers Can Get ITSO Redbooks

Customers may request ITSO deliverables (redbooks, BookManager BOOKs, and CD-ROMs) and information about redbooks, workshops, and residencies in the following ways:

- **Online Orders** – send orders to:

In United States
In Canada
Outside North America

IBMMAIL
usib6fpl at ibmmail
caibmbkz at ibmmail
dkibmbsh at ibmmail

Internet
usib6fpl@ibmmail.com
lmannix@vnet.ibm.com
bookshop@dk.ibm.com

- **Telephone Orders**

United States (toll free)
Canada (toll free)

1-800-879-2755
1-800-IBM-4YOU

Outside North America
(+45) 4810-1320 - Danish
(+45) 4810-1420 - Dutch
(+45) 4810-1540 - English
(+45) 4810-1670 - Finnish
(+45) 4810-1220 - French

(long distance charges apply)
(+45) 4810-1020 - German
(+45) 4810-1620 - Italian
(+45) 4810-1270 - Norwegian
(+45) 4810-1120 - Spanish
(+45) 4810-1170 - Swedish

- **Mail Orders** – send orders to:

IBM Publications
Publications Customer Support
P.O. Box 29570
Raleigh, NC 27626-0570
USA

IBM Publications
144-4th Avenue, S.W.
Calgary, Alberta T2P 3N5
Canada

IBM Direct Services
Sortemosevej 21
DK-3450 Allerød
Denmark

- **Fax** – send orders to:

United States (toll free)
Canada
Outside North America

1-800-445-9269
1-800-267-4455
(+45) 48 14 2207 (long distance charge)

- **1-800-IBM-4FAX (United States) or (+1) 408 256 5422 (Outside USA)** – ask for:

Index # 4421 Abstracts of new redbooks
Index # 4422 IBM redbooks
Index # 4420 Redbooks for last six months

- **On the World Wide Web**

Redbooks Web Site <http://www.redbooks.ibm.com>
IBM Direct Publications Catalog <http://www.elink.ibm.com/pbl/pbl>

Redpieces

For information so current it is still in the process of being written, look at "Redpieces" on the Redbooks Web Site (<http://www.redbooks.ibm.com/redpieces.html>). Redpieces are redbooks in progress; not all redbooks become redpieces, and sometimes just a few chapters will be published this way. The intent is to get the information out much quicker than the formal publishing process allows.

IBM Redbook Order Form

Please send me the following:

Title	Order Number	Quantity

First name	Last name	
Company		
Address		
City	Postal code	Country
Telephone number	Telefax number	VAT number
<input type="checkbox"/> Invoice to customer number	_____	_____
<input type="checkbox"/> Credit card number	_____	_____
Credit card expiration date	Card issued to	Signature

We accept American Express, Diners, Eurocard, Master Card, and Visa. Payment by credit card not available in all countries. Signature mandatory for credit card payment.

List of Abbreviations

APA	All Points Addressable
AIX	Advance Interactive eXecute
IBM	International Business Machines Corporation
FC-RSS	Fibre Channel Raid Storage Server
ITSO	International Technical Support Organization
I/O	Input / Output
JBOD	Just a Bunch Of Disks
LED	Light Emitter Diode
PCI	Peripheral Component Interconnect
PDU	Power Distribution Unit
RAID	Redundant Array Independent Disks
SAN	Serial Storage Network
SCSI	Small Computer System Interface
SMIT	System Management Interface Tool
SSA	Serial Storage Architecture
VHDC	Very High Density Cable
VSS	Versatile Storage Server
VTS	Virtual Tape Storage

Index

Numerics

2101 4
2101-100
 features 10
2102 Model F10 5
2102-D00 4, 7
2102-D00 Fibre Channel Storage Expandable Unit 49
2102-F00 49
2102-F10 4
 features 6
2103 Model H07) 10
2103-H07 Fibre Channel Storage Hub 49
2105 Versatile Storage Server 3
7131 SCSI Disk System 20
7133 Serial Disk System 3, 20
7204 external disk drive 19

A

abbreviations 131
about the authors xi
acronyms 131
adding drawers to the FC-RSS 93
ADSM 113
advanced configurations 105
AIX 71
application considerations 43
Architecture 5
Assigned drive group 75
assigning drive letters 61
Attachments
 HP - UX 67
 limitations 17
 RS/6000 71
 Sun Solaris 69
 Windows NT 61
Availability 43, 113

B

Backup 113
balancing LUNs 110
BOOTP 87

C

Cabling 52
cabling the control unit 53
cabling the FC storage unit 53
cabling the hub 52
cabling the rack 52
Cache 7
Cache Block Size 109
Cache Flushing 108
cache throughput 36
changing RAID groups 103
Changing RAID Level 81
clustering considerations 104

command line 100
common services 30
Components
 5
 hardware 5
 software 14
concentrators 33
Configuration 75
configurations 50, 54
connectors 32
Controller Module 6
Controller Names 87
current IBM storage products 19

D

data integrity 115
data warehousing applications 45
day-to-day operation 101
DBMS 44
Defragmenting 82
Deleting LUNs 82
Demand Flush Amount 108
Demand Flush Threshold 108
DHCP 87
Disk Array Monitor 88
Disk Array Monitor Service 88
drive groups 75
dual controllers 3

E

Environmental Requirements 12
Ethernet Address 86
Expandable Storage Unit 8

F

failures 113
FC GBIC 50
FC-RSS 3
FC-RSS management
 features 14
Fiber Cables 50
Fibre Channel
 Fabric 31
 FC-0 29
 FC-1 30
 FC-2 30
 FC-3 30
 FC-4 31
 Gigabit Interface Converters (GBIC) 33
 Layers 29
 Loop 32
 Point to Point 31
Fibre Channel Expandable Storage Unit 4, 7
Fibre Channel hubs 3
Fibre Channel Storage Hub 10
Fibre Channel Storage Manager 4, 15

Fibre-Channel
Performance Tuning 105
framing and signaling protocol 30

G

GBIC 11, 22
general considerations 43
Gigabit Interface Converter 22
GUI interface 100

H

Hewlett Packard 16
host attachment cables 7
Host Bus Adapters 22
Hot Spare drive group 75
hot spares 79
Hot-Spares 79
HP9000 K 16

I

IBM Fibre Channel Storage Manager 99
implementation 49
Independent Host 51
Informix 45
initial setup 101
Introduction 3

J

Jaycor adapter 69

L

lad 83
LEDs 52
loop topology 24
Low Voltage Differential SCSI 49
LUN segment size 111
LVD 6

M

maintenance 103
management 99
management software 99
measurement results 36
media interface adapter 53
MIB 90
migrating disks 94
migration considerations 103
minimal disruption 115
mirroring 25
modem/switch 7
Modifying
drive group attributes 80
LUN attributes 80
segment size 81
monitoring 102
MSCS 104

Multi-host 51

N

Netfinity 15
NMS 90

O

OLTP considerations 45
one SCSI bus 54
Oracle Parallel Server 44
others servers 16
Overview 3

P

parity 26
PCI FC Adapter 12
PCI Fibre Channel Adapter 11
performance considerations 35
Performance Tuning
Balancing LUNs 110
Fibre Channel 105
LUN Segment Sizes 111
Performance tuning
Cache 106
physical interface and media 29
Physical Requirements 12
plan for maximum availability 115
point-to-point 31
ports 50
positioning the FC-RSS 19
prefetching 106

R

RAID Module Names 86
RAID Technology 24
RAID-0 25
RAID-0+1 27
RAID-1 25
RAID-10 27
RAID-3 26
raidutil 100, 106, 107
random workload measurements 38
Read Ahead Prefetching 106
Recovery 113
Remote Service 7
Requirements
environmental 12, 13
power 12
summary 17
wiring 12
Resetting the Configuration 83
rmparms File 91
RS/6000 16
rule of thumb 107

S

SCSI-1 23

- SCSI-2 23
- Seascape 21
- segment size 77
- serial port setup 101
- signaling layers 29
- signaling protocol 30
- Single Host 50
- SMIT 100
- SMIT menus 119
- SNMP 90
- SNMP Trap Message 90
- Solaris 69
- Specifications 13
- SSA 23
- storage allocation 75
- Storage Area Network 20
- StorWatch Fibre Channel RAID Specialist 4, 15, 99
- Sun Enterprise 16
- Sun SPARCservers 16
- Supported Platforms 15
- Sybase 45
- systems attachment limitations 17
- Systems Management Interface Tool 100

T

- technical support 121
- throughput rates 36
- transmission protocol 30
- two independent SCSI buses 53
- types of failures 113

U

- Ultra2-SCSI 23
- Ultra-SCSI 23
- Ultrastar disk drive 4
- Unassigned drive group 75
- unassigned drive group 76
- upper layer protocol 31

V

- verifying LAN connectivity 88
- Versatile Storage Server 20

W

- Windows NT 15, 61
- Wiring Requirements 12
- write penalty 27

ITSO Redbook Evaluation

Configuring and Implementing the IBM Fibre Channel Storage Server
SG24-5414-00

Your feedback is very important to help us maintain the quality of ITSO redbooks. **Please complete this questionnaire and return it using one of the following methods:**

- Use the online evaluation form found at <http://www.redbooks.ibm.com>
- Fax this form to: USA International Access Code + 1 914 432 8264
- Send your comments in an Internet note to redbook@us.ibm.com

Which of the following best describes you?

Customer **Business Partner** **Solution Developer** **IBM employee**
 None of the above

Please rate your overall satisfaction with this book using the scale:
(1 = very good, 2 = good, 3 = average, 4 = poor, 5 = very poor)

Overall Satisfaction _____

Please answer the following questions:

Was this redbook published in time for your needs? Yes ___ No ___

If no, please explain:

What other redbooks would you like to see published?

Comments/Suggestions: (THANK YOU FOR YOUR FEEDBACK!)

SG24-5414-00

Printed in the U.S.A.

